Universidad
Rey Juan Carlos

TESIS DOCTORAL

Capital Theory, Capital Markets and Q

Autor:

Olav A. Dirkmaat

Directores:

Jesús Huerta de Soto

Juan Ramón Rallo

Programa de Doctorado en Ciencias Sociales

Escuela Internacional de Doctorado

2018

Abstract in Spanish (resumen en español)

I: Antecedentes

La crisis financiera de 2008 manifestó que los fundamentos teóricos detrás de muchas prácticas modernas en la economía son deficientes. En primera instancia, la crisis financiera demostró la necesidad de entender el capital en un contexto mucho más amplio que el de un agregado homogéneo usado en las muchas versiones de la función de producción y en su relación con los mercados financieros. La teoría del capital está universalmente presente en la economía. Sin embargo, no hemos avanzado demasiado desde las aportaciones de los economistas clásicos. Según Smith (1776) y Ricardo (1817), por ejemplo, el capital era un simple input, el cual, en combinación con otros inputs (el trabajo y la tierra), genera una cierta cantidad de output físico.

Esta idea fundamental del capital como input físico inserto en un proceso productivo, junto con los demás inputs, tiene sus raíces en el pensamiento económico clásico. Esta teoría material y física del capital continúa vigente: Piketty (2014) la usa para justificar su posición ricardiana de la concentración de riqueza; Solow (1956) y Swan (1956) la usaban para su modelo de crecimiento económico. Así las cosas, los cimientos de la ciencia económica se tambalean. ¿Cuánto ha influido esta visión sobre la teoría del capital aún por resolver en la actual economía aplicada? ¿De qué manera dependen de esta los modelos de predicción económica que se usan en la práctica de la teoría del capital? ¿Cuál es la importancia de los acuerdos y desacuerdos sobre la teoría del capital en relación con los mercados de capital, las regulaciones bancarias, el desarrollo económico (la pobreza), la supuesta desigualdad y las políticas públicas modernas?

Durante la crisis, los modelos de crecimiento que utilizan la función de producción Cobb-Douglas, basados en la misma teoría física y material de Smith y Ricardo, fallaron en forma espectacular. Fueron incapaces de indicar las verdaderas causas del crecimiento económico (ya que la riqueza no es igual a producción u output físico) y malinterpretaron el concepto del capital. Los desajustes en la estructura de capital pasaron desapercibidos antes de la Gran Recesión. La de 2008 fue la historia del fracaso de una teoría del capital falaz. La OCDE, por ejemplo, que utiliza la función de producción Cobb-Douglas para estimar el llamado *output* *gap* (el crecimiento económico potencial de una economía), no fue capaz de predecir, aunque por poco, la Gran Recesión de 2008 en ninguno de los mayores países-miembro de la OCDE (Turner, 2016).

La función de producción está en el corazón del *mainstream* macroeconómico (e.g., Solow, 1956; Swan, 1956): tanto los modelos (de predicción) de crecimiento económico como la teoría del ciclo real (*real business cycle model*) fallaron. Sin embargo, las críticas a este planteamiento sobre el capital, que es más bien una teoría de la producción (física), no son nuevas; ya en la década de los 60, los economistas de Cambridge en Inglaterra atacaron a Solow (1956), Swan (1956) y Samuelson (1966). Sin embargo, estas posturas han sido olvidadas o incluso ignoradas: los postulados económicos continúan iguales con los mismos fallos. La teoría del capital de Solow-Swan, efectivamente, *no* es una teoría económica del capital, sino una teoría de la producción física, cuyo sitio es el mismo que su origen: una planta industrial de EE. UU. de la década de los 20. Este trabajo se ha desarrollado, por lo tanto, con la idea de poner fin al constructivismo extremo que ha caracterizado la economía desde la Segunda Guerra Mundial: la obsesión por predecir y medir la economía es un fenómeno que tiene su génesis en la posguerra, como Robert Solow admite. Los constructivistas, después de la guerra, pretendían ocuparse del crecimiento económico.

Una de las razones por las cuales perduran las teorías falaces que pretenden respaldar la economía aplicada moderna es que la alternativa que propuso Cambridge en Inglaterra no fue mucho mejor: solo se repitieron los mismos argumentos falaces que David Ricardo utilizaba hacía 200 años. Desafortunadamente, la reacción de la ciencia económica ante la crisis fue en parte buscar refugio en las mismas falacias ricardianas: el capital genera riqueza en sí, lo cual era, según muchos economistas, el gran aporte de Piketty (2014). Fuimos de una teoría fracasada del capital, que fue incapaz de prever o entender la crisis de 2008, a una igual de fracasada, que es incapaz de entender el concepto de riqueza: ahora parece que la teoría del capital (ricardiana) sirve solamente como justificación política para imponer *más* impuestos y *más* restricciones a la libertad empresarial.

Se podría decir, sin exagerar, que la economía desde la década de los 60 se bifurcó con los debates entre los neoclásicos (o neokeynesianos) y los neoricardianos de dos diferentes posturas en Cambridge. Las respuestas ante esta crisis existencial de la economía se encuentran, no obstante, en las controversias sobre el capital *anteriores* a las de Cambridge.

En muchos sentidos, la teoría del capital sirve como un mapa. Como Taleb (2012) aprecia, es mejor encontrar el camino sin mapa alguno que con uno equivocado, lo cual es una buena caracterización de la teoría del capital aplicada a la vida real durante los últimos 70 años. Sin embargo, el objetivo de este trabajo es proponer un mapa correcto, que nos ayude a navegar en la increíble complejidad de una economía moderna.

Otro antecedente era la teoría o "hipótesis" de los mercados (financieros) eficientes, los cuales, según este paradigma dominante, están por definición en equilibrio. Según esta teoría, es un mito que se puede ganar retornos en excesos del mercado en la ausencia de mera suerte. Sin embargo, con una apreciación correcta del capital, se identifica desajustes desapercibidos por los demás. El principio de *q* es, efectivamente, una teoría de equilibrio para explicar el desequilibrio, y así es capaz de revelar sub o sobrevaluaciones, revelando los retornos futuros en diferentes mercados. El deseo de demostrar la utilidad de una buena teoría del capital, en forma de *q*, es otro antecedente de este trabajo.

II: Objetivos

El objetivo, por tanto, es colocar la ciencia económica sobre una base sólida, específicamente con la integración de una teoría *subjetiva* del capital. Sin embargo, antes de presentar una visión alternativa al *mainstream* actual, tendremos que exponer los errores de los "viejos" economistas sobre la teoría del capital e inevitablemente la teoría del interés. Los objetivos de este trabajo, por ende, son múltiples:

(1) Analizar los tres principales debates históricos de la teoría del capital desde un enfoque "austríaco" o desde la perspectiva de Mises.

(2) Elaborar un análisis comparativo de las principales teorías del interés.

(3) Elaborar y extender la teoría de Menger y de Mises sobre el capital, basada en el subjetivismo y las finanzas modernas, hacia a) la teoría del empresario, b) la teoría del ciclo económico y c) la teoría del tipo de cambio. Es decir, resucitar la teoría olvidada del capital de Menger (1888), utilizarla como fundamento y analizar sus ramificaciones en un contexto moderno de mercados financieros modernos.

(4) Formular una teoría del consumo, derivada de la teoría del capital de Menger y Mises, basado en el subjetivismo y en la idea de liquidación y/o consumo de servicios, que contraste con la costumbre en economía de considerar que "consumir" equivale a "obtener propiedad" si en dado caso es el consumidor.

(5) En línea con el punto cuatro: formular una teoría del capital humano, en oposición a la idea del trabajo como "factor de producción" completamente separado del capital, que sugiere *capitalizar* los sueldos futuros esperados de un individuo como parte de su propio "balance financiero." Se cree que este ejercicio ayuda a delimitar el capital como un concepto más amplio de lo que se considera usualmente: de esta forma, el capital humano es el resultado del cálculo económico de un individuo.

(6) Demostrar que la teoría austríaca del capital no fue "refutada" en la controversia sobre el capital en las décadas de los 50 y 60: específicamente el fenómeno de *reswitching* fue, en las palabras de uno de los participantes, el nobel Samuelson (1966), una clara prueba de que la teoría austríaca del capital es errónea. Así, se intentará aplicar la curva de tipos de interés (en lugar de una sola tasa de interés) a los ejemplos específicos presentados en estos debates importantísimos y controversiales.

(7) Entender nuestra teoría del capital en términos de equilibrio y desequilibrio. En concreto: analizar el principio *q* (un derivado del trabajo de Tobin y Brainard [1969]) de la teoría del capital y analizar la relación entre el *q* como métrica de valuación y los retornos futuros esperados en el contexto de distintas categorías de activos (por ejemplo, la bolsa de valores y el mercado inmobiliario).

(8) Combinar la teoría de la curva de rendimientos con el principio de *q*. De esta manera, tratar de llegar a estimar la probabilidad de una posible caída en la bolsa o una recesión de una manera más acertada que en la literatura existente (e.g., Wright, 2004; Spitznagel, 2011).

(9) Redefinir la teoría austríaca del ciclo económico con nuestra teoría del capital, según la teoría de *preferred habitats* de la curva de rendimientos y la noción de los descalces de plazos de los diferentes intermediarios financieros.

(10) Repasar algunos de los episodios de auge y recesión más importantes en la historia con la teoría reformulada del ciclo económico del punto anterior; ofrecer otro punto de vista para entender y analizar el fenómeno del ciclo económico, basado en la práctica de descalzar plazos (*asset-liability mismatching*) de los diversos intermediarios financieros.

(11) Presentar un argumento sólido para explicar cómo la calidad de los activos de un banco central (su "respaldo," es decir, el valor de reemplazo de sus pasivos) determina los retornos sobre los pasivos del banco central en la medida en que desvía su valor o precio del mercado, usando el mismo principio *q*. En igualdad de circunstancias, después de una caída en el valor de los activos de un banco central, se esperarán retornos negativos en la medida en que el tipo de cambio empeore o la inflación doméstica aumente.

(12) Abrir un nuevo campo de investigación relacionado con las interrelaciones entre el capital, el dinero y los diversos activos financieros y tiempo (plazos/duración). (Mencionaremos las posibilidades de futuras investigaciones más adelante.)

III: Metodología

La metodología del presente trabajo consiste en los siguientes pasos:

(1) Primero, se revisará la literatura existente sobre la teoría del capital. Es inevitable, en este contexto, tratar temas y teorías auxiliares como, por ejemplo, la teoría del interés, la teoría de la curva de tipos de interés, la teoría de la intermediación financiera y la banca, y la teoría de riesgo, entre otras. Se dividirá la literatura existente en tres capítulos según un orden cronológico: primero, con la primera controversia del capital de 1888 a 1907 entre Böhm-Bawerk y J. B. Clark, más las contribuciones de Menger y otros. Segundo, con la segunda controversia del capital de la década de los 20 y adelante: este período incluye los debates entre Fisher y Hayek, Knight y Hayek, Knight y Kaldor, y Hayek y Sraffa. También se incluyen en esta fase los aportes de Mises, Williams, Macaulay, Lutz y, por último, Lachmann. Los últimos no han participado en forma directa en los debates, pero han hecho contribuciones valiosas a la teoría del capital e interés a lo largo de estos mismos años. Tercero, con la tercera (y más conocida) controversia del capital de la década de los 50 y los 60 (las famosas controversias de Cambridge, que se refieren a las universidades involucradas en Cambridge, EE. UU. [MIT] y Cambridge, Inglaterra).

Esta controversia incluía, como protagonistas, a Solow (1956), Swan (1956) y Samuelson (1958) en el Cambridge de EE. UU., y Robinson (1953), Sraffa (1960), Kaldor (1961) y otros en el Cambridge de Inglaterra. En estos últimos debates, la discusión acerca de la teoría del capital se volvió cada vez más ideológica: sin embargo, los liberales (en el sentido europeo) deberían de pensar dos veces en aceptar una teoría del capital equivocado: como Piketty (2014) demuestra, una mala justificación de los mercados solo nos lleva por el camino equivocado del populismo. La idea detrás de esta revisión de la literatura precedente es encontrar aportes valiosos para la (re)formulación de una teoría del capital consistente con la ciencia económica como ciencia subjetivista (e.g., Hayek, 1955).

(2) Segundo, encontrar y analizar las interrelaciones entre los diferentes elementos que aparentemente están aislados. Se intenta conectar las teorías del capital e interés con las teorías del empresario arbitrista, con la curva de tipos de interés, con *duración* Macaulay, con la estructura (financiera) del capital, con la *q* de Tobin, de la demanda de portafolio de activos financieros y la de descalces de plazos o duraciones. En este sentido, el presente trabajo se propone sintetizar esta gran variedad de aportes directos e indirectos históricos a la teoría del capital.

(3) Tercero, proponer una teoría del capital basada en el subjetivismo, que evite errores pasados. Discutir sus implicaciones para varios fenómenos relacionados, como el consumo, el ahorro, el interés, la liquidez y el dinero.

(4) Cuarto, proponer una teoría general del *q*, la cual, como operacionalización de la teoría del capital, se puede aplicar a una amplia gama de diferentes sectores y tipos de activos. Esta teoría combina la teoría del empresario de Huerta de Soto (2010) y Kirzner (1960) con la teoría del capital de Menger (1888) y, al final, se convierte en un método de valuación robusto y predictivo de retornos futuros.

(5) Cinco, utilizar empíricamente esta misma teoría del *q* a casos concretos: primero, a la bolsa de valores (en EE. UU.) junto con la teoría del ciclo; segundo, al mercado inmobiliario (residencial) con un estudio empírico que cubre una serie de más de 60 años en 12 países diferentes; tercero, al mercado de *forex* (divisas), incluidos en un estudio

empírico 15 países latinoamericanos y sus correspondientes bancos centrales; cuarto, al mercado del oro físico (como activo monetario real no financiero) en el cual resaltamos brevemente la aplicación del principio del *q* a activos tangibles como el oro físico, abriendo paso a aplicaciones futuras a activos similares como las criptomonedas. Se resalta, en este punto de nuestro trabajo, que los estudios empíricos solo sirven como ilustraciones de la teoría del capital que proponemos. Estos estudios, además, demuestran unas aplicaciones prácticas reales de nuestra teoría del capital.

(6) Seis, reformular la teoría del ciclo económico sobre la base de la curva de tipos de interés y la práctica del descalce de plazos (y *duraciones*). Se usa la teoría reformulada del ciclo económico austriaco para demostrar brevemente las causas reales de cuatro episodios de crisis y recesión en la historia de EE. UU.: la Gran Depresión de los años 30, la crisis de los S&L en los años 80, la crisis asiática en los años 90 y la Gran Recesión de 2008.

De esta forma se intenta elaborar un análisis sólido de las diferentes teorías del capital mientras se observa su valor práctico y las implicaciones importantes que tienen para los profesionales.

IV: Conclusiones

Las conclusiones de este trabajo son múltiples, desde contribuciones teóricas hasta empíricas. En resumen, proporcionaremos una reseña de los principales aportes de este trabajo:

1. Böhm-Bawerk (1888) criticó las teorías del interés basadas en la “productividad” en forma persuasiva: si estas fueran correctas, los empresarios simplemente arbitrarían los precios hasta que desaparecieran los excesos en los retornos. Por tanto, estas teorías no pueden explicar el fenómeno del interés. Efectivamente, sus proponentes (e.g., Clark, 1899; Ricardo, 1817) no explicaban por qué los empresarios no eliminan estos beneficios, aprovechando los retornos que ofrece un bien de capital.

   Algunos proponentes de estas teorías basadas en la productividad, como Frank Knight (1934), adoptaron una visión pesimista sobre el futuro del capitalismo. Ellos creían que, debido a los retornos decrecientes sobre el capital, los beneficios en algún momento desaparecerían, lo cual de una u otra manera lleva a una crisis existencial del capitalismo. Los economistas más modernos, como Solow (1953), basan su teoría del retorno decreciente del capital no en la noción de que los empresarios son exitosos en arbitrar y eliminar las oportunidades de beneficio, sino en la noción de que cada unidad física adicional de capital (digamos, una máquina más) tiene un retorno *físico* decreciente. Hay una brecha enorme entre los economistas que intentan estudiar economía y los economistas que confunden la producción física con la economía.

2. Böhm-Bawerk, a pesar de ello, basó su teoría del capital en la triada clásica de la producción (e.g., Smith, 1776): trabajo (salarios), tierra (renta) y capital (interés). Según esta teoría material, el capital consiste en los "bienes producidos de la producción". El capital, por tanto, es una combinación histórica de tierra y maniobra. Además, el capital es heterogéneo, ya que *ningún factor de producción es el equivalente físico de cualquier otro*. Esta noción del capital como algo físico fue adoptado luego por Cobb y Douglas (1928) en su ensayo sobre la función de producción: estos dos economistas trataron de estimar la tendencia empírica entre unos inputs materiales (específicamente el trabajo medido en semanas de trabajo y el capital medido en términos físicos expresado en dólares según un índice de bienes de capital) y un output físico (productos producidos). Así, estimaron los coeficientes que aportaron las horas del trabajo y el número de bienes de capital físicos a la producción física.

   Desafortunadamente, este triste legado de los economistas clásicos sigue persiguiendo a la economía aplicada moderna. La misma función de producción, o sus diferentes versiones, sigue siendo la herramienta más popular entre los economistas modernos para estimar el crecimiento económico futuro, para proyectar las finanzas públicas y cualquier otro tipo de predicción macroeconómica, según un método que no cambió mucho desde el *paper* de Cobb y Douglas (1928), popularizado por los participantes de la tercera ronda de debates sobre el capital: las controversias de Cambridge (e.g., Solow, 1953; Swan, 1953).

3. La controversía entre Knight, Kaldor y Hayek, como todos los debates históricos del capital, no llegó a un colmo intelectual. Las críticas de Hayek a Knight eran claras: Knight, según Hayek, no tenía en cuenta el factor del tiempo. Y esta omisión derruma su esquema. Efectivamente, Knight, como su antecesor J. B. Clark, argumenta que la teoría de la preferencia temporal y el periodo promedio de producción no puede ser correcta porque el capital es "perpetuo," en el sentido de que desde el momento de su creación (su

"génesis") se renueva automáticamente. Ya que se considera la depreciación en el cálculo de beneficios, el capital es una fuente de ingresos inagotable. Ya no requiere abstinencia; la abstinencia solo fue necesaria en el principio, cuando los primeros seres humanos tenían que esforzarse para "establecer" el primer arroyo. Desde aquel momento, la teoría de la abstinencia ya no tiene un papel. Knight, como explicó Hayek, se equivocó. El mero hecho de que uno tiene en cuenta la depreciación en el cálculo de las ganancias/perdidas no le impide consumir todo el capital. Cada flujo de efectivo implica una nueva decisión intertemporal y, por tanto, una infinidad de nuevos actos de "abstinencia." Lejos de ser "perpetuo" y "automático", mantener el capital es una decisión diaria que requiere de una abstinencia deliberada continua. Incapaz de reconocer el papel del tiempo, Knight nunca consigió articular una teoría del interés coherente: el interés simplemente era igual a la tasa de ganancia. Sin embargo, Knight criticaba a Hayek y a Kaldor por sus teorías materiales del capital: un bien de producción no es diferente de un trabajador o un parcial de tierra desde un punto de vista económico. Distinguir, por tanto, entre el trabajo (salarios), la tierra (renta) y el capital (interés) carece de sentido. El error de Hayek, según Knight, era su orientación al pasado (y así arbitrariamente determinar qué es un bien producido y qué es un bien original), mientras que el capital solo se preocupa por el futuro. La teoría subjetiva del capital de Knight, en este sentido, es superior a la teoría material del capital de Hayek.

4. Fisher (1930) hace una contribución extremamente valiosa a la economía: contrariamente a las teorías "ingenuas" del interés, basadas en la productividad física inherente al capital (como los árboles de manzanas naturalmente rinden manzanas), complementa la teoría del interés basada en la abstinencia o, mejor dicho, las preferencias temporales (e.g., Mises, 1949). Fisher (1930) está de acuerdo con la teoría de la preferencia temporal, pero solo como una explicación del lado de la oferta en el mercado *intertemporal*. Sin embargo, no explica satisfactoriamente el otro lado de la ecuación, necesario para explicar el fenómeno de la tasa de interés de *mercado,* el cual es el lado de la demanda en el mercado intertemporal.

   La demanda, explica Fisher (1930), depende de las "oportunidades de inversión." Cuando los empresarios hacen un buen trabajo en arbitrar los diferentes precios en desequilibrio, habrá pocas oportunidades de inversión y, por tanto, poca demanda en el mercado *intertemporal* de los diferentes empresarios. En cambio, cuando hay muchos desajustes en la estructura de precios, existen muchas oportunidades de beneficio (es decir, altos retornos) y, por tanto, alta demanda en el mercado *intertemporal*. Los dos factores, las preferencias temporales subjetivas y las oportunidades de inversión, explican el fenómeno del interés. De esta forma, Fisher (1930) contribuye a las teorías del interés basadas exclusivamente en la preferencia temporal.

5. Los neoclásicos de la controversia de Cambridge se contradicen en su teoría del capital en varias formas: si el capital es un input que se mide en términos monetarios, y si se necesita una tasa de interés para descontar flujos para llegar a tal estimación en términos monetarios, y si la tasa de interés, según la costumbre neoclásica, equivale a la tasa de utilidad (*rate of profit*), entonces han caído en un razonamiento circular. Efectivamente, los neoclásicos (e.g., Solow, 1963) se han "autorrefutado". Otra crítica correcta a la teoría neoclásica del capital de los economistas de Cambridge en Inglaterra consiste en el hecho de que no se puede "agregar capital" y luego usar el resultado como input en un modelo "input-output."

   Es decir, como podemos apreciar en Menger (1888) y Mises (1949), el capital es una herramienta empresarial de cálculo económico que le permite al empresario estimar futuros beneficios y pérdidas. Es decir, el capital es un *resultado* del proceso productivo, ya que consiste en el valor presente de los ingresos que los diferentes bienes de producción pueden aportar en el futuro. El capital representa, de cierta manera, un "output" futuro más que un input del proceso productivo que genere un "output" presente.

6. El triángulo de Hayek está basado en un concepto de capital erróneo: la cercanía al consumo final no es lo importante, sino la *duración* ("weighted maturity") de los flujos de efectivo. Es decir, que un proyecto sea capital-intensivo y que tenga una larga duración no implica que esté próximo o lejos de las industrias o etapas cercanas al consumo. En este trabajo, vemos cómo los plazos de los activos productivos se han comportado a lo largo de los años en función de cambios en las tasas de interés, utilizando datos históricos.

7. El supuesto problema del *reswitching* del capital no es un problema. Los economistas de Cambridge en Inglaterra (e.g., Robinson, 1953; Sraffa, 1960; Pasinetti, 1966) pensaban que el problema de *reswitching* de capital refutaba completamente la teoría del capital de

Solow, Swan y Samuelson y, por tanto, todas las pretensiones neoclásicas de usar la función de producción como fundamento en su economía aplicada. El problema del *reswitching* consiste en el hecho de que, según la teoría de Böhm-Bawerk y Mises y la aplicación de la función de producción de Solow (1956) y Swan (1956), en la cual se "mide" el "stock de capital" en términos monetarios, que luego dependen de una tasa de descuento, es decir, la tasa de interés, algunas técnicas o ciertos valores netos presentes (VPN) parecen más rentables a una tasa de interés alta y menos rentables a una tasa de interés media, pero otra vez rentables a una tasa de interés baja. Esta paradoja o deficiencia refuta la idea de que una técnica/VPN es más favorable a una tasa de interés alta, mientras que otra técnica/VPN es más favorable a una tasa de interés baja. La noción de que la tasa de interés determina el *roundaboutness* de la estructura del capital o el "periodo de producción" es falaz, según los economistas de Cambridge en Inglaterra.

Se proponen dos soluciones a este problema: (1) utilizar la curva de tipos (en lugar de *una* sola tasa de interés) en los casos que arroja el fenómeno y (2) la metodología de las tasas de interés "múltiples" de Osborne (2014), que consiste en utilizar tanto las tasas ortodoxas como *no* ortodoxas en el cálculo de VPN, ya que las ecuaciones para descontar flujos son polinómicas. Varios casos ejemplifican que el fenómeno de *reswitching* desaparece. Los dos enfoques tienen mucho en común, ya que el segundo es una forma de "pesar" la tasa de descuento respecto del momento en el cual ocurren los flujos, lo cual incluso lleva a una mejor aproximación al concepto de duración (Macaulay, 1938). El fenómeno de *reswitching* fue percibido por primera vez por Fisher (1930), pero no le dio seguimiento (era más bien una simple curiosidad, de poca importancia), pero se volvió sumamente importante en la controversia de Cambridge.

8. El presente trabajo consigue una importante reflexión sobre el olvidado debate en economía de comienzos del siglo XX sobre el *shiftability*. Según los proponentes de la teoría del *shiftability*, los bancos comerciales podían invertir en activos ilíquidos (por ejemplo, hipotecas en el largo plazo, crédito del consumo en el mediano plazo y bonos corporativos en el largo plazo), ya que el desarrollo de los mercados secundarios en los mercados de capital significaba que un banco siempre se podía hacer "líquido" vendiendo sus "activos ilíquidos" a otros bancos más fuertes. Sin embargo, este punto de vista supone que la "iliquidez" ocurre en forma distribuida a lo largo del tiempo, pero la "iliquidez" en la mayoría de los casos no afecta a bancos en forma aislada, sino al sistema

bancario completo de golpe en un periodo donde la mayoría de los bancos permanecen "ilíquidos". Esto ocurrió en la Gran Depresión de la década de los 30. Desafortunadamente, los teóricos a favor del *shiftability* vencieron a los teóricos de la liquidez, a pesar de la Gran Depresión y debido al inicio de la Segunda Guerra Mundial, que favoreció la inversión en deuda pública en el largo plazo mediante la Reserva Federal para financiar el esfuerzo bélico de los aliados.

9. La teoría del interés de Fisher ha dado otro aporte importante: Fisher (1930), posiblemente por su trabajo no académico, reconoció que existe una dinámica entre los demandantes de bienes presentes y los oferentes de bienes presentes en el mercado *intertemporal*, lo cual explica por qué la tasa de interés en el largo plazo (contraria a la tasa en el corto) se mantiene tan estable en el tiempo. Cuanto más ahorro disponible hay para la inversión productiva, de más recursos disponen los empresarios para arbitrar precios, ganar utilidades y reducir las tasas de beneficios en el mercado. *Ceteris paribus*, si las inversiones se vuelven menos atractivas, se reduce la demanda de bienes presentes, lo cual hace caer las tasas de interés del mercado y disuade a los oferentes, en el margen, de seguir ahorrando al mismo ritmo o ahorrar más (las observaciones empíricas de una relación inversa entre las tasas de interés y el consumo sustentan este mecanismo de retroalimentación). Ahora, cuando los ahorradores empiezan a sustituir, en el margen, ahorro por consumo, reducen la oferta de bienes presentes en el mercado *intertemporal*. Esto aumenta la tasa de interés y hace que los empresarios arbitristas tengan menos recursos disponibles para llevar a cabo sus *empresas*[1], lo cual causa mayores desequilibrios en el sistema de precios por ausencia de arbitraje y, por tanto, produce utilidades y tasas de beneficios más altas. Esto provoca, de nuevo, la entrada de empresarios y la *demanda* en el mercado intertemporal.

   Este proceso dinámico en el mercado intertemporal se visualiza como un péndulo, que oscila a ciertos extremos pero, por la existencia de mecanismos naturales de retroalimentación, siempre gravita en torno a una media. Esta se denomina teoría del péndulo de retornos de Fisher (1930) y aplica, incluso, a otros ámbitos, como el reciente

[1] Uno debe de apreciar, más en el castellano, el énfasis que hace Huerta de Soto (2006) en la definición de la RAE de 'empresa', la cual actualmente es: "acción que entraña dificultad y cuya ejecución requiere decisión y esfuerzo".

debate de la inversión "pasiva" en los mercados de capital; estos inversores "pasivos" simplemente invierten en un índice, obteniendo el retorno "promedio" del mercado, en lugar de discriminar entre empresas. Cuanta más inversión "pasiva," más desajustes y más oportunidades de beneficio. De esta forma, cuando la inversión pasiva llega a cierto punto, entrarán inversionistas "activos" que buscan retornos más altos, hasta que desaparezcan y se reduzcan de nuevo.

10. La teoría del empresario es sumamente desarrollada (Mises, 1949; Kirzner, 1970; Huerta de Soto, 2010). El empresario es un arbitrista de oportunidades de beneficio que están implícitas en la estructura de precios presente y futura. Tiene una función coordinadora hasta que, en equilibrio, ya no hay ninguna diferencia en los precios que le pueda proveer de beneficios. El empresario cumple un papel económico que no requiere de recursos. Sin embargo, esta teoría del empresario se ha enfocado en la teoría del empresario *no financiero*. Los empresarios *financieros* determinan cuáles de sus homólogos *no financieros* reciben los recursos y bajo qué condiciones. Esencialmente, son iguales a sus pares no financieros, pero se mueven exclusivamente en los mercados de capital. De esta forma, son los intermediarios entre los proveedores de los recursos (los ahorradores o capitalistas) y los empresarios no financieros. Este enfoque abre un campo nuevo de investigación sobre las dinámicas entre los empresarios financieros y los no financieros. Así, se sintetiza la estructura de capital de Lachmann con la teoría del empresario de Mises, Kirzner y Huerta de Soto.

11. El *capital* se define como el *patrimonio financiero* ("net worth"), en línea con las definiciones de Carl Menger (1888) y Ludwig von Mises (1949), la única definición coherente desde un punto de vista subjetivista. En esta teoría, el empresario, en tanto arbitrista de precios y asesor del capital que le corresponde, es el protagonista e impulsor del rumbo de la economía, en lugar del economista, que arbitrariamente clasifica bienes de producción según algunas características físicas u objetivas desde su torre de marfil. Esta teoría financiera del capital está basada en el cálculo económico, emprendido por el propio empresario. En esencia es una teoría del capital "forward-looking" (orientada al futuro), que no "backward-looking" (orientada al pasado).

    No pretende clasificar bienes: solo establece las diferentes escalas del capital y el arbitraje empresarial entre sí mismas y postula que los mercados de capital (incluida la banca) están íntimamente conectados con la estructura de capital, ya que las decisiones de los ahorradores (los dichos capitalistas) determinarán el plazo o la duración de la inversión de los recursos no consumidos de una sociedad ahorradora capitalista. La estructura de capital puede ser más o menos líquida, según las preferencias temporales de los ahorradores y las oportunidades de inversión de los empresarios en cada plazo. La "escala" del capital se refiere a que es aplicable a muchos niveles: el nivel del activo solitario (por ejemplo, un edificio), el de la empresa (un conjunto de bienes múltiples) o el de la bolsa de valores (un conjunto de empresas). Entre cada "nivel," cada escala, existe un arbitraje. Esto es efectivamente igual a la ratio *q* de Tobin y Brainard (1976), quienes establecieron una teoría del arbitraje entre los derechos financieros (deuda y *equity*) y los activos subyacentes que los respaldan. Los precios de los activos productivos individuales se denominan (en conjunto) el "valor de reemplazo," mientras que los precios cotizados en bolsa de la deuda y *equity* representan el "valor del mercado." La diferencia entre los dos es el *q*. Esta teoría es una formulación estrecha de la teoría del capital (y del principio *q*) que proponemos aquí. En efecto, nuestra definición del capital, basada en la función empresarial, el cálculo económico y el subjetivismo, nos lleva naturalmente a la teoría de *q*.

Un bien específico, como un edificio, es un "capital" equivalente al valor presente neto de los flujos o servicios futuros que es capaz de generar. Una combinación de bienes, como una empresa, es un "capital" equivalente al valor presente neto de los flujos futuros que es capaz de generar. Ambos son "capitales" a diferentes "escalas." La ratio *q* surge a raíz de estas diferencias de "escala." Sin embargo, hay una tendencia a que los precios del conjunto se igualen a la suma de sus partes; de lo contrario, existirían oportunidades de para los empresarios en el sentido de Kirzner. Por tanto, el principio *q* es caracterizado por una regresión a la media: nunca puede desviarse permanentemente de su punto de "equilibrio." La teoría del capital y la teoría del *q* son una.

12. Se presenta brevemente la tesis de que el "*balance sheet approach*" a la economía es la metodología *sine qua non* de la escuela austríaca. Este método requiere la misma coherencia y solidez lógica que el método lógico-deductivo característico de los economistas austríacos (e.g., Mises, 1949). Lachmann (1956), en este contexto, hace un aporte valioso casi olvidado de una estructura del capital plasmada en los balances de tres contrapartes: la estructura de "activos", la estructura de "control" y la estructura de

"portafolio". A partir de definir un sistema de contabilidad lógicamente cerrado, podemos entender mejor las interrelaciones entre los diferentes partidos, en este caso cuánto influyen las decisiones de los ahorradores en las decisiones de inversión y bajo qué premisas, lo cual determina cómo y las condiciones en que los empresarios pueden obtener financiación y, por tanto, qué clase de procesos de producción (y a qué duración) pueden emprender sin arriesgarse y caer en la ruina. Otros enfoques a la teoría del capital basados en el "*portfolio approach*" de la demanda del dinero y activos financieros (en línea con la teoría de la demanda del dinero que presenta Ludwig von Mises [1912]). Concluimos que la tarea de los intermediarios financieros es, por ende, asegurar que las preferencias temporales coincidan con la estructura temporal del capital (Bagus & Howden, 2014).

13. Aplicando la teoría del capital y del *q* a la bolsa valores, llegamos a una ratio llamada *equity q*. En este caso, ya que en la bolsa de valores se compra y vende *equity* en lugar de deuda, el valor de reemplazo de una firma equivale a los precios presentes de los activos productivos subyacentes de una empresa *neto* de la deuda. Concluyamos que el *equity q* es capaz de predecir retornos futuros en la bolsa de valores y que, además, combinado con el *spread* de la curva de tipos (el diferencial entre el tipo de interés en el largo plazo y el tipo de interés en el corto), es un excelente indicador de las caídas bursátiles de más de 20% en el corto plazo. El *equity q*, por ende, tiene un atractivo teórico y práctico. Es más, se demuestra y argumenta en esta tesis que en el ciclo económico los primeros precios que se mueven son los financieros (es decir, los de los activos financieros).

14. En una nueva contribución (según el conocimiento del autor, ningún académico lo ha intentado antes) se extienden los principios de la teoría del capital, el empresario arbitrista y el concepto de *q* en el mercado inmobiliario (residencial). Igual al *equity q*, se muestra que el *housing q* predice los futuros retornos en viviendas. Es más, igual al *equity q*, se demuestra que en sus extremos –según datos históricos– consigue indicar retornos esperados extremadamente bajos. Es otra prueba de que el *q* funciona como un afán de los empresarios (financieros o no financieros). Se trata del indicador más robusto de los retornos futuros en los mercados de capitales.

15. También se aplica el principio del *q* en el mercado de las divisas y a los tipos de cambio. Este principio está fundado en algunas nociones de la teoría del "*backing*"; el valor de

mercado de los pasivos que un banco central emite dependerá de la calidad de los activos de dicho banco central. A pesar de que ya no existe convertibilidad *directa* entre los pasivos y activos del banco central (como existía en el antiguo sistema de reservas de oro), aún son importantes los activos que respaldan el dinero emitido por el banco central. Si son de mala calidad, por ejemplo, los retornos (ajustados por riesgo) sobre los activos del banco central no bastarían para defender su moneda con ventas en el mercado abierto, pagar intereses sobre sus depósitos y deuda o cubrir sus gastos operacionales. La única salida en estos casos es una devaluación contra las divisas extranjeras o tolerar una inflación doméstica alta.

16. Por último, al menos respecto de la teoría del q, se aplica el mismo principio en el mercado de los activos monetarios no financieros (reales), en este caso, el oro. En un primer intento, se intenta desarrollar una teoría basada en los mismos principios que pueden indicar el nivel de sub o sobrevaluación en el mercado de oro y de otros activos monetarios, como las criptomonedas. Estas se comportan muy parecido al oro, ya que sus cantidades son limitadas y no son el pasivo de nadie. Se demuestra que la misma lógica se puede aplicar a esta clase de activos.

17. Se refina la teoría austríaca del ciclo económico en cuatro puntos clave: el descalce de plazos en la estructura del capital, la interacción entre el arbitraje sobre la curva de rendimientos y los descalces de plazos, la dinámica entre el ciclo y la ratio *q*, y un enfoque financiero en *duración* Macaulay en lugar de la teoría de Hayek basada en la cercanía al consumo (o sectores cercanos al consumo). Así, se reformula la teoría del ciclo económico. En resumen, el ciclo económico se caracteriza por las siguientes etapas: a) ahorradores/capitalistas mantienen un portafolio de activos líquidos en el corto plazo (en lugar de mantener la mayor parte en inversiones ilíquidas en el largo plazo); b) bancos arbitran la curva de rendimientos, expandiendo los plazos de sus préstamos e inversiones, financiados por deuda líquida en el corto plazo (principalmente, depósitos a la vista); c) las tasas de interés *a largo plazo* empiezan a caer y el *spread* de la curva de tipos se reduce; d) los precios de los activos financieros aumentan, las ratios *q* suben; e) con tasas de interés más bajas, las compañías comienzan (en el margen) a invertir en proyectos con *duraciones* más largas que son más ilíquidos; f) la inversión en el largo plazo (en capital fijo) aumenta y sube (marginalmente) los precios de los activos productivos (entre estos, las materias primas), la tasa de beneficio *promedia* o agregada (es decir, por toda la

economía) alcanza un punto máximo; g) ahorradores/capitalistas empiezan a liquidar sus activos líquidos y (en el margen) consumir ahorro (la tasa de ahorro tiende a caer). *O* los precios de los bienes presentes suben *o* las tasas de interés en el corto plazo empiezan a subir en cuanto los bancos afrontan una menor demanda de sus depósitos; h) en el margen, las compañías empiezan a quebrar debido a pérdidas *ex post* como consecuencia del aumento en los precios o las tasas más altas en el corto plazo (en pocas palabras, un WACC más alto); i) el *spread* de la curva de tipos se vuelve negativo (la curva de tipos se invierte); j) los precios de los activos financieros colapsan y las ratios *q* caen (posiblemente por debajo de sus puntos de equilibrio por la liquidación de capital); k) liquidación masiva y una crisis de liquidez: aumento en los *defaults*, los prestamistas se debilitan y los prestatarios endeudados en exceso quiebran, despidos en el mercado laboral y el capital se reduce en forma violenta; l) fase de recuperación: las preferencias temporales se alinean con la estructura financiera temporal y la duración de la estructura productiva; m) una vez la recuperación sigue su curso natural y los intermediarios financieros vuelven a arbitrar la curva de tipos, el ciclo se repite.

18. Inversionistas, empresarios y otros profesionales que utilizan el método de flujo de efectivo descontado en sus presupuestaciones de capital, deben tener en cuenta los efectos de la curva de rendimientos (tipos de interés) y de la naturaleza polinómica de los cálculos de VPN (y, así, la posibilidad de *reswitching* en la práctica). En conclusión, insistimos en que se use la curva de tipos de interés para descontar flujos futuros en vez de utilizar una sola tasa de interés. Además, posiblemente se puede aplicar el método propuesto por Osborne (2014), el cual incorpora el producto de tanto la tasa ortodoxa como las tasas no-ortodoxas al momento en que deseen descontar flujos de efectivo futuros.

19. Se ha hecho un repaso detallado y preciso de algunos episodios históricos de suma importancia de auge y recesión (la Gran Depresión de la década de los 30, la crisis de los S&L de los años 80, la crisis asiática en la década de los 90 y la Gran Recesión de 2008). Generalmente, las explicaciones existentes dicen poco sobre el descalce de plazos: con nuevos datos y un enfoque renovado se elabora una narrativa, basada en datos y hechos históricos que ocurrieron en los sistemas financieros, del ciclo económico por descalce de plazos. Esto se podrá ver como el primer intento de un estudio más amplio sobre los diferentes capítulos de recesión en las últimas décadas o incluso siglos. Con un análisis de

balances, llegamos a la conclusión de que los cuatro episodios están profundamente influenciados por el fenómeno del descalce de plazos que precedió cada una de las recesiones. Merece mencionarse que parte del aporte consiste en haber estimado los datos de los plazos de los activos y pasivos de los principales bancos comerciales estadounidenses para calcular el grado del descalce de plazo en el sistema bancario en el periodo previo a la crisis de 2008.

20. Se concluye el presente trabajo con varias sugerencias de investigaciones futuras en el campo: a) ampliar el principio *q* a otras clases de activos y a otros mercados (por ejemplo, estimar una ratio equity *q* para la bolsa de valores en China, Rusia, Japón, etcétera, o, estimar una ratio *q* para el mercado inmobiliario español y hacer un estudio histórico de la Gran Recesión de 2008 en España), b) ampliar el estudio de capital con un análisis del papel del uso de colateral en los mercados de capital, la reutilización de colateral, *collateralization*, *hypothecation* y *rehypothecation*. Los contratos de recompra (inversa) en los mercados de capital son, prácticamente, igual a los préstamos con colateral. Estos temas de investigación, sumamente importantes en los tiempos modernos, han recibido poca atención en economía. Es necesario analizar como la teoría del capital vincula el colateral a los mercados de capital y cuáles son sus implicaciones, c) extender nuestro *forex q* a un estudio de caso del Banco Central Europeo (BCE), d) aplicar nuestra versión revisada del ciclo económico a otros episodios históricos de auge y recesión, e) los incentivos perversos de descalzar plazos en el contexto de los bonus bancarios. Como Taleb (2012) sugiere, parece que los banqueros exponen sus depositantes a riesgos enormes debido a la práctica de descalce de plazos, optimizando sus propios beneficios en el corto plazo a costo de perdidas futuras, f) extender nuestra aplicación de la curva de tipos de interés al ciclo económico examinando la convexidad de las tasas de interés a los precios de capital y su comportamiento a lo largo del ciclo, g) por último, estudiar en qué medida las prácticas en el mundo real con respecto a la producción de función difieren de las practicas teóricas sugeridas en los libros de texto. Además, la reciente popularización de los modelos de equilibrio general dinámico y estocástico (DSGE), los cuales dependen en parte de la función de producción, puede ser una nueva vía de investigación.

A continuación, presentaremos una tabla que reseña los aportes históricos más importantes de los pensadores y economistas considerados en nuestra revisión de literatura. Comparamos las diferentes líneas de pensamiento en función de seis componentes: la definición de capital, las diferencias entre el capital circulante y el capital fijo, el valor del capital, la explicación del fenómeno de la tasa de interés, el papel del "factor trabajo" y, para concluir, las contribuciones más importantes del autor (o de los autores) considerado:

Autor(es)	**Teoría del capital**	**Capital circulante / capital fijo**	**Valor del capital**	**Interés**	**Trabajo**	**Contribuciones**
Böhm-Bawerk (1888)	Concepto físico, orientada al pasado, fondo de "subsistencia" o fondo "salarial"	No, el "fondo de subsistencia" implica un enfoque en el capital circulante	Productividad física (*round-aboutness*)	Preferencia temporal, productividad física del "esperar"	Factor de producción separado, requisito para el capital	Crítica de las ingenuas teorías del interés basadas en la productividad
Clark (1899), Knight (1934)	Concepto financiero, orientada al futuro	Sin importancia	Valor del mercado de los activos productivos en términos del dinero	Tasa de beneficio, productividad física, ausencia completa del factor tiempo	El trabajo es capital humano	Crítica de las teorías materiales o físicas del capital, aportes a la teoría del capital humano
Fisher (1930)	Concepto financiero, orientada al futuro	Sin importancia, teoría cruda de la curva de rendimientos basada en las expectativas	Valor del mercado de los activos productivos en términos del dinero	Preferencia temporal (oferta), "oportunidades de inversión" (demanda)	El trabajo es capital humano	Teoría del interés, elaboración de la teoría financiera del capital
Menger (1888), Mises (1949)	Concepto financiero, orientada al futuro	Sin importancia	Valor del mercado de los activos productivos en términos del dinero (enfoque de *patrimonio neto*)	Preferencia temporal	Permite integrar el capital humano	El capital como concepto financiero, papel del empresario
Hayek (1941), Lachmann (1956)	Concepto físico, separado de los "recursos permanentes"	Cuestión de grado, capacidad de convertir en un bien de consumo (liquidez)	Desorientador, enfoque en el capital como complejo de bienes heterogéneos de diferentes "órdenes" y grados de especificidad	Preferencia temporal	Factor de producción separado (e.g., "efecto Ricardo")	"Estructura del capital" de Lachmann, liquidez de activos (el dinero a lo largo de un "continuo de liquidez")

Solow (1956), Swan (1956), Samuelson (1960)	Concepto físico (expresado en términos de dinero), orientada al pasado	Asume que no hay diferencias entre capital liquido (circulante) y fijo	Tasa de beneficio	Tasa de beneficio, productividad física	Factor de producción separado, con una unidad de medición diferente (horas de trabajo)	Ninguna
Robinson (1953), Sraffa (1960)	Concepto físico, orientada al pasado (inputs de trabajo)	-	Suma del trabajo anteriormente invertido (teoría del valor-trabajo)	Tasa de beneficio, superávit que recae sobre el capitalista a costo de los trabajadores	Factor de producción separado pero primario a todo	Crítica de las contradicciones neoclásicas del capital

Tabla I: Un resumen de los principales pensadores a lo largo del tiempo y sus aportes más importantes a la teoría del capital y a la teoría del interés

En conclusión, la teoría del capital que el presente trabajo propone es una síntesis de la teoría del capital de Carl Menger (1888), Ludwig von Mises (1949) y Irving Fisher (1930), complementado con la teoría del interés de Fisher (1930) y la teoría de '*preferred habitats*' de la curva de rendimientos de Modigliani and Sutch (1966).

Preface

Fools try to prove that they are right. Wise men try to find when they are wrong.

Many men have the 'courage of their opinions,' few the courage to abandon opinions.

The little man demands to be understood; the great man is content to be misunderstood.

Learn principles. Facts will then fall into their relations and connections.

Better capital in a man's head than capital in a bank.

~ Dickson G. Watts

These five principles by a seemingly obscure and little-known cotton speculator from the late 19th century have served me well. This piece of work was a prime example of abandoning opinions, finding out where my theory was wrong rather than right, separating principles from observations, and even acquiescing in being misunderstood from time to time.

I am forever indebted to my directors Prof. Dr. Jesús Huerta de Soto and Dr. Juan Ramón Rallo. Besides being of tremendous help during this work, they serve more than anything else as great examples. They can be considered my intellectual inspirations. Especially Prof. Dr. Huerta de Soto has left an important mark on my theories. His breadth of knowledge spanning many so different fields (epistemology, economics, law and even theology) and his passion in disseminating these ideas is truly unprecedented.

Likewise, Ludwig von Mises served as a great inspiration, albeit not directly. The successful publication of the Dutch version of his underappreciated but enormously valuable treatise, *Human Action*, already the second edition, was an initial stimulus to this final product. Other academics and practitioners that have influenced this work, either directly or indirectly, are Nassim Taleb, Mark Spitznagel, Jan Vis, Nicolas Cachanosky, Gabriel Calzada, Michael J. Osborne, Stephen Wright, Michael Sproul, Szabolcs Blazsek and countless others who I perhaps failed to mention. I am eternally grateful for their contributions.

I expect to revise the contents of this work considerably over time. Therefore, I see this work more as a starting point to a coherent treatment of capital than a finished product. Nevertheless, let this work be a monument of years of back-and-forth between theory and practice, incorporating what I have learned over the years, waiting to be revised repeatedly.

Biography

Olav Alexander Dirkmaat was born in Almere, the Netherlands on January 28th, 1988.

He received a Bachelor in Economics from the Hogeschool INHolland in Diemen-Zuid, the Netherlands. Because he wanted to pursue further academic work, he decided to enroll himself in a master's program of the Vrije Universiteit (VU) in Amsterdam, where he later obtained a Master in Marketing in 2011. However, he noticed that he was mostly interested in investing and financial markets, as a window to experience the ups and downs of millions of different businesses, rather than being exposed to and focus on one single business.

As an autodidact, he would as a twist of fate stumble upon the book *Economics in One Lesson* by economist and New York Times journalist Henry Hazlitt, which was an eye-opener to him. The Austrian school of economics was vastly different from the economics he learned at university, which mostly consisted of calculating price elasticities, drawing supply and demand curves and calculating the corresponding yet unobservable "consumer" or "producer" surpluses. As a result, he became obsessed with the work of Ludwig von Mises, Friedrich Hayek and others, such as Fritz Machlup and Huerta de Soto, to which he would dedicate countless hours in his spare time.

As he became increasingly interested in the 'Austrian' school of economics, he ended up writing his master's thesis on marketing and innovation in the context of the Austrian business cycle, titled *What Austrian Business Cycle Theory Teaches Us: The Role of Innovation and Advertising during Recessions*, under the supervision of Prof. Dr. Frambach. His thesis was later nominated best marketing thesis in the Netherlands. In addition, together with Dion Reijnders, he would translate Mises's *Human Action* into Dutch and set up the Pierson & Templeton Foundation, which would publish the first edition of the book. At the time of writing, the Dutch version of *Human Action* already made it to a second edition. The publication was applauded by Dutch politicians such as the classical liberal Frits Bolkestein and journalist/investor and best-selling author Willem Middelkoop.

As a result of his renewed interest in economics, he decided to move to Madrid, Spain to study economics under the wings of Prof. Dr. Jesús Huerta de Soto. Intellectually, this was an extremely fruitful period for him. He learned about fields he previously paid little attention to, such as epistemology and the history of economic thought. His main reason to move to

Madrid was, however, the Austrian theory of the business cycle. He became mostly interested in practical applications of this revered theory. This doctoral thesis is a result of that period in Madrid.

After successfully completing his master in (Austrian) economics from the Rey Juan Carlos University, he returned to the Netherlands and became part of a then recently founded start-up, GoldRepublic, which he helped grow into a million-dollar business. He was also closely involved with the founding and development of Nxchange, an equity and debt crowdfunding platform with secondary market, which was a spin-off of GoldRepublic. However, as he wanted to focus on his academic research and his personal ambitions in the world of investing, he decided to leave GoldRepublic and Nxchange in 2016 and move from Amsterdam to Guatemala City, where he, under the wings of his former professor Gabriel Calzada, became a part of the Business School of Universidad Francisco Marroquín, where he of today as continues to teach.

Table of Contents

List of Figures

Section I:
General Introduction

Chapter 1: Introduction

> "*Every important advance in economic theory during the last hundred years was a further step in the consistent application of subjectivism.*" (p. 31)
> **~ Friedrich Hayek** (1955)

1.1 Background: The Present-Day Importance and Impact of Capital Theory

The whole edifice of economic science is built on shaky foundations. Putting epistemological differences aside—the different housing constructions, colors of the walls, roof materials, etcetera, economists have made an art of ignoring that a lion share of their theories are grounded in soft clay. It is capital that moves the world, yet economists still know little about it. Economic science lingers in a dark veil, with misunderstandings across the board and little consensus.

The shaky foundations, of course, refer to capital theory. The edifice, perhaps, deserves some more elaborate explanation. In what way is capital theory relevant today? And how permeated are the unresolved questions on capital theory in applied economic science? In what way do modern economic forecasting models used in practice depend on capital? What is the importance of the agreements and disagreements on capital theory to financial markets, banking regulation, economic development (poverty), inequality and public policy?

In brief, we can observe three important ways in which erroneous theories of capital are leading (or have led) to catastrophic real-life implications. Two competing theories of capital seem to have their grip on our present-day world: (1) the "neoclassical" theory of capital and (2) the "neo-Ricardian" theory of capital. The consequences of both capital theories have been disastrous:

(1) The "neoclassical[2]" theory of capital (e.g., Solow, 1963) is mostly used in macroeconomic policymaking, economic growth forecasts and public finance projections (primarily fiscal deficits). Such growth models have been complete and utter failures. IMF, OECD and EC forecasts of budget deficits among G7 countries, for instance, typically suffer from forecast errors in the range of ±50% (Artis & Marcellino, 2001). The OECD overestimated economic growth in every area from 2007 to 2012, with average errors of one-year GDP growth projections of up to 280 basis points (Lewis & Pain, 2015). The lowest absolute forecast error equaled 170 basis points for OECD countries outside Europe. According to Turner (2016), the OECD failed to account for the 2008 crisis in their models, even in their estimates for the year *during* the recession. The OECD failed to predict declining or negative growth before the 2008 crisis for *any* of its major member-countries (the OECD uses a Cobb-Douglas production function to project "output" for each industry before aggregating industry outputs). As Mandelbrot (2004) describes the impossibility of forecasting human beings acting in the present while anticipating the future: "Anticipation is a feature unique to economics. It is (…) even harder to fathom than the paradoxes of quantum mechanics. Anticipation is the stuff of dreams and vapors." (p. 29).

While such prediction errors might seem innocent and harmless, important economic decisions are made on the basis of these pin-point precise growth estimates. They are used to justify economic policies, to "optimize" public spending (leading to over-optimization and procyclical public finances), to inform monetary policy (the output gap) and to feed (private and public) risk models. The production function formed the basis of the Solow-Swan growth model[3]. At its core, the neoclassical Solow-Swan growth model is an aggregate production function. The Solow-Swan growth model, or any given variation, is by far the most important model in present-day macroeconomics, whether we are trying to forecast or explain economic growth, business cycle fluctuations or the path of public finances (Kydland & Prescott, 1990). As the Cobb-Douglas production function was originally used to estimate *physical* output of physical *inputs*, it does not capture well non-material inputs and advances. The Cobb-Douglas production function, hence, will

[2] Even though Solow (1956) and others are strictly considered 'neo-Keynesians', commonly distinguished from 'monetarists', for the purposes of our work we consider both 'neoclassicals'.

[3] We will go to great lengths to criticize the Solow-Swan model in our literature review, see p. 182.

overestimate the contribution of labor to economic growth. More importantly, this overestimation bias will only grow worse over time, as non-material production becomes increasingly important, justifying *more* labor market intervention or even give rise to populist and Marxist movements[4].

(2) The byproduct of mainstream "neoclassical" capital theory in modern finance (e.g., Markowitz, 1952) has been a similar disaster. Most modern-day finance theory (derived one way or another from neoclassical capital theory) turned out to be a complete failure during the 2008 crisis, since it was unable to distinguish a sound, healthy and sustainable capital structure from an unsound unsustainable capital structure (and, likewise, sustainable economic growth from unsustainable growth). In fact, the neoclassical theory of capital is completely separated from money and finance (money is, thus, a mere "veil"). As a consequence, on one side we have a "neoclassical" theory of finance (which ignores the capital that underlies financial assets), while on the other side we have a "neoclassical" theory of economic growth (which completely ignores the financial side of the capital equation and is built on a naïve "money as a veil," which refers to the notion of the so-called "neutrality of money," quantity theory of money).

Perhaps one of the two universities that is most closely associated with the neoclassical point of view, the London School of Economics (the other is MIT in Cambridge, U.S.), was confronted after the 2008 crisis by Queen Elizabeth II[5] with the pertinent question why nobody had predicted the recession. LSE-professor Tim Besley then decided to send a three-page letter to the queen after convening a forum, apologetically stating that inflation was low and there were no signs of overheating. The imbalances in the capital structure were left unnoticed (and the subsequent losses on capital), partly due to the treatment of, what appears to be a parody to, *capital* in neoclassical models. In line with the failure to understand how the capital structure was fundamentally unsound, popular modern-day (neoclassical) finance theory also suffers many flaws. Especially the complete breakdown of financial and risk models should be noted. Popular theories and models, such as Markowitz' "Modern Portfolio Theory" (MPT), its later spin-off,

---

[4] In fact, the popularity of the neo-Ricardian capital theory (as evidenced by the positive reception of Piketty's *Capital in the 21st Century*) can be explained with the failure of neoclassical capital theory.

[5] From The Telegraph, Nov. 5, 2008, "*The Queen asks why no one saw the credit crunch coming*".

Sharpe's (and other's) "capital asset pricing model" (CAPM) and its ugly nephew, Garbade's and J.P. Morgan's "Value at Risk" (VaR) turned out defective and destructive ways of accounting for risk and the composition of capital[6]. These models largely lose their relevance when faced with a redefinition of capital theory.

(3) The "neo-Ricardian" theory of capital (e.g., Piketty, 2014) has been harmful, albeit in a different way, mostly because it has been used to justify (increasingly progressive) taxation, income redistribution and other welfare policies. In Venezuela, for instance, mining, oil and agricultural properties were confiscated from the hands of "foreign capital," as if such natural resources provide "monopoly rents" of their own[7]. Or, as another illustration, the nongovernmental organization Oxfam Novib presented a report titled *An economy for the 99 percent* at the Davos World Economic Forum 2017, which allegedly demonstrates an alarming trend of increasing wealth concentration among a handful owners of capital. Oxfam, as a result, calls for governments to end "the extreme concentration of wealth to end poverty."

What is colloquially called the "inequality crisis" is mostly backed by neo-Ricardian capital theory. The revival of Ricardian capital theory finds, perhaps, its origin at Cambridge, U.K., one of the sides in the Cambridge capital controversies (e.g., Robinson,

[6] Both Harry Markowitz and William Sharpe, alongside Merton Miller, received a Nobel Prize in Economic Sciences for their work on mostly neoclassical grounded (and here mentioned) finance and risk theories. Ken Garbade from Banker's Trust was one of the first to present a more elaborate Value-at-Risk (VaR) model based on Markowitz' work, although his work never truly circulated or gained prominence.

[7] In a strange twist of fate, Goldman Sachs bought $2.8 billion in bonds of Venezuela's state-owned oil company PDVSA (which partly comprises the confiscated and nationalized assets from ExxonMobil and ConocoPhillips in 2007) from the Venezuelan central bank in 2016, over which Venezuelan president Maduro ended up paying interest since it otherwise risks harming its virtually only source of revenue that is left after years of sabotaging free enterprise in the country. A PDVSA default would imply that Venezuela, in effect, is cut off from foreign capital markets. PDVSA accounts for almost all of Venezuela's foreign exchange inflows - about 93 percent, which are necessary to pay for imports. In other words, A PDVSA default would imply no dollars and no dollars imply no imports, which would lead to a social crisis of an immense scale. Such levels of poverty, starvation and lack of access to basic goods are unprecedented in absence of war.

> 1953; Sraffa, 1960). However, this very same neo-Ricardian theory of capital is based on fundamentally flawed economic thinking that must be addressed at its roots. A coherent theory of capital would demonstrate the futility of wealth and income redistribution, despite popular sentiment. A rekindling of the debate on the underpinning of the theory of capital could counter such destructive sentiment based on an erroneous Ricardian theory of capital, which only drives a wedge between wage-earners and profit-seekers.

In this light, the words once uttered by my mentor, Jesús Huerta de Soto, seem appropriate: "*Without capital theory, you will not understand anything that happens in an economy*." Of course, he is right. Yet with a *flawed* capital theory, you will not understand anything that happens in an economy all the while *you think you do understand* what is happening, which is in many ways even more pernicious and harmful. This is exactly what we can observe in the misapplication of neoclassical and neo-Ricardian capital theory to the real-world cases that we summed up above.

Quintus Cicero's "*men prefer a false promise to a flat refusal*" comes to mind, which is perhaps best captured by Taleb's (2015) "best map fallacy." According to Taleb (2015), the "best map fallacy" refers to the (unconditional) preference of a "false map to no map at all" (p. 38). "I know few people who would board a plane heading for La Guardia airport in New York City with a pilot who was using a map of Atlanta's airport "because there is nothing else." People with a functioning brain would rather drive, take the train, or stay home. Yet once they get involved in economics, they prefer professionally to use a wrong measure [the wrong theory of capital], on the ground that "we have nothing else." This idea, well accepted by grandmothers, that one should pick a destination for which one has a good map, not travel and then find "the best" map, is foreign to PhDs in social science." (Taleb, 2010, pp. 38-39).

In sum, a *wrong* capital theory is even worse than *no* capital theory, just as using the wrong map is worse than using no map. With no map, at least you have no pretentions of knowing the road, whereas having the wrong map might lead you to think that you actually do. To use another analogy, in any complex economy, looking at a market economy without an *accurate* theory of capital is akin to scuba diving in the Great Barrier Reef with your eyes closed. Much of the modern-day applications of fundamentally flawed capital theory are akin to using the wrong map.

Hence, capital theory is of utmost importance in understanding the world that surrounds us. Yet popular literature only confirms that “capital” is surrounded by mysticism. Most people have a vague association of factories and plants with the term “capital.” Others, associate it with money or someone’s wealth. Let alone any popular (often negative) associations with the terms “capitalism” or “capitalist.” Now, this mysticism would be rather innocent if it was strictly limited to a small subset of the population. It becomes potentially devastating, however, when even economic ‘experts’ that directly inform and sway top-down central planners *cannot* agree as to what capital is.

A clear definition of capital is therefore of great social importance. Many discussions evolve around “capital,” without a clear consensus on what capital is or what is meant by the term. As we will see, discussions back and forth between economists are fruitless without agreement to what capital is. Indeed, the famous *Cambridge Controversy*, which involved economists such as Sraffa (1960), Solow (1956), Samuelson (1966), Robinson (1953) and others, was never really resolved. The debates ended because most participants grew old and eventually passed away, not because the underlying disagreements were resolved. Indeed, there is a strong argument that "[w]hile many of the key Cambridge, England, combatants stopped asking questions because they died, the questions have not been resolved, only buried" (Cohen & Harcourt, 2003, pp. 212-213). Yet here we are, left with controversy and without answers. And our future depends on a solid understanding as to what capital truly means.

In a rather unfortunate turn of events, economic science has taken a wrong direction. Ever since the writings of the classical economists, economic science has persisted in treating capital as a physical stock of capital goods separate from labor or even land. Capital is, either implicitly or explicitly, seen as something “material” and “physical.” Despite numerous advances in demonstrating the subjectivist nature of economics, capital theory has remained behind. The consequences of this “wrong path” is still haunting economic science today. “Physical” theories of production are still the dominant paradigm. Production functions view capital as just another factor of production. An unsatisfactory starting point then leads to troublesome and sometimes insolvable contradictions.

Or as Mandelbrot (2004) puts it: “[P]erhaps (…) economics is about not just the physics of wheat, weather, and crop yields, but also the mercurial moods and unmeasurable anticipations of wheat farmers, traders, bakers, and consumers.” (p. 41). Yet modern-day applications of neoclassical and neo-Ricardian capital theory emphasize the physical side of the equation. Yet, economics is not about physical relationships between different sets of inputs yielding some output. Economics is about the relations between individual human beings, their preferences, their valuations, and their decisions.

1.2 Research Aim

Many modern misconceptions can be traced back to this tragic mistake. The classical legacy of capital versus labor (and land) led to devastating consequences. To begin this work with a rather bold hypothesis: is it possible that 80% of today's economic debates were unnecessary if the most important debate of all, that is, the debate on capital, would have been dealt with adequately? Bliss (1975) seems to agree, despite his skepticism on any consensus on capital. Striking an ironical note, he writes: "When economists reach agreement on the theory of capital they will shortly reach agreement on everything else." (p. vii)

It should be clear that, as even one of its protagonists Robert Solow (1963) would admit, the debate on capital has never been satisfactorily solved. Many present-day economists simply do not care. They ignore this unresolved theoretical divide and continue with business as usual. But the debate is bound to become front and center in economic science once more.

After numerous controversies throughout the 19th and 20th century, this is an attempt to revive the debate on capital theory and to provide a satisfactorily solution. This thesis is an attempt to "right the wrongs" and put economic science on a solid footing.

Before proposing a solution to the issues that encumbered capital over the past centuries, I had to become, as Solow (1963) would call it, a "learned reader." The history of capital theory is as a dark path through the treacherous Andes peaks and valleys. Navigating it requires incredible diligence. It is easy to lose oversight whenever crossing one of its valleys.

In sharing my explorations through the historical debates on capital, I saw two options to document my journey: either by distinguishing the various economic schools of thought, or by chronically portraying the most important debates so far. I opted for the latter. Even authors within the same economic school of thought hold highly opposing views on the nature of capital. In some cases, it even seemed some had more in common with fellow economists from other schools of thought then with their own "breed," at least when it comes to capital (take Fisher (1930) and Menger (1888) as an example). Hence, we must be careful to "group" economists and their ideas according to their supposed affiliation. I still summarize the key differences in how capital is seen, even though the preceding work is structured chronologically.

There exists substantial disagreement on *what* a sound theory of capital should be able to achieve. Bliss (2005) argues, for instance, that: "[c]apital can mean specific capital goods (as in the case of a piece of machinery); it can mean the finance that a particular project requires (as when we refer to the *capital market*) (…). A fully successful theory of capital should master both cases, because the market system allocates capital in both senses. It determines how and where a specific machine tool is used; and it decides how the financial resources directed to current investment will be translated into the purchase of specific machines and structures of what particular designs" (p. xi). Without getting into too much detail at this point, his idea of what the scope of capital theory ought to be can easily be refuted. If, *ceteris paribus*, a specific machine is used in a way or in a place where it is able to yield a higher profit, it will be reflected in the value of capital. Moreover, capital investment is an integral part of a satisfactory theory of capital, since assuming otherwise would lead to a capital theory that lacks an explanation of its origin. What is not up to economic science, however, are indications as to how (or where) a specific machine should be used to optimize its yield. Economics can only provide insight in where economic value, and thus capital, is derived from, but cannot provide insight in the specific cases of capital and individual capital goods.

In summary, the aim of this work is to reach a coherent definition of what capital *is* and what it *is not*. Moreover, it is to understand and study the *determinants* of capital and its structure, yet in an economic sense rather than a physical or material sense. Doing so, it is inevitable to study the inner workings of *capital markets*, which is the key determinant of capital and its structure. It is where capital is allocated and thus defines its composition at any given time[8].

If it is true that we cannot rely on a physical or material analysis of capital or a "capital stock", we are forced to single out value drivers within the economy. What brings about certain tendencies in a capital structure? The aim is to discover regularities and logical relations between two variables that form a common threat throughout the structure of

[8] The term *capital markets*, as used in this thesis, is understood as the financial markets (which thus comprise any financial intermediary including banks) that channel savings and investment between *suppliers* (or *providers*) of capital (i.e., savings) and *users* of capital.

capital, which are practically useful for practitioners on capital markets. We call these regularities "*q*", which is the subject of section IV.

Hence, the primary objective of this work is straightforward: to provide a solid foundation for economic science, specifically by incorporating a truly *subjectivist* theory of capital. Nonetheless, before proposing an alternative vision to the present-day mainstream, we will have to expose the errors of the "old economists" with regard to capital theory and, inevitably, interest theory. The purpose of this thesis is therefore multifaceted:

(1) To review the three most important historical debates on capital theory from an 'Austrian' or 'Misesian' perspective.
(2) To do a comparative analysis of the primary theories of interest.
(3) To extend the Mengerian and Misesian theory of capital, grounded in subjectivism and modern finance, to include (a) the theory of the (financial) entrepreneur, (b) the theory of the business cycle, (c) the theory of the term structure, and (d) exchange rate theory. In other words, to revive the forgotten theory of capital by Menger (1888), use Menger's theory as foundation, and analyze its broadest of implications in the context of modern financial markets.
(4) To formulate a coherent theory of consumption, derived from the Mengerian and Misesian theory of capital, grounded in subjectivism and in the idea of liquidation and/or consumption of "services." This in contrast to the commonly observed custom of mistaking "consumption" for "obtaining ownership."
(5) In line with the previous research objective: to formulate a theory of human capital, contrary to the idea of labor as "factor of production" completely separated from capital, which suggests capitalizing future (expected) wages of an individual as part of his own "personal balance sheet." It is thought that by developing and formulating such a subjectivist theory of human capital, the theory of capital can ultimately be better understood as something broader than what is usually considered: human capital becomes an *outcome* of the economic calculations by an individual.
(6) To prove that the 'Austrian' theory of capital was not "refuted" in the controversy on capital in the 1950s and 1960s: specifically the phenomenon of reswitching was, according to one of its participants, the Nobel laureate Samuelson (1966), clear proof that the 'Austrian' theory of capital was fundamentally flawed. Thus, to attempt to apply the term structure of interest rates (instead of one single interest rate) to the

specific examples that were used over the course of these historically important and controversial debates on capital.

(7) To understand our theory of capital in terms of equilibrium and disequilibrium. More specifically: distill the principle of *q* (which is derived from Tobin's and Brainard's [1969] work) from the theory of capital and study the relationship between *q* as a method of valuation and futured expected returns in the context of various asset classes (such as equities, real estate, et cetera).

(8) To combine the term structure of interest rates with the principle of *q*. As such, to attempt to estimate the probability of a possible (large) drawdown in equities or a recession in a more effective manner than in the existent literature (e.g., Wright, 2004; Spitznagel, 2011).

(9) To complement and rephrase the 'Austrian' theory of the business cycle, assisted by our capital theory, according to the 'preferred habitats' theory of the term structure and the phenomenon of maturity mismatching by the different financial intermediaries.

(10) To review various important historical episodes of economic downturns by applying the reformulated theory of the business cycle as proposed in our previous research aim, providing a different perspective to understand and analyze the phenomenon of the business cycle, based on the practice of maturity mismatching (asset-liability mismatching) by the different financial intermediaries.

(11) To present a sound reasoning to explain how the quality of a central bank's assets (its "backing," that is, the replacement value of its assets), ultimately determines the returns on central bank liabilities to the degree that the value of the underlying assets diverge from their market value, applying the same *q* principle. All other things equal, after a fall in the market value of the assets of a central bank, we would expect negative future returns to the degree that the exchange rate depreciates or domestic inflation increases.

(12) To open up a new field of investigation related to the intricate interrelations between capital, money, the various other financial assets and time (maturity/*duration*).

1.3 Contributions

The author does not pretend to add completely new or novel ideas to existing theory. In fact, this work is better to be considered a grand synthesis of a plethora of theories. The strength, therefore, of this work, is to integrate various theories from different academic fields, such as economics, finance and banking, to build a coherent theoretical structure which allows someone to understand the real world. I attempt to make various theoretical contributions:

(1) To revive the time preference theory of interest in a modern finance context and integrate the term structure of interest rates into capital theory. Few authors have attempted to integrate the yield curve into a more comprehensive explanation of interest and its consequences for the structure of capital.

(2) To summarize, outline and document the various contributions made to capital theory over time, as well as revisit the roots of modern-day misconceptions with regard to capital theory.

(3) To provide a coherent, truly subjectivist theory of capital without falling in the pitfalls of classical economists.

(4) To reconcile the views of a truly subjectivist theory of capital with a coherent theory of banking that is not separate from, but rather integral to the theory of capital.

(5) To resolve the problem of *reswitching* which, allegedly, refuted 'Austrian' capital theory.

(6) To resolve the controversy on diminishing returns on capital by incorporating the 'Austrian' theory of entrepreneurship (e.g., Kirzner, 1960; Huerta de Soto, 2010) and to explain the apparent stability of long-term rates and the continued existence of profits with our theory of Fisher's "pendulum of returns."

(7) To criticize and refute the neo-Ricardian theory of capital that underlies much of present-day policies geared toward "wealth inequality."

(8) To criticize and refute the neoclassical theory of capital and interest that underpins much of today's growth models (e.g., Solow, 1956).

(9) For the first time ever, to provide a satisfactory and comprehensive theoretical explanation of the yield curve and recessions. While there exist papers that focus on maturity mismatching as a source of financial instability, all these models always require an exogenous factor to trigger a crisis. I attempt to depict an endogenous theory of the business cycle based on maturity mismatching, by integrating (1) the theory of the portfolio demand for money, (2) the theory of time preference, (3) the theory of Kirzner and Huerta de Soto of the (financial) entrepreneur, (4) the "preferred habitat" term structure hypothesis, (5) the law of reflux, (6) the "Austrian" theory of malinvestment (e.g., Huerta de Soto, 2006), (7) the theory of *q*, and (8) the theory of capital, including a notion of "scales of capital", and arbitrage between the various "scales."

(10) To empirically illustrate that the principle of *q*, derived from our theory of capital, can predict future expected returns.

(11) To offer an innovative, modern explanation of the Great Depression of the 1930s, the S&L crisis of the 1980s, the Asian financial crisis of the 1990s, and the Great Recession of 2008.

1.4 Methodology

The methodology of this present work can be broken down into the following steps:

(1) First, we will conduct an extensive literature review about the theory of capital. It is inevitable, in this context, to avoid related topics and theories of interest, for instance, the theory of interest, the theory of the term structure, theory of banking and financial intermediation, and the theory of risk, among others. We will divide the literature into three chapters according to a chronological order: first, with the first capital controversy from 1888 to 1907 between Böhm-Bawerk and J.B. Clark, besides the important contributions by Menger and others. Second, with the second capital controversy of the 1920s and beyond: this period includes the debates between Fisher and Hayek, Knight and Hayek, Knight and Kaldor, and Hayek and Sraffa. We also discuss the important contributions of Mises, Williams, Macaulay, Lutz and, ultimately, Lachmann. This latter

group of authors did not participate in the second round of debates directly, but has made extremely valuable contributions to the theory of capital and interest during this period. Third, with the third (and best-known) capital controversy of the 1950s and 1960s between, which are commonly referred to as the Cambridge Controversies (which alludes to the fact the universities involved were located in Cambridge, U.S. [MIT] and Cambridge, U.K.). This controversy included, as key participants, Solow (1956), Swan (1956) and Samuelson (1958) from the Cambridge, U.S. side and Robinson (1953), Sraffa (1960) and others from the Cambridge, U.K. side. In this last round of debates, the discussion about the theory of capital became increasingly ideological: nevertheless, the classic liberals (in the European sense) should think twice before siding with Cambridge, U.S. and accepting a completely flawed capital theory: as Piketty (2014) proves, a bad justification of markets only leads us down the wrong path of populism. The idea behind this review of existent literature is to find important contributions toward our goal of (re)formulating a theory of capital consistent with economics as a subjectivist science (e.g., Hayek, 1955).

(2) Second, find and analyze the interrelations between the different elements that, at first sight, seem unrelated or isolated. We intend to connect the varying theories of capital and interest to the theories of the (arbitraging) entrepreneur, the theory of the term structure, the theory of (Macaulay) *duration*, the (financial) structure of capital, to the theory of Tobin's q, to the portfolio theory of money, and to the theory of maturity mismatching. In this sense, this work proposes to synthesize this large variety of both direct and indirect historical contributions to the theory of capital.

(3) Third, propose a theory of capital grounded in subjectivism, which avoids past mistakes. Discuss its implications in relation to various related phenomena, such as consumption, savings, interest, liquidity and money.

(4) Fourth, propose a general theory of *q*, which – as an operationalization of the theory of capital – can be applied to wide array of different sectors and types of assets. This theory combines the theory of the entrepreneur of Huerta de Soto (2010) and Kirzner (1960) with the theory of capital of Menger (1888) and which, ultimately, becomes a robust method of valuation predictive of future returns.

(5) Five, to use empirically the same theory of *q* to concrete applications: first, to the stock market (U.S.) combined with business cycle theory; second, to the (residential) real estate market with an empirical study that spans over 60 years and 12 different countries; third, to the foreign exchange market (currencies), which includes an empirical study including 15 Latin-American countries and their respective central banks; fourth, to the physical gold market (as real nonfinancial monetary asset), in which we briefly highlight the application of the principle of *q* to tangible assets such as physical gold, paving the way for future applications to similar assets such as cryptocurrencies. These empirical studies illustrate our work on our proposed theory of capital. They show various real-world applications of our theory of capital. We use different statistical techniques for our estimations, mostly panel data models.

(6) Six, to reformulate the theory of the business cycle (e.g., Huerta de Soto, 2006), emphasizing the role of yield curve spread and the widespread practice of maturity (or *duration*) mismatching. The theory is then applied to four different historical episodes of crisis and recession to show their real causes: we will discuss the Great Depression of the 1930s, the S&L Crisis of the 1980s, the Asian financial crisis of the late 1990s, and the Great Recession of 2008.

In this manner, this work attempts to conduct a solid understanding of the different theories of capital while their practical value and implications are at the same time also considered. This allows us to build a sound capital theory from the ground up.

1.5 Outline

This work begins with an extensive literature review of capital and any auxiliary concept that is deemed important by the author (Section II). This literature review is, as mentioned above, a chronological account of the most important historical debates on capital theory. Three rounds (clusters) of debates were identified in approximately the following periods:

– The first round from 1888 to 1907, mainly between Böhm-Bawerk and Clark, but we also look at the contributions of Menger, Wicksell, Fisher, Mises and Machlup. The later debates on shiftability (which were aimed at, as we will see, capital theory but from a finance and banking perspective) are also discussed, which is part of a novel approach to understanding capital theory.

- The second round from 1925 to 1942, mainly between Hayek and Knight, but also with important roles for Kaldor, Sraffa, Mises, Hicks and Fisher. However, many auxiliar contributions were identified: Macaulay's work on duration is of utmost importance for capital theory, as well as Williams's (1938) work on investment and Culberton's (1957) work on the term structure of interest rates.
- The third round from 1953 to 1969, famously known as the Cambridge controversies. Hayek did not get involved, yet Kaldor and Sraffa took the side of Cambridge, U.K., together with Joan Robinson and Pasinetti, who heavily criticized the neoclassical theory of capital, represented mostly by Solow, Swan and Samuelson. Both the absence of Hayek (and other 'Austrian' thinkers) and the alleged issue of reswitching dealt a deathblow to the Austrian School of Economics, which was largely considered obsolete after the 1940s and even refuted after the 1960s. Paradoxically, many of the "Austrian" insights (represented by Menger, Mises, Machlup and to a lesser degree Hayek) are needed to resolve the important issues of this third round of debates on capital theory. In addition, over the same period, various contributions were made with regard to the term structure, which are also discussed in this chapter.
- The fourth round from 1970 to the recent past, which technically cannot be seen as a round of debates. In essence, this period should be characterized as a complete lack of controversy on capital (and perhaps even deliberate avoidance from capital theory), yet interesting auxiliary developments in related areas: principally, portfolio and asset pricing theory (finance), capital budgeting (finance), commercial and central banking, asset-liability mismatching (in particular maturity and duration mismatching), (re)hypothecation and (reuse of) financial collateral (few theoretical advances but many new practices arising in the real world), developmental economics, and wealth inequality.

We will then conclude the literature review in a final chapter of Section II that summarizes all of the important tendencies in varying theories of capital, interest, money (in its narrow context to capital), financial intermediation, highlighting the (historical) contributions made by each author.

In the remainder of this treatise, we will develop our line of thought in the following sequence:

- In Section III, we will propose our own theory of capital that should be viewed as a grand synthesis of earlier work on capital and interest. We begin in Chapter 7 by proposing a comprehensive theory of capital. In Chapter 8, we will propose and summarize a "portfolio approach" to the demand for money and financial assets. Then, in Chapter 9, we will outline the well-developed theory of entrepreneurship (e.g., Huerta de Soto, 2010) and extend it by including the *financial* entrepreneur or *capitalist*-entrepreneur. In Chapter 10, we will review pressing issues on the return of capital and present a solution to previously unresolved issues. Last, in Chapter 11, we will show how the value of capital fluctuates and changes, very different from what other economists suggest.

- In Section IV, we will start in Chapter 12 by summarizing the recent breakthrough in the equity q ratio, before extending the theory by incorporating the term structure of interest rates. By including in our measure the yield curve spread, we are better able to predict stock market returns. In Chapter 13, we apply the same principle to residential real estate markets with a longitudinal study on housing q and housing returns. In Chapter 14, we repeat the same methodology, but this time apply our q principle to foreign exchange rates, by analyzing and examining central bank losses and future foreign exchange returns (estimated by the rate of inflation and the exchange rate). In Chapter 15, we propose a preliminary approach to using the q principle to the physical gold market. We provide some anecdotal prove that the principle of q also functions in monetary nonfinancial asset markets.

- In Section V, we will propose a reformulation and extension of existing 'Austrian' business cycle theory in Chapter 16, which was already impressively advanced by Huerta de Soto (2006), by incorporating some of the theories that were discussed in our work. In Chapter 17, we will review various historical recessionary episodes equipped with our theory. Specifically, we look at the Great Depression of the 1930s, the S&L crisis of the 1980s, the Asian financial crisis of the late 1990s, and the Great Recession of 2008. Many of the observations coincide with our theory of the business cycle.

- In Section VI, which spans Chapters 22, 23 and 24, we will conclude our work by summarizing our main conclusions, implications and suggestions for future research.

Section II:
A Historical Account of the Controversies on Capital:
A Critique of Other Theories of Capital and Interest

Chapter 2: The First Round (1888 – 1907): Böhm-Bawerk versus Clark

The history of capital abounds with head-on collisions between intellectual giants. It was at the close of the 19th century that two of those giants reached a crescendo of scholarly activity. At the time, they were not even aware that they were working toward the pinnacle of their careers, simultaneously, without any consciousness of each other's existence, akin to the battle between Thomas Edison and Joseph Swan to reap eternal fame for inventing the light bulb. Eventually they found out about each other's work, only after one of both discovered that he was not the only economist working on unraveling the secret behind this earthshattering thing called "capital." They exchanged letters. And more letters. And then, in all friendliness, they went after each other in public. It was at the start of the 20th century that their debates were getting fierce, even though they never ceased to be full of praise of one another.

These early intellectual giants colliding over capital were Englishman John Bates Clark and the Austrian economist Eugen von Böhm-Bawerk, later joined by economists such as Knut Wicksell and Carl Menger. As a curious fact, this debate would have never happened if Clark and Wicksell did not possess an excellent command of the German language. Hence, their understanding of the German writings of Böhm-Bawerk led to the first explicit controversy on capital theory.

We will begin by summarizing the contributions made prior to this first round of capital controversy, mainly by referring to important classical economists. Then we will summarize the debate between Clark and Böhm-Bawerk, highlighting some key contributions made by other economists, such as Wicksell, Fisher, and, principally, Carl Menger.

2.1 Prior to the First Round of Capital Controversies: The Classical Legacy

2.1.1 Adam Smith on Fixed versus Circulating Capital and Interest as "Rent" of Capital

Adam Smith was largely responsible for the separation of capital, labor and land, which each having their own "income categories," in this case interest (on capital), wages (on labor) and rent (on land). As Smith (1776) writes himself: "[T]he price of the greater part of commodities resolves itself into three parts, of which one pays the wages of the labour, another the profits of the stock, and a third the rent of the land which had been employed in producing and bringing them to market." (p. 286).

This separation still haunts economics today. Menger (1888) was very critical of what he calls the *Smithian doctrine*: "[W]ho would claim that a naturally grown tree used to build a ship is not capital while a purposefully planted, equally constituted tree used for the same purpose is? That natural mineral water is not capital, but that humanly refined water is?" (p. 11). Menger (1888) laid bare the fallacies behind Smith's classical notion of capital rather quickly. We could call Smith's view on what constitutes capital and what not, a "backward-looking" view. It takes into the account the historical origin of an economic good, not its ability to increase future well-being. Menger (1888) concludes: "It is an utterly indefensible position to claim that all other goods, including sheer objects of nature, may become products and capital once labor has been used on them as long as they are devoted to further production while immovable objects of nature and human labor power may not." (p. 13).

Moreover, Smith (1776) was responsible for introducing the dichotomy between fixed and circulating capital. Smith (1776) argues that: "Every fixed capital is both originally derived from, and requires to be continually supported by a circulating capital." (p. 283) In fact, fixed capital requires circulating capital in three ways:

(1) To create fixed capital;
(2) To maintain fixed capital;
(3) As input to fixed capital (Smith has in mind the idea that raw materials are processed by a machine and the machine, as fixed capital, that thus require inputs, that is, raw materials).

Smith (1776) begins classifying several economic agents according to how much circulating or fixed capital they use. The merchant, to Smith (1776), exclusively uses circulating capital. On the contrary, a manufacturer depends heavily upon fixed capital, etcetera.

Not only is circulating capital according to Smith necessary to sustain fixed capital. Circulating capital (more "liquid" assets) is also necessary to support the labor force over a given period of production. This latter idea would later influence especially Böhm-Bawerk's work when he tried to explain the *sources* of time preference and deviated to capital as a "labor fund." Nevertheless, Smith (1776) himself falls into some contradictions. First, Smith seems to define capital as capital goods. Then he says that a lender views his loan to some borrower also as "capital" and that the borrower, in turn, invests the loan in "capital."

Adam Smith recognizes, in line with Menger (1888), the relation between money and capital. Indeed, contrary to many other economists, Smith (1776) at a very early point in Book II (specifically Chapter 2), aptly titled *Of the Nature, Accumulation, and Employment of [Capital] Stock*, relates capital to banking operations. Laidler (1980) summarizes Smith's contribution in the following way:

> "In short, when gold is displaced by paper money it can (...) make a permanent contribution to the stock of circulating capital. What we have here then is not an analysis that confuses money and capital, (...) but a clear and correct account of the social gains to be had from replacing commodity money with paper." (p. 11)

Laidler refers here to the idea that people can hold various instruments as money. Specie, or gold in this case, is a very inefficient transmitter of saving. Replacing it by paper money, however, is more efficient. Any person willing to hold paper money (and, thus, not spends it), allows a bank to issue an equivalent of paper money to a borrower. Since the likelihood of the person holding the paper money spending it (and causing reflux) is high, a bank would not risk investing in fixed capital, but is willing to invest it in circulating capital. As such, any monetary savings not by hoarding gold but rather by hoarding liabilities of financial intermediaries, allows for a "contribution to the stock of circulating capital". This contribution is, however, not permanent. It is contingent upon the act of saving. As soon as the holder of the paper money decides to spend rather than save, it will compel the bank to withdraw from the stock of circulating capital.

As Adam Smith (1776) himself puts it:

> "**It is not by augmenting the capital of the country, but by rendering a greater part of that capital active and productive than would otherwise be so, that the most judicious operations of banking can increase the industry of the country**.

> That part of his capital which a dealer is obliged to keep by him unemployed, and in ready money, for answering occasional demands, is so much dead stock, which, so long as it remains in this situation, produces nothing either to him or to his country. **The judicious operations of banking enable him to convert this dead stock into active and productive stock**; into materials to work upon, into tools to work with, and into provisions and subsistence to work for; into stock which produces something both to himself and to his country. The gold and silver money which circulates in any country (…) is in the same manner as the ready money of the dealer, all dead stock. (…) **The judicious operations of banking, by substituting paper in the room of a great part of this gold and silver, enables the country to convert a great part of this dead stock into active and productive stock; into stock which produces something to the country**." (p. 247)

In other words, Smith (1776) shows convincingly that savings are transmitted into investment, and that a smoothly running banking system is much more efficient in transmitting savings *directly* into investment than *indirectly* through, what Smith (1776) calls, "dead stock."

Moreover, Smith (1776) *did* recognize the limitations that financial intermediaries run into when issuing deposits or paper money whilst lending freely, even though a complete formulation of the law of reflux was only developed later[9]:

> "**A banking company, which issues more paper than can be employed in the circulation of the country, and of which the excess is continually returning upon them for payment**, ought to increase the quantity of gold and silver, which they keep at all times in their coffers, not only in proportion to this excessive increase of their circulation, but in a much greater proportion; their notes returning upon them much faster than in proportion to the excess of their quantity." (p. 301) [emphasis mine]

Additionally, Smith (1776) recognized the importance of maturity matching of individual Scottish banks. To lend at longer maturities than to borrow would imply that: "the whole of the returns is too distant from the whole of the outgoings, and the sum of repayments could not equal the sum of its advances within such moderate periods as suit the conveniency of a

[9] We will address the law of reflux at a later point, beginning on p. 268

bank." (p. 307). A banking system is thus understood to be "liquid" as long as the maturities of assets do not exceed the maturities of bank liabilities (principally, in Smith's case, paper bills).

According to Smith (1776), advances made on real bills by banks *tended* to equal corresponding savings (in banks' cash balances). As Glasner (1992) argues: "[T]he sum advanced by a bank to customers borrowing on the security of real bills is unlikely to exceed the amount of "ready money" that the customer would have already been holding." Such "ready money" holdings would thus be holdings of bank deposits or bills (i.e., the "demand for bank money"), offsetting an increase in the "supply of bank money." Hence, any bank lending on real bills would tend to be backed by savings. Visualizing Smith's point with the discounting of a 30-day real bill drawn against highly demanded consumer goods:

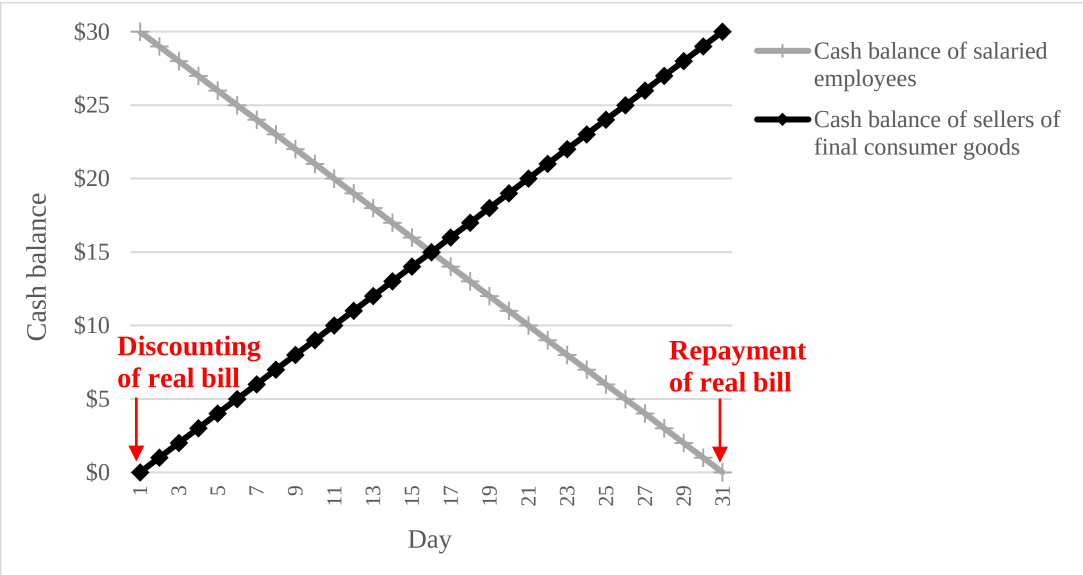


Figure 1: A different interpretation of Smith's real bills doctrine: consumers tend to have short-term savings that are aligned with a gradual liquidation of such savings and thus the discounting of the underlying real bill. In fact, any commercial credit would do, not just real bills.

Thus, one of the great merits of Adam Smith (1776) was to recognize and describe the intimate connection between capital theory and banking theory (or, in present-day economics, a broader theory of financial intermediation). Yet one of the great flaws of Smith was to distinguish between capital (interest), labor (wages) and land (rent), while turning a blind eye to the role of time and time preferences in his analysis.

2.1.2 Ricardo’s Capital Sins

David Ricardo (1817), perhaps like no other, was responsible for introducing the to many modern economists apparently holy trinity of land, labor and capital. In fact, as we will see later[10], Ricardo was largely responsible for the persistent myth that the “rent” on land would increase as the population grows, which justifies his proposed “land tax.” Moreover, the “cost theory” of prices was a product of Ricardian thought, which made its comeback in the third controversy on capital theory (e.g., Robinson, 1953; Sraffa, 1960), almost one hundred and fifty years after Ricardo’s work. In fact, the Italian economist Piero Sraffa (1960) is generally “credited” with the revival of Ricardian thought and the onset of the “neo-Ricardian” school. Later Marxian thought was also largely influenced by Ricardian capital theory.

With our present-day knowledge, it is a relatively straightforward task to identify the fallacies implied in Ricardo’s work. Stigler (1952), for instance, was remorseless: “Economics is the body of substantive generalizations on the workings of economic systems. **Ricardo did not enlarge much this body of knowledge**: his one addition to Smith's work was the systematic, though only partial, recognition of diminishing returns.” (p. 206) [emphasis mine]. Far from advancing economics, Ricardo misled economic thought.

If we limit ourselves at this point to an analysis of Ricardian capital theory, a mention of Ricardo’s theory of profits is inevitable. Yet as Edelberg (1933) argues, Ricardo’s theory of profits is “obviously a mere confusion” (p. 51). Ricardo (1817) was principally concerned with the distribution of income in society. To develop his theory of profits, he first developed a theory on the nature of capital. To Ricardo, capital or a capital good was a mere embodiment of past labor. Ships, buildings and machines were products of previous labor. Therefore, all capital goods can be reduced to past labor input[11] (Edelberg, 1933). Ricardo viewed capitalistic production as an “indirect method of applying labour to the production of consumption goods” (Edelberg, 1933, p. 52). Moreover, the distinction between fixed and circulating capital, according to Ricardo (1817), was arbitrary and a mere question of degree. To Ricardo (1817), capital would thus encompass all the goods used in production, which includes the sustenance of laborers (a crude notion of the later popularized view of capital as

[10] Piketty (2014) is a prime example of neo-Ricardian capital theory, see p. 304.

[11] We will see later how the Ricardian theory of capital has been revived in the Cambridge controversies on capital by Piero Sraffa, see p. 221.

“subsistence fund”). Simply put, Ricardo argues that the value of a consumption good depends on the labor spent on creating such good. In the case of capital goods, the total labor spent on creating the consumption good would merely include the labor spent on creating the capital good. Capital is “accumulated labor.”

Nevertheless, Ricardo (1817) also adds a “time” component. If a capital good is accumulated labor, then that capital good can only be recreated by expending the same amount of labor again, which involves time. The difference between a present and future capital good is the time needed to create it by means of labor. Hence, the value of a consumption goods ultimately depends on labor expended and the period of production (that is, time). In Ricardo’s (1895) words:

> “After the best consideration which I can give to the subject, I think that there are two causes which occasion variations in the relative value of commodities; first, the relative quantity of labour required to produce them; second, the relative time that must elapse before the results of such labour can be brought to the market. All questions of fixed capital come under the second rule, which I will endeavour to explain to you if you should wish it.” (p. 65)

According to Ricardo (1817), capital is an embodiment of past labor inputs and profits are therefore:

> “[I]n all countries and at all times, profits depend on the quantity of labour requisite to provide necessaries for the labourers on the land or with that capital which yields no rent.” (p. 70)

To understand Ricardo’s statement, we should notice that according to his analysis a natural tendency toward *lower* profits and *higher* wages exists: “The natural tendency of profits is to fall; for (…) the additional quantity of food required, is obtained by the sacrifice of more and more labour.” (Ricardo, 1817, p. 133).

To paraphrase Ricardo (1817), he essentially thought that land has diminishing returns (as we use marginal land to grow corn and feed a growing population).

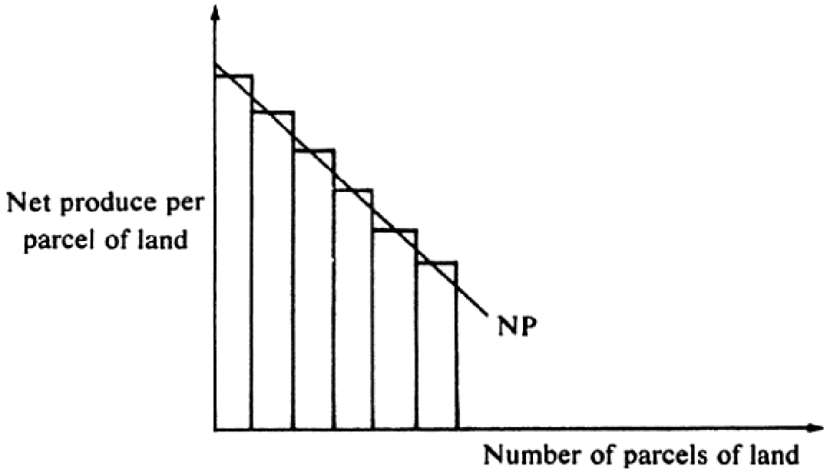


Figure 2: A key part of Ricardo's (1817) capital theory is the idea that food production suffers from diminishing marginal returns as increasingly less fertile parcels of land are used.

In sum, "(…) profits depend on high or low wages, wages on the price of necessaries, and the price of necessaries chiefly on the price of food (…)" (Ricardo, 1817, p. 119). This implies that surplus rents simply exist as yield to capital (similar to the way an apple tree by its very nature yields apples) and is divided between workers (who earn *wages*), entrepreneurs (farmers, in Ricardo's terms, who earn *profits*) and owners (or landlords, as Ricardo would call them, who earn *rents*).

In Ricardo's (1817) words: "The whole value of their commodities is divided into two portions only: one constitutes the profits of stock, the other the wages of labour. Supposing corn and manufactured goods always to sell at the same price, profits would be high or low in proportion as wages were low or high." (p. 117). However, the bounty (profits) left to farmers (entrepreneurs) diminishes and disappears over time, since an increase in population can only be served by expanding food production to land with marginal diminishing returns, only to be postponed by productivity gains through better machinery and/or greater agricultural knowledge. In Ricardo's (1817) words: "The natural tendency of profits then is to fall; for in the progress of society and wealth, the additional quantity of food required is obtained by the sacrifice of more and more labour. This tendency, this gravitation as it were of profits, is happily checked at repeated intervals by the improvements in machinery, connected with the production of necessaries, as well as by discoveries in the science of agriculture which enable us to relinquish a portion of labour before required, and therefore to lower the price of the prime necessary of the labourer." (p. 121).

The point that Ricardo (1817) drives home, however, is the notion that both farmers (entrepreneurs) earn less and less and workers earn less and less in real terms. Profits (which accrue to entrepreneurs) completely disappear; wages (which accrue to laborers) fall to their subsistence level. It is only landlords (land owners) of marginally superior (more productive) land who accrue the benefits of population growth and increased production, as their relative share of total produce increases. As the population grows and marginal parcels of land are being introduced (with diminishing returns on labor input), residual profits go down, as the rate of profit is arbitraged between farmers (and thus between parcels of land) until a uniform rate of profit is achieved. The differential between the profit earned by farmers of more productive land and the profit earned by farmers of the least productive land accrues to the landowner in the form of rent. Hence, profits have a tendency to fall and disappear; wages will fall to a subsistence level; and only privileged landowners will reap net gains. Only temporary "productivity shocks" due to innovation could postpone this dire scenario. Of course, since such gains by land owners of marginally superior plots of land are permanent, they will also accumulate excessive "wealth," although their share of "wealth" is bound by the subsistence level of workers[12].

Interest in Ricardo's mind effectively equated to profits: "**The rate of interest, though ultimately and permanently governed by the rate of profit**, is however subject to temporary variations from other causes." (Ricardo, 1817, p. 349) [emphasis mine]. The "other causes" Ricardo refers to consist mainly of changes in the supply and value of money. This is basically Cantillon's (1755) notion of the non-neutrality of money. As Ricardo (1817) expresses it: "If by the discovery of a new mine, by the abuses of banking, or by any other cause, the quantity of money be greatly increased, its ultimate effect is to raise the prices of commodities in proportion to the increased quantity of money; **but there is probably always an interval, during which some effect is produced on the rate of interest**." (p. 350) [emphasis mine].

This pessimistic and naïve view of capital (or by singling out land and labor as some special subsets of capital) led, first, to a foundation of Marxian thought on the inevitably of capitalism's demise and, second, to a revival of interventionist thought in the late 20th

[12] We will later see that his Ricardian idea forms the basis of Piketty's $r > g$ (that is, the return on capital exceeds the rate of growth) and served as a political justification of wealth-destroying wealth taxes, see p. 304. Indeed, Ricardo himself proposed a land tax as a result of his theory of capital.

century. Curiously, Ricardo (1817) himself was far from being a pessimist. He was a classical liberal and, as we can infer from his writings, thought that the moment of reaching such a calamitous state would rather lie in the distant future.

In sum, according to Ricardo (1817), capital equals stored up labor. Interest is equal to profit; and tends to disappear over time. The owners of the factors of production, however, reap a benefit, since their rents will not decline and even grow in relative terms. Hence, summarizing the Ricardian triad between land (rent), labor (wages) and capital (profit/interest), land has almost magical, automatic and permanent gains, whereas wages tend to fall and profit or interest tends to disappear completely.

2.1.3 Karl Marx and the Modern-Day Divide Between the "Real Economy" and the "Financial Economy"

Marx (1894) made some interesting contributions on capital theory that have been largely ignored or misinterpreted by either Marx himself or his many disciples. While Marx's economic thought and Marxism are often seen as a completely separate line of thought, Marx's work can be traced back to Ricardo's (1817). As Schumpeter (1954) writes:

> "**This obvious truth is that, as far as pure theory is concerned, Marx must be considered a 'classic' economist and more specifically a member of the Ricardian group**. Ricardo is the only economist whom Marx treated as a master. I suspect that he learned his theory from Ricardo. But much more important is the objective fact that Marx used the Ricardian apparatus: he adopted Ricardo's conceptual layout and his problems presented themselves to him in the forms that Ricardo had given to them. No doubt, he transformed these forms and he arrived in the end at widely different conclusions. But he always did so by way of starting from, and criticizing, Ricardo (…)" (p. 368) [emphasis mine]

Marx uses a concept called the "organic composition of capital" to explain the rate of profit. The organic composition of capital consists of variable capital and constant capital. Variable capital means, to Marx, simply the outlays in wages, whereas constant capital refers to the outlays made for production goods (which would include both working and fixed capital). However, the former (variable capital), if well-directed, can create a *surplus* value. Such surplus value, however, accrues to the capitalist (that is, the entrepreneur) rather than the worker. In short, an entrepreneur invests variable (v) and constant capital (c) and obtains a profit (*s*), hence the rate of profit is *s* / (*c*+*v*). Now, different industries have different ratios of *c* / *v*. Some industries require higher *c*: in other words, industries with a higher *c* / v ratio are more capital-intensive industries. In such industries the surplus is lower (eaten away by higher *c*), lowering the *rate* of profit (the surplus value relative to *c* + *v*). In other words, the rate of profit is *per definition* lower in capital-intensive industries than less capital-intensive industries. In Marx' (1894) own words: "[D]ifferent lines of industry have different rates of profit, which correspond to differences in the organic composition of their capitals. (…) There is no doubt, on the other hand, that aside from unessential, incidental and mutually compensating distinctions, **differences in the average rate of profit in the various branches of industry do not exist in reality, and could not exist without abolishing the entire system of capitalist production. It would seem, therefore, that here the theory of value is incompatible with the actual process, incompatible with the real phenomena of production, and that for this reason any attempt to understand these phenomena should be given up**." (p. 151) [emphasis mine]

This conclusion was, hence, not a good fit with reality. In reality, a tendency toward an equalization of profits existed. Marx correctly identified that "capital" (what rather should be called investment) is invested in industries in which the rate of profit is high relative to other industries, bringing down profits.

The component *c* (constant capital) has a tendency to increase over time, especially with technological advances, eating away the rate of profit (since more capital investment is necessary relative to *s* or surplus value). More fixed capital would thus be required and the rate of profit would come down. This, according to Marx (1894), would bring about a crisis of capitalism, as obtaining profits becomes harder and harder.

This entire analysis is flawed. What Marx (1894) fails to explain, is why the prices of variable capital (labor) and constant capital (largely fixed capital) are not bid up by competing entrepreneurs and why, consequently, profits do not disappear entirely. In fact, this is the basis of modern profit theory which explains that profits can only arise from maladjustments between essentially input prices and output prices (e.g., Kirzner, 1960). Entrepreneurs have a selfish incentive to identify such maladjustments, which are in fact latent profit opportunities, and act upon them, thus bidding up prices of inputs and bringing down prices of outputs until no profits remain. Profit can only be earned when there are maladjustments and when maladjustments seize to exist, no profits exist. Nevertheless, our continuously changing world is marked by constant new maladjustments (due to innovation or changes in consumer preferences or changed in natural conditions) which generate new profit opportunities. In fact, entrepreneurial arbitrage per definition creates new profit opportunities, as new information is created and transmitted (Huerta de Soto, 2010). **In Marx's terms, the rate of profit does not depend upon a physical input *c*, but upon price differentials between inputs *c* and *v* and product *s***.

Curiously, Marx, without noticing, completely explains in *Capital* (1894) how profits are discounted in the prices paid for capital goods and/or combination of capital goods (that is, businesses). Whenever a certain combination of capital goods has a higher rate of profit, its price is bid up, thus lowering its rate of profit until it falls to the average market rate of profit. Vice versa, whenever a certain combination of capital goods results in a lower than average rate of profit, its market price declines, thus pushing up its rate of profit toward the market average. Hence, the market tends toward an equalization of rates of profit.

Nevertheless, Marx (1894) does not explain in his example why or how an economy-wide rate of profit of 22% could possibly persist[13]. If he would, he would arrive at the same conclusion as, for instance, Mises (1949). The rate of profit would be arbitraged down by profit-seeking entrepreneurs, eager to make the most out of price differentials between inputs and outputs, until the rate of profit equals the rate of interest (that is, the "value" of time). However, Marx (1894) completely ignores the role of time.

Another curiosity arises in Marx' (1894) discussion on *money* capitalists versus *industrial* capitalists. What Marx calls money capitalists are in fact to a great extent individual savers. Such money capitalists put the resources at the disposal of industrial capitalists. This divide has, nevertheless, been often lost by other economists (e.g., Mises, 1949). Yet Marx's (1894) analysis is quite sophisticated. In fact, there are two "layers" of entrepreneurs: "industrial" entrepreneurs who invest productive assets and "capitalist" or financial entrepreneurs who manage savings and transmit savings into investment, thus deciding which "industrial" entrepreneurs are trusted with a society's saved resources.

[13] Marx (1894) uses this precise 22% average rate of profit in his example. Even though he is possibly unaware that at such a rate of profit a capitalist would double his wealth approximately every three years, 22% seems sufficiently outrageous to portray capitalists as unethical exploiters.

Marx (1894) argues that the rate of interest is bound by:

(1) The rate of profit (on the upside, that is, the rate of interest cannot be higher than the rate of profit);
(2) The supply of savings (on the downside, that is, the rate of interest cannot be lower than the supply of savings or "money capital" dictates)[14].

Profits, or *surplus value*, have some type of independence and preordained existence and are basically a bounty, unfairly expropriated from workers, that is divided between industrial and money capitalists.

While Marx's explanation of profits and interest are highly unsatisfactory (profits and interest tend to disappear or become "depressed" over time, akin to Ricardo's thesis[15]), his recognition that *profits* are linked to interest in the sense that for interest to arise "industrial" capitalists should demand savings (later formulated time preference theories of interest only explain the "supply side" of savings) and that demand for savings decreases when profits are low is valid. Marx (1894) did not, however, understand the feedback mechanism between both users of savings (entrepreneurs or "industrial" capitalists in Marx' words) and providers of savings (capitalists or "money" capitalists in Marx' words). Profits reflect maladjustments between input prices and output prices and competitive entrepreneurial arbitrage tends to eliminate profits (Huerta de Soto, 2010). When fewer savings are supplied, less funds are available for entrepreneurs to arbitrage away maladjustments[16]. Profits would be higher. Conversely, when savings are amply supplied, entrepreneurs dispose of more funds to arbitrage away maladjustments. Profits would be lower. When profits are higher, the demand for savings by entrepreneurs would be higher, which would lead to higher rates of interest that serve as incentive for savers. When profits are lower, the demand for savings by entrepreneurs would be lower, which would result in lower rates of interest that serve as disincentive for savers. Hence, there is a feedback mechanism between profits and interest:

[14] Marx (1894) wrote that "the progressive concentration of these savings in amounts which can serve as money-capital, must also depress the rate of interest." Just as there exists a long-run tendency of depressed profits, there also exists a long-run tendency of depressed interest, both of which would lead to a "crisis of capitalism."

[15] For our earlier discussion on Ricardo's theory of capital, see p. 55.

[16] We will later coin the term 'Fisher's pendulum of returns' for this phenomenon.

high rates of profit and high rates of interest would lead to low rates of profit and low rates of interest and vice versa. The historically stable long-term rate of interest suggests that such feedback mechanism ought to exist[17].

In short, Marx (1894) continues Ricardo's (1814) legacy in an important way: the labor-value theory explains prices of consumer goods and, similar to Ricardo (1814), capital goods are simply stored up labor. Marx (1894) is unable to provide a satisfactory account of profits and does not understand that profits tend to equal the rate of interest (and *equal* the rate of interest in equilibrium). However, individual owners of capital cannot simply earn rates of profit higher than the market rate *ad infinitum*, since profits are the result of arbitrage (Huerta de Soto, 2010), instead of being some inherent characteristic of material production goods. What is interesting in Marx (1894) is, on the contrary, his distinction between the "real economy" (and *industrial capitalists/entrepreneurs*) and the "financial economy" (and *money capitalists/entrepreneurs*). Far from being exploitative, however, these two different worlds are complementary and allow for a better use of scarce resources in function of a society's individuals' ends. The difference between the two is often underappreciated in modern (Austrian[18]) literature (e.g., Kirzner, 1996; Huerta de Soto, 2008). Moreover, Marx (1894) seems to define *money capitalists* as resource providers (savers), whereas there exists a large difference between resource providers (or savers/capitalists) and *financial entrepreneurs* (who intermediate savings and investment). However, in the historical context of Marx (1894), financial markets were not anywhere near their modern level of sophistication.

---

[17] For instance, since as early as 1630, long-term interest rates in the Netherlands have fluctuated surprisingly little (Van Winden, 2002). Contrary to Marx's dire predictions, long-term interest rates have not structurally declined over the past five centuries but actually, on the whole, remained within narrow bounds. A notable exception is the most recent period, in which long-term interest rates fell to all-time lows. Far from vindicating Marx, the current low interest rates are evidence of low demand by entrepreneurs since returns have been low.

[18] I refer to "Austrian" literature, since the neoclassical school of economics completely lacks a theory of the entrepreneur (Huerta de Soto, 2008). However, some neoclassical finance theory *does* capture the entrepreneurial spirit of the *financial* entrepreneur (or *money capitalist* in Marx's words): see, for example, Mehrling (2000) on Fischer Black's work and Black's own attempts to profit from his own analysis.

2.1.4 Karl Marx versus Silvio Gesell: The Return of a "Monetary Crank"

While the name of "forgotten thinker" Silvio Gesell made a comeback in the present-day era of negative (central bank) interest rates, his debates in the late 19th century with Karl Marx on capital theory remain largely in obscurity.

According to Gesell (1918), Marx (1894) mistakenly thought that "real capital" equaled the physical means of production. Money, according to Marx (1894), was a mere medium of exchange. Gesell (1918), however, argued that money is superior over commodities or physical means of production, since money (in the case of gold) does not decay. Moreover, looking at money as a mere medium of exchange, ignores hoarding. Hoarding is the very opposite of exchange. Gesell (1918) therefore criticizes Marx (1894) on the basis that Marx's capital theory amounts to a theory of material objects. In fact, interest on money is some type of opportunity cost that arises out of the sheer fact that material means of production and commodities decay and become worth *less* over time. Yet, Marx, according to Gesell (1918) "made a false assumption." "Like the orthodox apologists of interest, he [Marx] assumed that money and commodities are equivalents." (p. 170). As a result, "[t]hrough this fatal mistake Marx went astray at the outset" (ibid). Yuki (2015) summarizes Gesell's (1918) critique in the following way:

> "Gesell's capital theory is that money can collect interest from material things that degrade naturally in order to use unnatural property as a use value of immortality. Therefore, 'material things as capital theory' is incorrect. This is because considering the means of production as material things can be positioned against inferior natural property, which makes exploitation possible. The interest of real capital must essentially be explained from the interest of money." (p. 23)

One of the most interesting (albeit scarce) contributions of Gesell (1918), is the idea that money provides optionality compared to the natural decay of physical goods. In Yuki's (2015) words: "Owners of material things take into account their asymmetrical natural property in the relationship of exchange, whereby the commodity owner considers avoiding the degradation loss by selling quickly; otherwise, the owner of money waits for a market situation in which she could buy favourably through the natural superiority of money." (p. 19). However, Gesell (1918) quickly goes astray himself, as he argues that the fact that a merchant would increase his cash balance to wait for lower commodity prices would lead to a negative spiral in which every *general* decrease in prices leads merchants to wait *even longer*, such that there arises a general shortage of money and thus a crisis of "underconsumption." Gesell's (1918) solution was to impose a tax on money that equaled the natural rate of decay of material objects[19].

Yet, according to Gesell (1918), the great error was to be found in attempts to explain the phenomenon of interest by reference to productivity or abstinence. Gesell (1918) wrote: "[I]nterest has never been determined by thrift, order, industry and efficiency. I shall reject the [other] theories of capital and show that interest springs from the ancient form of money (…) and that it is protected by the physical, or legally acquired advantages of that form of money." (p. 170). As we have discussed above, interest merely arises out of the bargaining power of holders of money: "The merchant can therefore force the possessors of wares to make him a special payment in return for the fact that he refrains from arbitrarily postponing, delaying, or, if necessary, preventing the exchange of wares by holding back his money." (p. 171). Hence, the sheer possession of money allows to *manipulate* sellers of physical goods by withholding money and imposing losses due to decay on the holders of physical goods. "[M]oney is capital because it can arbitrarily interrupt the exchange of commodities." (Gesell, 1918, p. 172). Yet money produces interest not as it is being lent, but rather by abstaining from buying goods. According to Gesell (1918), interest is a monetary phenomenon, whereas profits are not and ought to be distinguished from interest (Marx and many others saw profits and interest as equals). Apparently, no competition between holders of money exists, which would avoid such a negative spiral into the abyss of deflation and underconsumption. The only limit to this arbitrary power of holders of money to postpone purchases and reap interest, is substitution into barter and *de facto* abandoning the medium of exchange *or* using bills of exchange instead of money. This "caps" the rate of interest.

Hence, Gesell (1918) quite appropriately criticizes Marx (1894) for his physical or material theory of capital, yet proposes to replace it with a rather confused concept of capital as

---

[19] This proposal has reemerged in modern times as central banks are suffering from something we could call a "reverse" Gresham's law. Since the rate of interest of currency is effectively zero (or slightly negative, accounting for insurance, transport and storage), central banks are unable to impose negative rates of interest on deposits *beyond* the carrying cost of currency, since member banks tend to substitute deposits for physical currency. The idea of taxing paper money is often contributed to Gesell, yet Gesell's proposal was different in many ways. The idea of estimating an average rate of decay of physical goods is, however, ludicrous and completely unfeasible.

money-capital. Physical capital is impaired due to wear and tear, yet money capital is not. It never dawned on Gesell (1918), however, that what he refers to as (interest-bearing) “money capital” are, in absence of 100% reserve banking and/or physical gold, simply claims on underlying “physical capital.” Hence, any claim on some underlying assets decays to the extent that the underlying asset decays, unless allowance is made for such decay (depreciation/amortization). More importantly, according to Gesell (1918), no holder of money is ultimately induced to “take profits.” When the rate of interest increases because holders of money refrain from using it, nobody considers after an *X* increase in purchasing power to convert its holdings into pleasure, that is, some type of consumption. This involves a willingness to completely abstain from *any* consumption by money holders (i.e., capitalists), which moreover is constant regardless of interest received. As Mises (1949) showed, this cannot be true *per definition*, since it implies a willingness to delay consumption *ad infinitum*.

Moreover, Gesell’s (1918) complete ignorance regarding the time element in explaining interest becomes apparent when he contributes the discounting of real bills of exchange by the fact that they “cannot replace money at all.” (p. 173). “[This] is apparent from the fact that they are frequently exchanged (discounted) at the bank for money, although they suffer thereby a deduction. This would not happen if the bill of exchange could always replace ready money.” (ibid). If real bills could completely substitute “money,” money holders would not have the same unfair bargaining power as they, according to Gesell (1918), hold. Gesell (1918) completely fails to grasp the “price of time” as explanation of the observed phenomenon of banks discounting bills. The existence of bills of exchange *caps* the rate of interest on money, rather than that bills of exchange are discounted *with* a market rate of interest. Indeed, Gesell (1918) never explains what caps the rate of discount of bills. He merely mentions the fact that bills are discounted somewhat more or less depending on the creditworthiness of its counterparty.

To sum up, Gesell (1918) rightly criticizes Marx (1894) for focusing on a physical or material concept of capital. Nevertheless, he replaces it with a less satisfactory concept of “money capital,” which is a result of some arbitrary withholding of money by holders. Modern money is simply a claim on underlying assets, thus any physical decay is already priced in. Interest can therefore never be a result of manipulation by money holders, contrary to what Gesell

(1918) says. Keynes's praise of Gesell therefore seems to be quite exaggerated (Keynes, 1936).

The whole idea that interest as income is some type of exploitation of workers (e.g., Marx, 1894) or non-money holders (e.g., Gesell, 1918) was completely refuted by various authors (e.g., Böhm-Bawerk, 1888), but their critique was perhaps best formulated by Fisher (1930):

> "Socialists would cease to think of interest as extortion if they would try the experiment of sending a colony of laborers into the unreclaimed lands of the west, letting them develop and irrigate those lands and build railways on them, unaided by borrowed capital. The colonists would find that interest had not disappeared by any means, but that by waiting they had themselves reaped the benefit of it. Let us say they waited five years before their lands were irrigated and their railway completed. At the end of that time they would own every cent of the earnings of both, and no capitalist could be accused of robbing them of it. But they would find that, in spite of themselves, they had now become capitalists, and that they had become so by stinting for those five years, instead of receiving in advance, in the shape of food, clothing, and other real income, the discounted value of the railroad." (p. 25)

In sum, both Marx's and Gesell's theories of capital, interest and money grounded in exploitation and manipulation are highly unsatisfactory in explaining their existence and nature.

2.1.5 On the Dichotomy between Labor and Capital

In many instances (e.g., Ricardo, 1817; Marx, 1894), economists view labor as a subtractive rather than an additive element, even in modern times (e.g., Piketty, 2014; Knoll et al., 2017). Any increase in labor wages, reduces returns on or impairs capital[20]. In effect, the entire literature on capital is permeated with, what Huerta de Soto (2008) calls, "static efficiency", which is focused on *given* resources. With labor and capital as pure inputs, which are given, one is only concerned with optimizing the output, which is in most cases operationalized as economic growth. However, Huerta de Soto (2008) convincingly shows that what matters

[20] While many economists held this view, it was Karl Marx who turned it into an ethical phenomenon with his notion of "exploitation of labor". These very same views can be found in present-day authors, such as Thomas Piketty (2014).

most in real life, opposed to in equilibrium, is "dynamic efficiency," in which resources are not given and optimization is not a mechanical or mathematical exercise. In this case, labor should be seen not as a mere "cost", but rather as something akin to an "investment" that can add to capital and/or output, however one might define output.

In many cases, as a consequence, "capital" is pitted against "labor." Whatever "capital" accrues, comes at the expense of "labor." Hence, wealth concentration is a chronic problem as long as the rate of return on capital (r) exceeds the increase in wage rates or economic growth rate (g). However, Piketty (2014) does not seem to distinguish a mean from the composition of mean. Concentration implies the same individuals accumulate wealth at the average rate of return, but in fact, the variability (or mean absolute deviation) of returns[21] is very wide (Bach, Calvet, & Sodini, 2017). This means that concentration is largely a myth and is intimately connected to the general ebb and flow of returns on capital markets (principally the stock market).

Fisher (1930) and Knight (1938) as a result included human capital in their capital theory, as we will see further below. Besides the obvious truth that human capital and nonhuman capital are often complementary, these authors show that human capital has a return on capital just as nonhuman capital does. Whenever the returns on human capital are low, wages are bid up and the "share of labor" of (total) income increases. Contrarily, whenever the returns on human capital are high, wages are bid down and the "share of labor" of total income decreases. However, both the contributions of humans and nonhuman "productive" assets mean-revert to their respective value-added as governed by consumer sovereignty.

2.2 The Böhm-Bawerk/Clark Controversy on Capital Theory

Interestingly, the publication of Böhm-Bawerk's most important books *Capital and Interest* (1884) and *Positive Theory of Capital* (1889) sparked the interest of J.B. Clark, who published various essays on capital in response to Böhm-Bawerk that were later recollected in Clark's *Distribution of Wealth* (1899). The debate began with a critique by Clark (1893) on Böhm-Bawerk's *Positive Theory of Capital* and continued in a fierce back-and-forth between both economists (Cohen & Drost, Böhm-Bawerk's Letters to J.B. Clark: A pre-Cambridge

[21] Bach *et al.* (2017) write: "The standard deviation of the financial portfolio held by the top 1% households is about 24% per year, as compared to 12% for the median household." (p. 2)

controversy in the theory of capital, 1996). Common tenets were noticed, but as time proceeded, key differences surfaced. Böhm-Bawerk took a break from the debate as he begun dedicating most of his efforts to his work as Austrian minister of finance until 1904, the year he resigned, before returning to the debate on capital with J.B. Clark in 1906. In 1906, Böhm-Bawerk published his (delayed) critique on Clark's *Distribution of Wealth* (1899), which led to a new series of exchanges which ended in 1907, albeit many issues remained unsolved.

Hence, the first historic debate on capital theory became a fact. Blaug (1997) went as far as calling the Böhm-Bawerk / Clark controversy "one of the three great controversies that have marked the history of capital theory" (p. 547). Many of the unresolved issues reappeared in later capital controversies in the 1930s (the "second round") and the 1960s ("the third round").

According to Cohen (2007), the Böhm-Bawerk / Clark controversy has its origin in the *dual nature of capital*: "Economists (including Böhm-Bawerk and Clark) have long conceived of capital both as a heterogenous collection of specific capital goods used in production, and as a homogeneous fund of financial value that flows among alternative uses to establish a uniform rate of return" (p. 2). Cohen (2007) argues that Clark emphasizes the monetary (or financial) nature of capital, whereas Böhm-Bawerk emphasizes concrete capital goods, while both sooner or later end up wrestling with, what Cohen calls, capital's *dual nature*, on which both authors agreed. Nevertheless, both economists struggled with breaking away from the legacy of the classical economists' fallacies, even though Clark puts a greater emphasis on capital as a "fund."

2.2.1 Early Contributions by British Economist John Bates Clark

John Bates Clark began publishing on capital theory at the end of the 19th century, which resulted in the publication of a series of articles that culminated in *The Distribution of Wealth* (1899). Clark (1891) distinguishes between "true" capital (that is, financial capital) and concrete capital goods (that is, production goods). The former produces interest, the latter produce rent. Said otherwise, on the one hand there is financial capital with interest as income, and on the other hand there are capital goods with rent as income[22].

As we will see later, John Bates Clark's article on capital titled *Capital and Its Earnings* (1888), initially shows some resemblances with Menger's essay on capital (which we will discuss further below). Both begin by showing that, from the point of view of an entrepreneur, capital is usually expressed in a dollar-amount. It is invested in capital goods, but capital, in Clark's words, is "an abstract fund" (p. 10). In Clark's (1888) own words:

> "Take an inventory of a hardware merchant's stock. Make a complete list of saws, hatchets, nails, etc., that his shelves and store-rooms contain. Have you determined what is his capital? Not, according to his own view, **until you have attached to each article on the list the figure that represents its market value**, and added the figures into a sum total." (p. 10) [emphasis mine]

Capital is permanent, then, since even though the underlying assets cease to exist, they are in the course of normal business operations replaced but the "value" remains intact. Clark (1888) continues by distinguishing circulating capital from fixed capital, calling goods that are considered circulating capital "passive instruments" (such as inventories of raw materials), while calling goods considered fixed capital "active instruments." He then provides a critique on the wage fund theory, which argues that capital is a "fund" from which current wages are being paid to cover a given period of production. It is silly, according to Clark (1888) to think that somebody saved all future wages of the entire workforce and that this fund is ought to be called capital. Rather, businesses have, as part of their circulating capital, some allowance to cover the wages for the months necessary as they match cash *inflows* with cash *outflows* (in this case wages). Yes, abstinence initially gave rise to capital,

---

[22] We will see further below that this distinction makes sense, but we will use a more apt modern terminology. Mises (1940) shows that in an equilibrium state, the cost of capital equals the return on capital and no arbitrage opportunities exist. Thus, in equilibrium, the income on Clark´s "financial capital" would equal the rate of interest. Curiously, his concept of "capital goods" that produce rent only makes sense in a non-equilibrium state, where arbitrage continues to exist between the price paid for the capital good (or combination of capital goods) and its corresponding return ("rent", according to Clark). Further below we will see, however, that his distinction between interest and rent is untenable.

and partially to circulating capital, of which a share is used to pay ongoing wages, but it is far from being the defining factor of capital (we will delve into Clark's notion of capital as "permanent" further below).

However, as soon as J.B. Clark (1888) delves into the subject of rent and interest in his chapter *The Earnings of Capital*[23], he makes some unjustifiable theoretical claims. Before defining rent and interest, he argues that land is a production factor equal to tools, ships or buildings and should therefore not be treated any different from a theoretical point of view (even though, at a later point, he singles out land by arguing that part of its rent is derived from monopoly power). So far so good. However, Clark's errors quickly accumulate.

First, Clark (1888) defines rent as the "amount earned by concrete productive instruments of any and every kind" (p. 29). He argues that rent should be expressed as a (dollar) lump sum, and not as a rate or percentage. This is perhaps odd for modern-day readers, as we are very much accustomed to express *any* return, without exception for specific rents, percentagewise for the sake of comparing returns. However, Clark suddenly takes the entirely opposite view and claims that if you list various productive instruments (in other words, assets), take their market value, and then compare them to the earnings they produce (their returns), we *are* allowed to express it as a (percentage) rate of return. A mindful reader would be surprised to see Clark apparently approving of expressing rents as rates of return in case of *a group of assets*, while disapproving the very same act in case of *a single asset*. To add insult to injury, Clark (1888) then makes a case for considering "interest" as a synonym of "rent" whenever it refers to the rate of return of a group of productive assets.

Consequently, Clark (1888) explains that any increase in wages ("the reward of labor") diminishes the rate of return on invested capital. We will see that the interaction between capital on the one side and wages on the other side is a recurring theme in historical controversies on capital (Joan Robinson, for starters, revisits this idea in her treatment of capital in the Cambridge Controversies of the 70s, which we will discuss further below). In fact, Clark's treatment of wages in the context of capital theory appears puzzling. While capital goods are heterogenous, only the assumption of capital as being a "homogenous" fund allows us to express different goods in terms of a single unit (in this case, in terms of money).

[23] In line with modern-day terminology, we refer to the "return on capital" (or, in short, ROC or ROIC), whereas Clark (1888) uses the phrase "earnings of capital".

Labor, however, is homogenous, and *can* be measured directly *without* any resort to a common denominator: "Labour (…) is capable of being measured in units, as though it were homogeneous; and there is a practical method of measuring the product of all of it" (Clark, 1899, p. 332). Clark (1899) persists in the same classical pre-industrial fallacies of distinguishing between land, labor and capital, albeit in a slightly different sense than most classical economists. Land is just as much capital as any other capital good, but is different since it gives rise to monopoly rents. Hence, land is different from all other capital.

Moreover, labor is also different from capital, since it can be directly measured in terms of homogenous labor units. As Clark (1899) writes with regard to labor, "we do not consider acquired abilities of workmen as a part of **the fund of productive wealth [i.e., capital]**. Man does not add to his capital, when he spends money in training or educating himself for a useful occupation. He gets something, indeed, that increases his productive power; and in getting it he is obliged to practise abstinence. (…) There is, it must be admitted, a certain similarity between the effects of money spent on a technical education and those of money spent in buying a tool." (p. 116) [emphasis mine]. As a consequence, "(…) capital is never a quality of man himself (…). The capital of the world is, as it were, one great tool in the hand of working humanity (…)" (ibid, p. 117). Hence, labor is different from all other capital. Human capital is ignored.

In addition, Clark (1899) formulates the theory of marginal diminishing returns on capital, which was kept alive in very much the same shape as his neoclassical and Keynesian followers (e.g., Solow, 1956; Samuelson, 1960): "[N]o increment of capital can get for its owner more than the last increment produces." (p. 105). Yet, if capital is a "value" fund, rather than a physical means of production, then Clark's notion of diminishing returns can only be valid if returns are equal everywhere (zero variance) and absolutely no further changes occur. However, Clark (1899) seems to have a more physical notion of diminishing productivity in mind.

In one of his main contributions, Clark (1888) identifies something he refers to as "pure mercantile profit," which he separates from his concept of interest, which is "a premium for (…) the perfecting of industrial organization" (p. 30). Nevertheless, he argues, such "pure" profits tend to disappear due to competition. These profits, sooner rather than later, are absorbed by wages and interest. Although Clark fails to explicitly describe the source of such “pure” profits, his recognition of “pure profit” as a separate category from “interest” is encouraging.

Clark (1888), in an effort to elaborate on his concepts of rent, interest and wages, resorts to an example: He says that the rent of a ship is simply “its product minus (...) wages and interest”, to which he refers as the rent formula. Interest, in this case, refers to the returns on the separate isolated productive assets and not to actual interest payments. Interestingly, Clark (1888) explains how arbitrage leads to an “equalization of the earnings of pure capital” (p. 45). “Pure capital gravitates to the points of greatest returns; it seeks out and vests itself in concrete forms that, as tested by the rent formula, give the greatest earnings” (ibid, p. 45).[24]

This arbitrage process, so aptly described by Clark (1899), requires an actor, which is the entrepreneur. In J.B. Clark’s theory of capital (1899), the entrepreneur has a rather prominent role: “Normal prices are no-profit prices. They afford wages for all the labor that is involved in producing the goods (…). They afford, also, interest on all the capital that is used in the business (…). Beyond this there is no return, if prices stand exactly at their normal rate; and the reason for this is that *entrepreneurs* compete with each other in selling their goods, and so reduce prices to the no-net-profit level” (p. 111). (Notice that “normal” for Clark means “equilibrium.”)

This rather profound insight of Clark involves two conclusions that still hold: (1) profits are arbitraged down until no pure entrepreneurial profit remains, so that nominal profits only reflect interest on capital (isolating interest as a recompense for time from a return on capital), and (2) wages are bid up by competing entrepreneurs to reflect their contribution to profits. In other words, wages depend on the marginal productivity of workers; if workers would be underpaid, competition between entrepreneurs would assure that this “wage gap” is

[24] Clark (1888), at the end of his essay, shows that returns on capital (“earnings on capital”) can only be excessive when there exists a monopoly which impedes competition and new capital from entering and reducing such returns (“Outside capital then presses from every direction upon the artificial barrier that speculation has erected, like rising water upon a coffer-dam. The breaking of the barrier is, in the end, inevitable, and the task imposed upon the government in its protection of the equities of capital is lightened” p. 65). He advocates government to leave capital “free”: “Free purchase and sale, the unimpeded flow of capital to the points of large reward, -this is the safeguard against monopoly.” (ibid).

corrected, not out of compassion for the workers, but out of self-interest given the latent profit in doing so. Wages are, effectively, arbitraged by entrepreneurs up to their marginal contributions to an entrepreneur's profit. "[Competition] annihilates the profit that an employer might make on the last increment of labor that he hires" (Clark, 1899, p. 112).

Moreover, studying the distribution of wealth, Clark (1899) does away with the "traditional" distinction between land/rent, capital/interest and labor/wages and replaces it with, what he calls, his "natural law." Clark (1899) writes: "It causes the whole annual gains of society to distribute themselves into three great sums – general wages, general interest and aggregate profits, These are, respectively, the earnings of labor, the earnings of capital and the gains from a certain coordinating process that is performed by the employers of labor and users of capital. This purely coordinating work we shall call the *entrepreneur's* function, and the rewards for it we shall call profits." (p. 3). Hence, what Clark (1899) calls *earning of capital* is mere *interest*, a compensation that reflects the productivity of individual production goods. On top of interest, there are "pure" entrepreneurial profits, which – however – tend to disappear over time due to competition. The entrepreneur, from Clark's perspective, is a coordinating force, very much like the Kirznerian entrepreneur (e.g., Kirzner, 1960).

All this adds up to a rather confusing and ambiguous treatment of rent and interest. What we wind up with, is a rather dissatisfactory treatment of rent and interest by this British economist. On the one hand we have a brilliant exposition of capital as a fund denominated in terms of money, a profit-arbitraging entrepreneur, interest as something distinct from pure entrepreneurial profit (return on capital) and competition as a way of assuring that no surplus profits accrue to entrepreneurs, while on the other hand we have inadmissible ambiguities with regard to land, labor, the exclusion of circulating capital from the concept of capital, and individual capital goods which, when considered in isolation, are no longer capital according to Clark. More importantly, Clark's treatment of interest is flawed; he does not incorporate the time element, but explains interest by referring to the returns (rents) on *individual* assets (accruing to their owners, that is, *capitalists*), separate from the (residual) returns on *a combination* of assets (which he calls profits, which accrue to entrepreneurs). Interest is, however, determined by the marginal productivity of the underlying assets.

This latter view cannot be possibly true: while Clark (1899) assumes competition between entrepreneurs, he assumes *zero competition* between owners (capitalists, in Clark's terms). In other words, Clark (1899) fails to explain *why* the prices of the underlying assets are not bid up to account for their marginal productivity or rents. This is the main reason why Mises (1949) would later call Clark's theories of capital and interest a product of a "naïve" physical productivity theory. Unfortunately, despite some advances (which we have mentioned above), Clark's treatment of interest is particularly poor.

2.2.2 Böhm-Bawerk's Agreements and Disagreements with J.B. Clark

Böhm-Bawerk (1890) defined capital as "a complex of produced means of acquisition, that is, a complex of goods that originate in a previous process of production, and are destined, not for immediate consumption, but to serve as means of acquiring further goods." (p. 6). He goes on by saying that: "Objects of immediate consumption, then, and land (as not produced) stand outside our conception of capital." Note the traditional distinction Böhm-Bawerk (1890) makes between land, labor and capital, a distinction that still sways neoclassical production functions and growth models, in contrast to Clark (1899), who at least considered land and capital to be of the same nature (even in Clark, labor was something separate and measurable in terms of "homogenous" labor units). In fact, Böhm-Bawerk (1891) later defines capital in different words as "an intermediate product of nature and labour, nothing more" (p. 86). Yet Böhm-Bawerk, when he began elaborating his theory of interest, tacitly defects to a different definition of capital: namely, to a homogeneous subsistence fund (wage fund) expressed in terms of money (Cohen & Drost, 1996).

Clark (1899) on the other hand, defines capital as consisting "of instruments of production, and these are always concrete and material" (p. 116). Later, Clark (1899) implies that, even though capital consists of instruments of production, it is different from such instruments of production, as we have seen earlier above. Clark's own understanding of capital appears to be confused at times. For instance, as noted by Groenewegen (1999), Clark pictured capital as an immense pile of money ready to be invested: "sixty-five billion dollars (…) free to invest themselves in other things (…)" (Clark, 1899, p. 113). In this sense, Clark viewed capital more as a "fund," although not in the sense of for instance Fisher (1930), who viewed capital simply as the net present value of the future income that the underlying capital goods are able to yield. Rather, Clark sometimes appears to "confuse" capital as being a "liquid fund" that can readily be invested in whatever lines of production are most profitable and capital as the dollar value of capital goods.

Capital goods are "perishable" and rightly need to be so, whereas capital is "permanent," understood as that it *must* last "if industry is to be successful" (Clark, 1899, p. 117). "Capital goods must perish in order that capital may abide" (ibid). Capital is permanent in the sense that depreciation charges allow for a continual replacement of the underlying material capital goods. In addition, capital goods are far from mobile, but capital is: "It is possible to take a million dollars out of one industry and put them into another" (Clark, 1899, p. 118).

Böhm-Bawerk took aim at Clark´s assumptions of *permanency* and *mobility*. Clark (1899) was most famous for his waterfall analogy, with which he attempted to prove that capital is "permanent": a waterfall consists of particles of water. Capital goods are like such water particles; they condense in clouds and thus the individual particles disappear. But the waterfall is permanent, since its source, say the sea, is constantly replenished with new water particles. Böhm-Bawerk (1895) countered with an analogy of a watermill: if a large stone would splash away the water of the creek, no water would reach the millwheel and the mill would stop. The fact that capital appears "permanent," is because entrepreneurs and capitalists deliberately decide to reinvest cash flows and replace the "nonpermanent" capital goods. Moreover, capital can be moved into other industries, but often not at a zero cost. Hence, by criticizing Clark's (1899) idea of *perfect mobility,* the notion of capital as a "permanent" fund also becomes a sheer impossibility. If the assumption of perfect mobility is erroneous, the value of the fund can fall when capital is moved from one industry to another. Indeed, as Cohen (2007) sums up Böhm-Bawerk's position: "Böhm-Bawerk agrees that there is mobility between sectors through the wearing out of old capital goods and investment in new and different capital goods. But he registers two objections. First, that this process is limited and could not take place on a massive scale all at once, contradicting Clark's claim of perfect and absolute mobility." (p. 18). Second, Clark's propositions are mere "figure of speech" and do not add to our understanding. In effect, Böhm-Bawerk took Clark's illusion of capital as a "permanent" fund to pieces, albeit it is clear that Clark's notions of permanency and mobility in reality *do* have some grain of truth. As a result, with regard to the notions of permanency and mobility, Clark (1899) seems much more "mechanic" in his approach to capital than Böhm-Bawerk. The notion of permanence also led Clark (1899) astray in his elucidation of interest. Clark (1899) explained that abstinence was only necessary at the conception of capital. After a first act of abstinence, capital was able to maintain itself indefinitely. No new acts of abstinence are therefore necessary. As Cohen (2007) writes: "Clark's error stems from his assumption of the permanence of true capital – that, once created, capital permanently generates both funds for its own replacement *and* a stream of interest." (p. 20) [emphasis mine]. As we will see later, this is sheer nonsense, since the act of replacing worn-down capital goods is not indifferent. An owner can decide to *not* replace his capital and consume whatever he saves by avoiding the need to reinvest capital to sustain production. Any replacement of capital, which would refer to both "true" and "concrete" capital, requires a renewed act of abstinence.

So, what then does Böhm-Bawerk precisely mean with capital as the produced means of acquisition? Böhm-Bawerk (1891) explicitly states that the means of acquisition simply equal the means of production plus consumption goods lent to foreign countries. But we can see how Böhm-Bawerk's critique of Turgot explains what he implies by *means of acquisition* with even greater clarity.

Böhm-Bawerk (1891) criticizes Turgot for designating all saved goods "indiscriminately" as capital. Here we can see how, beyond any doubt, Böhm-Bawerk's own view of capital differs from the French 18th century economist Turgot. Böhm-Bawerk follows in the footsteps of Adam Smith (1776), by arguing that "stocks of consumer goods" yield no (monetary) income, and therefore should not be considered capital. Other saved stocks, which are capable of yielding (monetary) income to its owner, are rightfully considered capital. In Böhm-Bawerk's own words, only goods capable of yielding interest are capital.

The argument against Turgot, and in favor of Adam Smith, boils down to Smith's distinction between *national* and *private* capital. Private individuals, according to Smith, "can make a gain, not only by the production of goods, but also by lending to other individuals (…) which are destined in themselves to immediate consumption, such as housing, masquerade dresses, furniture, etc." (Böhm-Bawerk, 1891, p. 26). However, on an aggregate level, a community as a whole cannot become richer other than by producing goods. "For the community, then, the conception of "means of acquisition" coincides with the otherwise narrower conception of "means of production" (p. 26).

Regrettably, Böhm-Bawerk then sets himself the task of classifying each and every economic good (means of production): when is a good considered to be capital and when not? This hopeless exercise culminates in some notable exclusions from Böhm-Bawerk's concept of capital (Endres, 1997): "debts and other claims, 'goodwill', (…) certain rights and relations such as patents, customer-relationships, legal claims." (Böhm-Bawerk, p. 74). While Endres (1997) correctly points out that double counting should be prevented in the case of debts or claims, it should be noted that Böhm-Bawerk is swayed by an extremely physical or material capital theory, where no room is left for intangible assets because they do not fit his incoherent justification of *why* people would save more if there are no gains in physical output (that is, the physically superior productivity of more roundabout methods, which are only feasible with an increase in savings and lower time preference). Böhm-Bawerk is essentially forced to exclude certain goods from his concept of capital to be able to justify his theory of the greater physical productivity of more roundabout production methods, thereby deviating further from a truly subjectivist view of economics.

Yet, this physical conception of "production" is completely and utterly wrong. Physical production is not the subject matter of economics, but rather coordination or, in different words, *value* production. Kirzner (1960) would say that economics is about "the necessity to reconcile numerous conflicting ends under the shadow of an inescapable scarcity of means" (p. 17). This unfortunate legacy of Adam Smith (1776), who essentially analyzed a primitive corn economy, is not merely reflected in Böhm-Bawerk's work on production, but also in his work on capital as a subsistence fund. Schumpeter (1954), in contrast, does not trace Böhm-Bawerk's (and others') confusion regarding "physical capital" back to Smith, but even further back to Nicholas Barbon (1690), while noting that Böhm-Bawerk essentially commits the same error as Smith (1776):

> "But most of them [that is, economists from 1870 to 1914) still took the view that we have traced to Nicholas Barbon (…), namely, that interest constitutes the bulk of business gains—**the part of business gains that results from the application of physical capital and is a return to physical capital in the same sense in which rent is a return to land and wages are a return to labor**. In this respect, it is highly significant that Böhm-Bawerk, in his critical history of interest theories, dealt with Ricardo's and Marx's theories of 'profit' without raising the question whether the returns thus denoted were really the same thing as 'interest.' He would have answered this question much as A. Smith or J.S. Mill had answered it. Monetary interest remained for him simply the shadow of the interest that is earned by supplying physical goods—which really were what, though perhaps 'in the form of money,' the capitalist owned. **This is all the more remarkable because Böhm-Bawerk's own work was principally instrumental in dissolving this schema**. (Schumpeter, 1954, pp. 891-892).

Schumpeter (1954) is apparently just as confused as us in noting this paradox in Böhm-Bawerk's work. While Böhm-Bawerk's work in refuting naïve (physical) productivity theories of interest was instrumental, he persisted in the same obsolete, untenable distinction between physical capital, land and labor, something Menger (1888) avoided. Böhm-Bawerk's view of capital as "stored-up labor and land" was precisely untenable because it is backward-looking, whereas decision makers are only concerned with the future and not about the past (Kirzner, 1996).

Besides the ancient idea that capital goods yield interest, labor yields wages and land yields rent, one should note the narrow conception of what, according to Böhm-Bawerk (1891), income represents. Income is monetary and is the equivalent of "interest." The fact that "income" could also be something non-monetary (that is, a house owned by worker providing housing services over time) is not considered. In fact, many of these "stocks of consumer goods" can be turned into income-generating assets (for instance, a house can be rented). Income can be narrowly defined as interest, but it can also be defined broader, including non-monetary income. As Huerta de Soto (2010) shows, income is a broader concept which includes any improvement in the wellbeing of an individual. Böhm-Bawerk only considers the former whilst ignoring the latter. Böhm-Bawerk (1890) confirms this conclusion by stating: "The income that flows from capital (…), we shall simply call interest." (p. 7)

Later economists, such as Ludwig Lachmann (1976), were unapologetic with regard to Böhm-Bawerk's capital theory. As Lachmann (1976) reflects: "Böhm-Bawerk never meant to be a capital theorist. He was essentially a Ricardian who asked a Ricardian question: "Why are the owners of impermanent resources able to enjoy a permanent income and what determines its magnitude?" The notion of a temporal capital structure consisting of a sequence of stages of production was a mere by-product of an inquiry into the causes and the magnitude of the rate of return on capital and not the main subject." (p. 89). Indeed, Böhm-Bawerk was mostly concerned with the distribution of income, not capital theory *per se*.

As a result, Böhm-Bawerk gets inexcusably caught up in a self-imposed dilemma to reconcile capital theory fully with his interest theory, that is, defining capital in such a way that interest is derived from that capital. Since production takes time (Mises, 1949), the theory of interest is without a doubt to a very important extent related to production. But the same goes for consumption. In fact, the theory of interest is just as important to production as consumption, especially when we refer to durable consumer goods.

Böhm-Bawerk (1890) proceeds by discussing the phenomenon of interest. He distinguishes between:

- Gross interest
- Net interest

These concepts have nothing to do with their modern-day equivalents. In modern finance terms, Böhm-Bawerk's *gross interest* roughly equates to Earnings Before Interest and Taxes, (EBIT). As Böhm-Bawerk (1890) himself writes:

> "The expression gross interest covers a great many heterogeneous kinds of revenue, which only outwardly form a whole. **It is the same thing as the gross return to the employment of capital**; and this gross return usually includes, besides the true interest, such things as part replacement of the substance of capital expended, compensation for all sorts of current costs, outlay on repairs, premiums for risk, and so on. Thus the Hire or Rent which an owner receives for the letting of a house is a Gross interest; and if we wish to ascertain what we may call the true income of capital contained in it, we must deduct a certain proportion for the running costs of upkeep, and for the rebuilding of the house at such time as it falls into decay." (p. 7)

On the other hand, *net interest*, according to Böhm-Bawerk, is akin to the "originary rate of interest" of Mises (1940), which reflects a society´s time preference. However, Mises (1949) argues that Böhm-Bawerk's mistake in his concept of *net interest* is his explanation of *where* interest comes from: "The higher productivity of more time-consuming roundabout methods of production which is referred to by Böhm-Bawerk and by some later economists in the explanation of interest, does not explain the phenomenon. It is, on the contrary, the phenomenon of originary interest that explains why less time-consuming methods of production are resorted to in spite of the fact that more time-consuming methods would render a higher output per unit of input." (p. 523).

Curiously, Mises (1949) first credits Böhm-Bawerk with "unmasking the fallacies of the naïve productivity explanations of interest," only to accuse him for committing the exact same mistake moments later. Mises (1949) wrote:

> "Böhm-Bawerk has once for all unmasked the fallacies of the naive productivity explanations of interest, i.e., of the idea that interest is the expression of the physical

productivity of factors of production. **However, Böhm-Bawerk has himself based his own theory to some extent on the productivity approach**. In referring in his explanation to the technological superiority of more time-consuming, roundabout processes of production, he avoids the crudity of the naive productivity fallacies. But in fact he returns, although in a subtler form, to the productivity approach. Those later economists who, neglecting the time-preference idea, have stressed exclusively the productivity idea contained in Böhm-Bawerk's theory cannot help concluding that originary interest must disappear if men were one day to reach a state of affairs in which no further lengthening of the period of production could bring about a further increase in productivity. This is, however, utterly wrong. Originary interest cannot disappear as long as there is scarcity and therefore action." (p. 525) [emphasis mine]

As Cohen (2007) notes: "It is easy to show that a capital good yields additional physical output and even revenue, but why is the *value* of the capital good not bid up to eliminate any payment of interest? What must be explained is not the gross return, but the net return to capital" (p. 6). On a similar note, Mises (1949) adds, directly addressing his criticism to Böhm-Bawerk's theory of interest: "[T]he phenomenon of originary interest explains why pieces of usable land can be sold and bought at finite prices. **If the future services which a piece of land can render were to be valued in the same way in which its present services are valued, no finite price would be high enough to impel its owner to sell it**. Land could neither be bought nor sold against definite amounts of money, nor bartered against goods which can render only a finite number of services. Pieces of land would be bartered only against other pieces of land. A superstructure that can yield during a period of ten years an annual revenue of one hundred dollars would be priced (apart from the soil on which it is built) at the beginning of the second year at none hundred dollars, and so on." (p. 523) [emphasis mine]. Hence, the only reason why the price of a capital good is not bid up to eliminate excess profits (referred to as *gross interest* by Böhm-Bawerk) is the phenomenon of time preference (referred to as *net interest* by Böhm-Bawerk).

Fisher (1930) effectively joins Mises in his critique of Böhm-Bawerk's theory of interest:

"Böhm-Bawerk presented the agio theory, or what is here called the impatience or time preference theory, clearly and forcibly, and disentangled it from the crude and incorrect notions with which it had previously been associated. It was only when he attempted to explain the emergence of this agio by means of his special feature of

> "technical superiority of present over future goods" that, in my opinion, he erred greatly." (p. 152)

Physical productivity or more roundabout methods of production are irrelevant in explaining interest. Yet the idea that capital yields interest as income *per se*, is a persistent myth in economic literature, even up till today (e.g., Piketty, 2015). They view capital analogous to an apple tree; capital yields interest as an apple tree yields apple, regardless of the price paid. A machine might yield products, but if the incurred costs to build or buy the machine exceed the income that its products yield, it is economically worthless at that very price point and thus, most certainly, *not* capital.

While, as noted, Böhm-Bawerk largely avoids the "naïve" productivity explanation of interest, he partly resorts to the "technical superiority" of more roundabout processes of production in explaining the existence and origin of "time preferences." Böhm-Bawerk attempts to prove the technical superiority by his concept of the "average production period." The longer the average production period, the greater the output. Therefore, production processes with longer production periods yield, per definition, greater output. Nevertheless, Böhm-Bawerk's own elaborate tables would not show this statement to be true (Fisher, 1930). Indeed, according to Böhm-Bawerk's (1888) productivity tables based on labor inputs (pp. 266-267), a month of labor available in 1888 was most profitably invested in the production process that yields 840 in value in 1890. Yet a month of labor available in 1899 yielded the highest value, 720, when invested for 1893. Böhm-Bawerk (1888) thus concludes that labor in 1888 is more productive than labor available in 1890, 1891, et cetera: "(…) methods of production which take time are more productive. (…) The lengthier the productive method employed the greater the quantity of products that can be obtained." (p. 260). Hence, the fact that one production process, according to Böhm-Bawerk's tables, yields 840 and another 720 is, according to Böhm-Bawerk, due to technical superiority of a lengthier production process, independent from time value. Nevertheless, if we flip Böhm-Bawerk's table upside-down, so that the longer the period the smaller the return (Fisher, 1930), the table will still show that a month of labor invested in 1888 gives the highest return, directly contradicting Böhm-Bawerk's thesis of the superior physical productivity of more roundabout (longer) periods of production. Hence, even though he explicitly denies he does, Böhm-Bawerk sneaks in his two other "causes" of time preference (that is, the fact that we discount future goods and that we plan to provide for our wants at different points in time). If

Böhm-Bawerk would keep these two other causes equal in his "technical superiority" analysis, he would be forced to conclude that technical superiority *cannot* be a cause of time preference and that he, effectively, falls into the same trap as other physical productivity theories of interest, which he so extensively criticized. As Fisher (1930) explains: "[T]he conclusion is that, if we eliminate the "other two circumstances" (relative underestimate of, and overprovision for, the future[25]), we eliminate entirely the superiority of present over future goods. The supposed third circumstance of technical superiority, in the sense that Böhm-Bawerk gives it, turns out to be non-existent." (p. 154). Therefore, "(…) the only reason any one does prefer the product of a month's labor invested today to the product of a month's labor invested next year is that today's investment will mature earlier than next year's investment." (ibid).[26] As Cohen (2011) concludes: "Böhm-Bawerk is wrong in claiming the independent effect of the third cause of interest [that is, his explanation of the higher physical productivity of more roundabout production methods as a source of time preference and thus interest], and his insistence on explanation beyond a consistent simultaneous equations model seems wrongheaded or, at best, confounding." (p. 31)

Clark's (1899) explanation of interest, however, was what Mises (1949) perhaps had in mind when he referred to "the naïve productivity explanations of interest." Kirzner (1996) compares Clark's "naïve productivity explanation" with Böhm-Bawerk's "less naïve productivity explanation:

> "J.B. Clark (…) saw capital as providing a flow of productive services, of which interest is the irrepressible expression. Competition does not erode it; ownership of a stock of capital inevitably confers title to a corresponding income flow. [For Böhm-Bawerk] (…) ownership of capital expresses the provision of a special productive service ('waiting') required in order to enjoy the enhanced fruits of more lengthy ('roundabout') processes of production. Competition cannot erode interest income: it *has* to be offered if potential capitalists (with positive time-preference) are to be

[25] Fisher (1930) refers here to the idea that people have subjective time preferences: they provide for future want satisfaction at different points in time *and* to the idea that future goods tend to be discounted against present goods.

[26] Fisher (1930) would reformulate the "technical superiority" of production to a theory of "investment opportunities" to explain the origin of the demand side in intertemporal markets, see p. 143.

persuaded to provide the waiting (needed in order to be able to enjoy the enhanced output available through capital-using production). And, given the productivity of waiting, it *pays* to offer interest in order to elicit that waiting." (p. 101).

Böhm-Bawerk's (1888) treatment of capital and interest is inherently contradictory. At some point, he defends capital as the structure of means of production, excluding durable consumption goods and money, while on the other hand he defines capital as a "subsistence fund" out of which the present consumption of workers can be paid. For instance, Böhm-Bawerk (1888) writes:

> "If the existing stock of subsistence is so great as to defray four million years' pay—in which case, as we know, where production is by stages, an initial capital amounting to two millions of wages only would be required—and if there are one million labourers in the country, then it is shown that an average four years' production period must be taken. For if, say, a three years' period were taken, the three years' payment of one million of workers would take up only a capital of one and a half millions of wage, and the rest of the capital would have to go idle. In a five years' production, again, an initial fund of two millions of wages would only defray the subsistence of 800,000 labourers for five years, and the remaining 200,000 would go starving—a position which evidently is as untenable." (p. 386)

The error in Böhm-Bawerk's thinking should be clear: in book VI and VII, he explains that the rate of interest can only be explained by time preference (what Böhm-Bawerk calls the "agio" theory of interest) or, put differently, by the practice of discounting of future goods against present goods. Then, Böhm-Bawerk mistakenly interprets this discounting of future income or value as a present supply of physical consumer goods, which are needed to sustain workers' wages over the entire "period of production." Savings and capital, thus, become a "subsistence fund," out of which the present consumption must be paid to workers *during* the period of production. This whole notion is, nevertheless, self-contradictory. First, Böhm-

Bawerk (1888) seems to mistake the fact that businesses pay to a certain extent[27] wages "up-front" (before the product that they produce is sold), which is therefore one of their operational expenditures, for the idea that businesses create a giant fund in which *all* future wage outlays are accumulated *ex ante*. Savings are, hence, a giant investment fund, represented "by the total sum of (…) wealth (exclusive of land)" (Böhm-Bawerk, 1888, p. 319) to cover wages until the final product is ready. This may be the case of some individual projects, but certainly not the case for an entire economic system. Moreover, there is absolutely no need to pre-finance a variable cost as labor expenses for the entire duration of a project. Since such variable costs are variable (workers can be laid off in the worst of scenarios), they can be perfectly financed with short-term savings, rather than long-term savings (in his own, example, Böhm-Bawerk uses a 10-year period). However, since Böhm-Bawerk (1888), unlike Fisher (1930), does not distinguish between short-term and long-term interest rates, he is unable to capture such details in his analysis. Any understanding of how businesses fund wages would show the absurdity of his idea of capital as a massive "subsistence fund." Much of the circulating capital businesses use to pay expenses, among which are wages, is obtained by intermediate sales of products. The idea that a production period of ten years would apply to all businesses in a society is ludicrous. No sane business would pre-finance years of wage outlays, nor would any entrepreneur abstain from beginning a long-term venture because he cannot borrow the entire sum necessary to cover wage expenses over the entire duration of his venture. What Böhm-Bawerk (1930) implies is akin to servicing a swimming pool that needs an *annual* replacement of its entire water content by filling it up with 15 times the required amount of water, given the fact that the swimming pool has an average life expectancy of 15 years. To reap the benefits of employing labor, there is no need to save up whatever an entrepreneur will pay out to a worker over time. Second, capital is the present value of the stream of income of society's productive assets. Hence, capital, in one way or another, simply represents the *value* of future production in present terms. This future production encompasses both final consumption goods *and* production goods. Therefore, the idea that capital is equal to a subsistence fund out of which

[27] Typically, monthly, or if a given project has absolutely *zero* revenues during its entire duration (which, in a certain sense, amounts to a zero-coupon bond), for any practical period until maturity. The latter are, however, odd examples that are generally not observed in practice. Within economic history, it would be an interesting avenue for research to pinpoint how Böhm-Bawerk's experience as finance minister led him to his concept of capital as a subsistence fund for workers.

“the supply of subsistence” can be provided should be completely rejected. Capital cannot be eaten; moreover, the period of production is already accounted for in the value of capital, not the other way around, as Böhm-Bawerk (1888) seems to imply. **It is not that *given* some value of capital, the period of production is determined. Rather, given a set of time preferences (and a set of Fisherian “investment opportunities”) that gives rise to a rate of interest, the period of production is determined, from which ultimately the *value* of capital is derived.** Pretending otherwise, would be akin to turning the world upside down. Hence, Böhm-Bawerk’s (1888) notion of capital as subsistence fund is sheer nonsense. Hayek (1941) would later call the idea of capital as a “subsistence fund” as “misleading in a number of ways.” (p. 85). “It is of course not a stock of actual means of subsistence, but only a stock of resources which can be turned into means of subsistence, i.e., into consumers’ goods.” (ibid). Yet, to be clear, investment is in no way dependent upon a produced stock of consumer goods. Investment requires hiring workers and paying wages, and such wages are (partly) converted into consumers’ goods. Yet some investment is demanded to provide for consumers’ goods. It is beyond me what the notion of capital as a subsistence (or wage) fund adds to our knowledge of how an economy truly works (or ought to work).

There have been some recent attempts, however, to reconcile the Böhm-Bawerkian concept of roundaboutness with modern-day finance theory. Cachanosky & Lewin (2014), for instance, distill Böhm-Bawerk’s concept of roundaboutness into two components: (1) *duration* (the weighted average term to maturity of the corresponding expected cash flows) and (2) *capital-intensiveness* (the amount of invested capital). But as Cachanosky & Lewin (2014) themselves conclude with regard to their “reformulation” of the Böhm-Bawerkian concept of roundaboutness: “This interpretation keeps the spirit of the concept but does away with the problems that arise in the traditional Hayek–Böhm-Bawerk approach discussed above. *The latter was a vain attempt to approximate a purely physical measure of ‘time taken’ or ‘quantity of capital invested,’* for example as the amount of homogeneous labor-time applied.” (p. 8) [emphasis mine]. Remarkably, Böhm-Bawerk already referred to production stages in his time structure of production[28] (visualized by his “concentric circles”) by calling them first, second, third, et cetera *maturity classes* or *more and less remote maturity classes*. The present-day financial concept of term to maturity, at a first

[28] For a discussion on Böhm-Bawerk’s circles and its more widely known successor, the Hayekian triangle, go to p. 163.

glance, appears very much in line with Böhm-Bawerk's terminology, despite our objections to Böhm-Bawerk's attempt to reconcile the subjective view of economics (specifically interest) with a physical or material theory of productivity.

Clark (1891), in contrast, argued in his *The Distribution of Wealth: A Theory of Wages, Interest and Profits* (1899), that whereas concrete capital goods might have a period of production (which we certainly do not deny and agree with), "true capital" has no such period of production. "[T]rue capital eliminates all roundaboutness or waiting between the beginning and end of a production process" (Cohen, 2008, p. 12). This idea is referred to as "synchronization," by which true capital permits the synchronization of production processes *without* the need for a subsistence fund. Clark (1899) uses a metaphor:

> "A tree will mature in twenty years; and the forest must be kept intact (…) or the supply of wood will fail. Each year we plant a row of trees along one side of the forest, and cut a row from the other. The planting and the cutting are, in a way, simultaneous. We do not burn today the tree that we plant today; but we do burn a tree, the consuming of which is made practicable by today's planting. (…) **The forest is a synchronizer of labor and its virtual fruit**. The fact that is of practical consequence is, that if we have once secured the permanent forest, we need do no waiting for fuel." (pp. 313-314) [emphasis mine]

While this is generally true, and precisely part of the critique on Böhm-Bawerk's notion of capital as a stored-up subsistence fund to cover the needs of laborers, the concept of duration *can* generally be applied to an entire economy. However, it is important to note that the concept of duration does *not* imply that capital, one way or another, equals a subsistence fund out of which wages are paid. While we have mentioned this fact above, it is important enough to repeat again. Hence, *duration* (roundaboutness) *is* important, but not in the sense of a subsistence fund and what maximum period of production such a subsistence fund might be able to cover. The notion of synchronization (Clark, 1899) is, to an extent, a refutation of any theory of capital that involves a subsistence or wage fund (Böhm-Bawerk, 1888).

In sum, the heart of the differences between Bohm-Bawerk and Clark did not boil down to, as generally is thought, the idea of capital as a "fund", but rather to the characteristics of this "fund" (Cohen, 2007). Whereas it is true that Böhm-Bawerk mostly refers to capital as "means of production" in his writings, capital becomes a homogenous subsistence fund denominated in money in his quantitative model of interest rate determination (Cohen, 2007). Indeed, Bohm-Bawerk begins describing an accumulated stock of wealth as "subsistence fund" for current laborers and even embarks upon calculating the optimal size of that stock of wealth to provide subsistence to a community (assuming a certain period of production), as we have discussed above. Hence, **Böhm-Bawerk does defend capital as a "fund,"** similar to Clark, although more as a "wage fund." Yet, on the other side, Clark ascribed permanence, asynchronization and perfect mobility to capital (Cohen, 2008), something Bohm-Bawerk fiercely opposed. According to Clark (1899), capital was equal to the market value of productive assets, including production goods *and* land, but excluding labor. While this idea of "true capital" does involve some degree of "*permanence*" (all other things equal, allowances are made to replace worn-out production goods), "*synchronization*" (outlays of one business are the income of the other business, which are subsequently used to pay, among other things, wages) and "*mobility*" (capital can, to a degree, be "taken out" one industry and invested in another), Clark (1899) largely exaggerated these characteristics.

Perhaps just as important, as mentioned above, Clark argued that "true" capital produces interest. Bohm-Bawerk took another approach. He distilled the phenomenon of interest to the concept of time preference (individuals always prefer some present consumption versus future consumption), while then trying to explain time preference with the "superior physical productivity" of production processes that require more time. In contrast, Clark viewed interest as a result of the productivity of capital. The former can be called a "less naïve" productivity theory of interest, whereas the latter can be called the "naïve" productivity theory of interest.

The trade-off between present and future consumption was not something inherent to interest, according to Clark. It was, to be sure, the initial source of capital, but not necessary to sustain the capital fund. This capital fund is "self-sustainable," because capital generates income and part of that income is automatically put aside to maintain the initial capital. In Clark's (1899) own words: "The laws of matter, in short, make capital productive. (…) Paying interest is buying the product of capital, as paying wages is buying the product of labor. *The power of capital to create the product is, then, the basis of interest*." (p. 135). However, as Taussig (1908) correctly points out and as we have repeated earlier, any income derived from capital can be consumed instead of reinvested, rendering Clark's argument null. Put differently, every decision to not consume the (what Clark calls the "initial") capital requires abstinence.

In modern finance terms, the act of *not* consuming cash inflows and *not* reducing capital expenditures (CapEx) to zero requires subsequent acts of abstinence.

There is thus nothing permanent about capital. Nevertheless, we should admit that Clark is, effectively, describing an equilibrium state. The return on capital should, in this equilibrium, equal the value of time (pure time preference) *after* depreciation and amortization. It would therefore not increase nor decrease the (total value of the) capital. Or as Clark (1899) himself writes: “Each year we plant a row of trees along one side of the forest, and cut a row from the other. The planting and the cutting are, in a way, simultaneous” (p. 314). In this sense, in a stationary state or equilibrium, capital is indeed synchronous. No waiting is necessary; the consumption and replenishment of the capital are simultaneous acts.

This is important to emphasize, because Clark in effect says interest is equal to the return on capital. Böhm-Bawerk says interest is, among other things, part of the return on capital[29]. If and when in an economy all profit opportunities have been arbitraged away, the (gross) return on capital would equal the rate of interest (time preference). However, Clark fails to recognize this, and seems to assume that returns on capital, besides the time value of money, are as permanent as, according to him, capital is.

Nevertheless, Clark (1899) makes an important point here, since Böhm-Bawerk’s physical period of production is essentially backward-looking (the time it takes or took to create a specific capital good), not forward-looking, which is a deviation from the subjectivist economic view. Yet Clark himself forgets the fact that the value of capital is essentially forward-looking and involves discounting some future net cash flows to the present with a discount rate that represents the value of time (or the rate at which an individual, or society, discounts future consumption).

Indeed, in this sense, Böhm-Bawerk’s critique on Clark’s capital theory is devastatingly pertinent. He objects to Clark’s concept of “true capital” because it completely eliminates the time element. As a consequence, Böhm-Bawerk refers to Clark’s concept of “true capital” as a “mythology of capital.” According to Cohen (2007), Clark makes no allowance for time because he describes and explains a purely stationary state, that is, an equilibrium state.

---

[29] Please note that we refer here to the gross or nominal return on capital. In other words, a pure theory of capital would not include interest in the (pure) return on capital.

Cohen (2007), therefore, concludes that Böhm-Bawerk really wanted to refer to the concept of "true capital" as a "mythology of static equilibrium." However, if this is truly the case, then Clark failed to account for entrepreneurs bidding up the prices of assets so that they reflect their future rents. If Clark wanted to use equilibrium to explain the dynamic forces behind capital that move it toward equilibrium, his omission amounts to a capital sin. As Mises (1940) explains, there would be no equilibrium if the prices paid for capital do not include or incorporate the future rents that this capital is able to generate, since arbitrage opportunities would continue to exist.

Rather than the problem being that Clark (e.g., Clark, 1888; Clark, 1891) uses a stationary state to construct his capital theory, the rub rather lies in the fact that Clark tries to explain the value of capital as a result of its rents, that is, the return on capital. Moreover, since he defends capital as a homogeneous, self-sustaining fund of "value," all production seems synchronous and instantaneous. However, this is not the core of Clark's mistake. **Rather, Clark's mistake lies in the fact that he tries to explain interest by referring to the "earnings of capital," instead of by referring to the value of time**. Böhm-Bawerk's accurate critique on Clark's concept of "true capital" centers on the complete omission of time in his interest theory. Indeed, Clark never attempts to explain what the origin of the "market value" of the various assets (or capital goods) is. He would instantly recognize that this market value can only be derived from the utility of these goods, which would simply mean the present value of their ability to earn or generate future income. **To be able to express that future income in present-day terms, you would need a rate of interest that equals the value of time**. That is, the rate of interest, far from being a return on capital, would be a result of time preference (Gunning, 2005).

Huerta de Soto (2006) summarizes the above arguments in the following way: "Clark considers production and consumption to be *simultaneous*. In his view production processes are not comprised of stages, nor is there a need to wait any length of time before obtaining the results of production processes. Clark regards capital as a permanent fund which "automatically" generates a productivity in the form of interest. **According to Clark, the larger this social fund of capital, the lower the interest.** The phenomenon of time preference in no way influences interest in his model." (p. 514) [emphasis is mine]. The complete omission of time taunts Clark's work on capital theory, even though Böhm-Bawerk's treatment of time is not without flaws.

2.2.1 Perfect Mobility of Capital, Degrees of Liquidity

Clark (1899), completely confused by the veil of money in his concept of true capital, goes on to define capital as "perfectly mobile" (p. 118). Clark (1895) explains by example: "A whaling ship cannot be made to spin cotton; but capital has, in fact, transferred itself from the whale fishery of New England to cotton spinning. Ships were allowed to decay, and mills were built in place of them. (…) You can get "money" out of one industry, and put it into another." (p. 265). He refers to the fact that the *value* of his "true capital" remains constant, as income (in the form of interest) is invested in other lines of production while other lines become worthless.

Clark is partly right. That is, if we have goods with higher degrees of liquidity[30] (both Clark and Smith would classify such goods as *circulating capital*[31]), they can be easily moved in and out industries. The cost of liquidating positions should be relatively low since "time to liquidity" is short. The difficulty is when part of the capital is invested in goods with lower degrees of liquidity. Lachmann (1956) would refer to such differences in liquidity as differences in the "specificity" of capital.

Clark (1895) goes on by arguing:

> "[Changes] in the amount of capital itself, and not a lengthening or shortening of productive periods, are the causes that affect the rate of interest. Make the social fund larger, and you make the rate of interest smaller." (p. 277)

This is a largely a fallacy of composition. Clark's single-minded focus on capital as an immense social fund prevents him from seeing that his "true" capital has an important time element; his "capital" is invested in different lines of production with inherently different maturities or durations (Adam Smith [1776] distinguished between fixed and circulating capital). These maturities must match the maturities of producers; otherwise a part of savings is squandered into providing goods at the wrong temporal point. On the margin (that is, new

---

[30] We discuss the historical debates on the definition and meaning of liquidity at p. 118.

[31] In modern times, we refer to circulating capital as working capital. Inventories are often assumed to be working capital for the mere reason that they are inventories, yet it makes intuitive sense to distinguish between noncurrent inventories and current inventories (which are expected to be converted into cash within a year).

investment projects), the average duration of a society’s productive efforts can be increased or decreased, depending on what the rate of interest signals to (financial) entrepreneurs.

Put differently, capital arises from renouncing consumption, as Turgot argued. If we would stop renouncing consumption, no capital would exist. We would consume all the income and, eventually all the capital, because we have no intent to supply for the future, Hence, as we save, we defer consumption, which gives rise to the concept of capital. However, we defer consumption for a certain period of time. For instance, we can remain as liquid as possible by opting for financial instruments such as currency or deposits with a zero maturity, as opposed to investing it in an instrument with a ten-year maturity. We can also opt to remain liquid at zero maturity, and roll over our zero maturity instruments. But doing so comes at a cost. The opportunity cost is a longer maturity instrument that has a higher (expected) yield. Interest rates for certain maturities go up when the demand for liquidity becomes excessive.

That would also lead to the conclusion that the theory of time preference is not enough to explain the phenomenon of interest, since it only explains the supply side. Both entrepreneurs and households form part of the demand side. And the entrepreneur’s demand depends on his assessment of the expected return on capital, which depends on expected future consumption. In modern-finance terms, if the return on invested capital exceeds the cost of capital, entrepreneurs have an incentive to step in as willing payers of interest and demand investable funds. This is one of the fundamental weaknesses of the explanation of interest of Böhm-Bawerk and later Mises (1940).

John Bates Clark (1895) briefly touches upon this point:

> “If the capitalist's forward glance, at the moment at which he is deciding how much capital to save, has the effect of fixing the amount of capital and thereby influencing the rate of interest, that rate is not what it would be if the capitalist were comparing personal gains alike in kind and in amount. Interest is not, as a mathematical fact, an equivalent offset for the sacrifice entailed by mere delay.” (p. 262)

Something that characterizes the literature up until the great Böhm-Bawerk/Clark debate, still haunts economics today: the simultaneous use of two different definitions of capital throughout economic texts (e.g., Mises, 1949). On the one hand, capital refers to the whole of (heterogeneous) physical production goods, whereas on the other hand it refers to the monetary value of these production goods. In other words, is capital the machines, tools and equipment, or the net present value that such production goods embody? We will see that this is largely a question of "scale." The contribution of an individual productive asset can just as much be expressed as the net present value (NPV) of all its net future cash flows (or contribution to net cash flow) as a set of productive assets (e.g., a business). Hence, capital equals the NPV of all productive assets while every individual asset equals its individual NPV. Since capital is much more a tool of economic calculation rather than a class of physical goods, the very same concept can be applied to large or small clusters (up till the individual level). None of both recognized this fact, since both more or less adhered to the classical economists' fallacy of distinguishing between land, labor and capital (or labor and capital in the case of J.B. Clark).

2.2.2 Roundaboutness, Periods of Production and Böhm-Bawerk's "Physical Delusion"

We will now in greater detail look at Böhm-Bawerk's theory of the period of production, Böhm-Bawerk referred to the "increased productivity" of more "roundabout" (time-using) methods of production. Interest rates indicate time preferences. If time preferences would not exist, production would always strive to maximize the "roundaboutness" of production. He attempts to prove that production processes with a "longer production period" are more "roundabout" and therefore more productive than shorter, less roundabout processes. We have already criticized this Böhm-Bawerkian notion above, by showing that his own examples are self-contradictory and that the whole idea of increased *physical* productivity (*compared* to value productivity) is misleading. Whereas including time in our analysis is key, Böhm-Bawerk erred on the side of attributing higher physical productivity related to waiting. We will now turn to what Clark (1895) specifically had to say about Böhm-Bawerk's idea of a period of production and compare both points of view.

Clark (1895) makes a valid objection to this theory of the period of production:

> "The only practicable lengthening of the interval is really a multiplying of the goods existing within the interval. The time that figures in Professor v. Böhm-Bawerk's analysis translates itself into quantity [of capital goods available]. (...) Lengthen the periods of production without increasing the amount of capital, and the law will not hold true." (p. 266)

Clark (1895) criticizes here the idea that capital goods "ripen" into consumer goods over time. In a certain sense, combining capital goods helps us to produce consumer goods. Yet, these capital goods work all at the same time to crank out consumer goods. In other words, car manufacturers produce both the cars of today *and* the cars of tomorrow. Hence, it is not so much the interval (especially when we look at the economy as a whole rather than a single venture), but the quantity of capital goods used in that interval. As counterfactual, Clark (1895) argues that merely taking larger intervals without additional capital goods would not lead to greater production. As such, the idea of a (mechanical) period of production as a relevant economic concept goes against the idea of *value creation*. Implicitly, Böhm-Bawerk adheres to a physical conception of production when he argues that a larger period of production allows for more roundabout methods of production and therefore higher returns. If it is true that value is subjective, then the value of production can either exceed or fall short against its costs – which represent an opportunity cost, and therefore *per definition* less roundabout and shorter periods of production *can* add more economic value than the product of more roundabout and longer periods of production. Clark's (1895) synchronicity collides with Böhm-Bawerk's production intervals.

2.2.3 Time Preference, Abstinence and Clark's Error

As we have seen above, Clark claims that abstinence (foregoing present consumption) was a prerequisite for capital to arise. However, since the moment capital was "created," abstinence is no longer necessary. The return on capital suffices to maintain the capital fund. Hence, time is of no meaning: all production is synchronous. Moreover, production does not require future abstinence. As Clark (1895) explains:

> "If he has let the surplus earnings of the machine accumulate, as a sinking fund for buying another, the time when he draws this fund from the bank means no new enjoyment for him. It is a self-renewing time for the instrument. Its accumulated earnings take the shape of a successor in the definite series of instruments that, in its entirety, constitutes a bit of permanent capital. The only thing the owner looks forward to with anticipation of enjoyment is the unending series of other earnings that come to him, not from this machine only, but from the whole series; and that succession of earnings constitutes permanent interest." (p. 273)

Clark completely rules out the possibility of an owner of means of production – or a combination of means of production, that is, a business – *not* replacing the worn-out means of production. Sure, owners usually make allowance for depreciation and amortization, and consume the income, not the capital. But that this rule usually applies, does not mean that owners necessarily refrain from consuming the entire capital. An owner can simply consume the allowances he made for depreciation and amortization and *not* buy a replacement for his worn-out production good(s). In fact, in practice, businesses are liquidated all the time. Capital is returned to its owners, the shareholders, constantly. Also, shifts in consumer preferences might induce a decrease in capital, which in Clark's theoretical model is impossible. Clark seems oblivious to this fact and persists in his fallacious critique of abstinence (that is, a crude formulation of the later theory of time preferences).

Whether or not capital is consumed is due to time preferences, the willingness to abstain from consumption (in short, abstinence) of the owner or owners in question. If a business is liquidated and the proceeds are not reinvested into other businesses, but consumed, it reflects beyond any doubt a change in time preference, with greater preference for present consumption. Ascribing permanence to capital only makes sense if we assume that no consumption of capital could occur. Despite Clark's objections against the "abstinence theory of interest," Schumpeter (1954) concluded that only few theories of interest came out alive after the first round of controversies, and the "abstinence theory" was one of them, as we will see in the next controversy on capital regarding interest.

2.2.4 The Verdict on Clark versus Böhm-Bawerk: A Tie

Böhm-Bawerk defended with vigor his theory of time preference as the basis of interest. The return on capital, therefore, does not equal interest, in contrast to what Clark posits. Clark, in effect, held a "naïve" productivity theory of interest. However, in explaining the *origin* of time preference, Böhm-Bawerk fell in the same trap as Clark. In trying to elucidate the motives behind the time preferences of "buyers of capital," Böhm-Bawerk argued that production methods with longer periods of production are more roundabout or technically superior to production methods with shorter periods of production.

Besides omitting the origins of the demand for capital (Fisher's "profit opportunities"), Böhm-Bawerk made even greater errors, many of which remain present in modern economic science. The errors committed by Böhm-Bawerk and many present-day economists are a result of viewing capital as a complex of capital goods, which arose by combining labor with "original" factors of production, such as land and raw materials. This, in turn, led Böhm-Bawerk astray in his attempt to "classify" factors of productions according to their physical origin. The classical legacy of distinguishing between land, labor and capital led both Böhm-Bawerk and J.B. Clark away from a coherent subjectivist theory of capital and interest, even though J.B. Clark (1899) partly recognized that land was no different from other capital goods when viewed from the notion of capital simply as *net worth*. This first controversy between Clark and Böhm-Bawerk was, therefore, a step backward in the subjective revolution of economic science. Fortunately, Menger and Fisher took it upon themselves to defend the subjective view of capital.

2.3 Enter Carl Menger and His Forgotten Contribution on Capital Theory

In 1888, Carl Menger published an article in reaction to Böhm-Bawerk, titled *Zur Theorie des Kapitals* (*Toward a Theory of Capital*). Carl Menger's intervention should not be taken lightly. According to Schumpeter (1954), Menger told him that the "time will come when people will realise that Böhm-Bawerk's theory [of capital and interest] is one of the greatest errors ever committed." Friedrich Hayek, who joined the debates on capital theory in a later stage[32], confirmed that Menger "did not quite agree" with Böhm-Bawerk's definition of capital. Hayek would later write in an introduction to Carl Menger's *Principles of Economics* (1871):

> "It is pretty certain that we owe this article [Menger, 1888] to the fact that Menger did not quite agree with the definition of the term capital which was implied in the first, historical part of Böhm-Bawerk's *Capital and Interest*. The discussion is not polemical. Böhm-Bawerk's book is mentioned only to commend it. But its main aim is clearly to rehabilitate the abstract concept of capital as the money value of the property devoted to acquisitive purposes against the Smithian concept of the "produced means of production." His main argument that the distinction of the historical origin of a commodity is irrelevant from an economic point of view, as well as his emphasis on the necessity of clearly distinguishing between the rent obtained from already existing instruments of production and interest proper, refer to points which, even today, have not yet received quite the attention they deserve." (p. 27)

Unfortunately, as we will see in the second round of controversies on capital, Hayek himself appeared to have forgotten Menger's lessons.

[32] For the Hayek-Knight-Kaldor controversy, see p. 149.

Schumpeter (1954) himself, however, thought that the debates on capital were mostly a waste of time. He saw the first round of debates on capital as, to a great extent, mere verbal gymnastics on unimportant, unpractical and time-wasting semantics. "Throughout the period, economists of all countries displayed a propensity to adhere to the deplorable 'method' of trying to solve problems by means of hunting for the meaning of words. There was a controversy about the concept of capital, or rather there were several of them, in particular one in which the chief figure was Böhm-Bawerk and another in which the chief figure was Irving Fisher." (p. 865).

Yet his disdain for definitions was quite misplaced. Since it was the very confusion regarding the *nature* of capital, and a coherent understanding as to what capital is, that led to the widespread modern use of economic growth models that incorporate and continue incorporating outdated notions on capital. Instead of discarding the model, economists (e.g., Phelps, 1961; Solow, 1963) continue adding variables and variations to the model, very much akin to an oil driller who keeps drilling deeper and deeper without much of a result. It seems as if a greater use of this "deplorable method" would have avoided present-day headaches and model errors. Fortunately, Menger (1888) did not appear to agree with Schumpeter.

2.3.1 Menger Corrects Menger

With his 1888-contribution, Menger contradicts his earlier work on capital theory, principally embodied in his *Principles of Economics* (1871). Unfortunately, many modern-day reviews of Mengerian economics are based on a reading of his *Principles of Economics* and not on his later work on capital theory (Menger, 1888). As a result, some authors (e.g., Horwitz, 2011; Huerta de Soto, 2006; Garrison, 1990; Skousen, 2007) wrongly conclude that Menger would define capital as "all of the goods of higher orders" (Horwitz, 2011, p. 11) with an emphasis on its "heterogeneity"[33]. This historical reading is wrong and erroneously links Menger's name to Böhm-Bawerk's legacy. Braun (2013) confirms this thesis and writes that the "common interpretation of Carl Menger's take on capital theory rests upon a few sentences in his *Principles of Economics*" (p. 1). He concludes that: "Menger would have opposed all attempts to define capital as a heterogeneous structure of higher-order goods – a definition that is associated with his name today." (p. 1). It is unfortunate that there have been attempts

[33] The "heterogeneity" of the capital structure was something that was later elaborated by, for instance, Lachmann (1956).

(e.g., Kirzner, 1996; Garrison, 1990) to contrast Menger's views on capital with Böhm-Bawerk's views, solely based on Menger's *Principles of Economics* while ignoring his later contributions. To Braun (2013), this is a grave mistake:

> "For Menger, once he had dealt with the issue in more depth, capital did not consist in the heterogeneous structure of producer goods, as is usually maintained. On the contrary, Menger was of the opinion that capital must be interpreted in the way common parlance does, namely as a homogeneous concept depicting sums of money on ordinary business accounts. In fact, he vigorously opposed all theories that dissented from this ordinary business view on capital, including the one that is commonly imputed to him." (p. 2).

Indeed, we can see how Menger's own ideas about capital have changed since the publication of his *Principles of Economics* (1871), when in an appendix he tries to justify the difference between capital on the one hand and wealth on the other hand. It seems as if Menger, over time, noticed the same inconsistency in his own work. In 1871, Menger wrote the following:

> "The most important difference between capital on the one hand and items of wealth that yield an income (land, buildings, etc.) on the other is that the later are *concrete* durable goods whose services themselves have both goods character and economic character, whereas capital represents, directly or indirectly, a *combination* of economic goods of higher order (i.e., complementary quantities of these goods) whose services also have economic character and therefore yield income, but whose productivity is of an essentially different nature than that of durable wealth that is not capital." (Menger, 1871, p. 304)

This distinction that Menger held in 1871 is, however, untenable. Menger apparently thought (at that point in time) that some individual assets (higher-order goods) should not be considered capital *despite* being able to yield income, such as office buildings, while only *combinations* of individual assets (higher-order goods) should truly be considered "capital". "Capital," hence, becomes a subset of the broader concept of "wealth." Nevertheless, in the case of a business owner who exclusively dedicates himself to renting out office space, the building actually *is* a combination of various "higher-order goods," for instance when we take into account the maintenance equipment, decoration, lightning, legal and permit fees, et cetera, which economists would customarily call supplementary goods. It would be completely arbitrary, if we now consider a company not outsourcing its office needs, to suddenly treat it in a completely different manner. When we talk about a "concrete durable good" according to modern accounting practices, we refer not only to the construction or purchase of the good, but to any expenditure that is considered both ordinary and necessary to get the good in condition for its intended use. What use would an office building without the required permits and thus the ability to use it have? The "wealth" items Menger (1871) refers to, do not exist in a vacuum. Hence, any asset or given combination of assets should be considered capital. Luckily, Menger (1888) seems to correct himself after, what seems, fierce debates with Böhm-Bawerk on the nature and consequences of capital.

2.3.2 Capital as Financial Net Worth

Indeed, Menger (1888) moved toward a definition of capital as the *monetary value* of the productive assets from which we expect an income ("active assets"). This would then include labor, for instance, but exclude stocks of consumer goods (including, usually, durable consumer goods), since such goods are not "productive" and do not "yield income." His emphasis on "monetary value" is especially important, as Menger (1888) argues that: "[T]he physical nature of the goods of which the assets of the for-profit economy consist fade from the spotlight and their "monetary value" comes to the fore of our economic observation and considerations: in that case, active assets – whatever the physical nature of their components – constitute monetary values for the purpose of economic calculation, and more specifically sums of money devoted to income generation" (p. 29). Indeed, Braun concludes that: "Assets [or production goods] become capital only in so far as they are homogeneously expressed in terms of money. That they are heterogeneous is of importance only from a technical point of view, that is, for concrete production processes. But this has nothing to do with the way the term "capital" has to be understood." (p. 20)

F.A. Hayek notes how Menger (1871) changed his definition of capital after *Principles*, striving "clearly to rehabilitate the abstract concept of capital as the money value of the property devoted to acquisitive purposes against **the Smithian concept of the "produced means of production[34]**." (p. 27). "His main argument [was] that the distinction of the

[34] We consider this our key takeaway from our review of the first round of controversies on capital and have, therefore, chosen to repeat and emphasize Hayek's reference to the "Smithian" concept of capital as "produced" means of production.

historical origin of a commodity is irrelevant from an economic point of view (…)" (p. 28). Menger (1888) himself confirms Hayek's reading:

> "For the size of the net yield generated by productive assets, it is irrelevant whether the goods devoted to income generation constitute products of labor or natural objects or (assuming an identical quality and quantity of the respective goods) whether much or little labor has been used on them. Whether fruits or timber devoted to revenue generation have grown naturally and only become economic goods as a consequence of relative scarcity or whether some amount of labor had to be employed in order to cultivate them, whether the land had always been naturally fertile or whether it had to be ameliorated in a labor-intensive process turns out to be inessential for the *yield* and the *capital value* of said goods under the above condition of identical quality and quantity." (p. 17)

Moreover, although implicit in Menger's work, Menger (1888) appears to refer to capital as the monetary value of a combination of income-generating productive ("active") assets *net* of debt. Menger does not explicitly explains it as such, but it is implied in his (almost desperate) attempt to separate money, monetary assets and other "interest-bearing" instruments from his theory of capital. The reason is simple: it would imply a type of double-counting of assets, since such debts (including monetary assets such as commercial bank deposits) are simply (direct or indirect) claims to productive assets.

Note also that Menger (1888), closely following his subjectivist theory of value, calls an asset "productive" (or "active") when it is able to generate income (yield) and/or command an exchange value when being bought and sold on the market. In other words, to Menger (1888), the material origin or nature of a production good is no determinant in whether something is capital or not. Nor is it an entrepreneur's completely arbitrary subjective appraisal of an asset. What determines whether an asset is "productive" and thus capital, is the fact that it generates income (i.e., somebody on the market is willing to pay a price for its services) or can be bought and sold on the market (i.e., somebody on the market is willing to pay a price for the asset). Both cases are two sides of the same medallion, however, as nobody would be willing to pay a price for an asset that does not generate income (or if the buyer at least would not *expect* the asset to generate future income), besides some potential subjective consumption value.

This subjectivity is a revelation compared to Böhm-Bawerk's strictly material understanding of capital. However, yield is defined as purely monetary. Yield (or profit), however, can also be understood in a broader sense, as the psychic appraisal of an end exceeds the psychic appraisal of the cost of reaching that end (Huerta de Soto, 2010).

With regard to the *dual nature of capital*, which we have discussed earlier[35], Menger (1888) notes that practitioners clearly distinguish between the underlying productive assets and "capital.": "By themselves [factories, machines, etc.], he [the entrepreneur] calls these goods productive assets, or assets in general, but not "capital." (p. 33). Moreover, the term interest, as used by businessmen, is interest on capital, and not the yields or rents that the underlying assets generate. Indeed, Menger (1888) sees *this* duality as the most pressing problem of economic science:

> "**The strict difference between (productive) assets and capital**, as defined by everyday life, **is of the greatest importance especially for economic theory and its misapprehension constitutes a main cause of the backward state of the doctrine of active asset yield [rents and yields]**. Only the confusion of these two important categories of economic life could have led to the misconceptions of a great number of revisers of our science who put a mere theory of interest on capital where a universal theory of asset yields belonged and who believed to have solved the much broader problem of the explanation of asset yields when in fact, they had merely tried to interpret the phenomenon of interest on effective capitals." (p. 34)

Instead, Menger (1888) ultimately defines capital not as a heterogenous structure of higher-order goods, but rather as a homogeneous financial or accounting concept "that only makes sense in economic calculation" (Braun, 2013, p. 3). Apparently, Schumpeter (1954) seems to agree, since this idea "may be expressed by saying that our capital-goods market is really a market of streams of perpetual net revenues, from which standpoint all capital goods are on the same footing **irrespective of their physical shapes** (p. 983) [emphasis mine]. Menger's financial or accounting approach to capital was also noted by Endres (1997):

> "From an accounting perspective, capital is a fund embodied in business assets. The fund pertains to an anticipated value of saleable goods produced by these assets as

[35] On the dual nature of capital as discussed in the Böhm-Bawerk/Clark controversy, see p. 68.

they are combined in a particular business. Menger's entrepreneurs create capital in the present on the basis of an *expected* value of the output to be forthcoming at a future date (…). From the entrepreneur's point of view, capital is an accounting concept—a fund of prospective values which can be estimated in the present." (p. 169).

Moreover, according to Endres (1997), Menger "readily accepted that capital was a 'fund' in the financial, acquisitive sense, not a list of items constituting possible instruments of production. Capital was described in terms of the money value (*Geldwert*) of the present worth of business assets productively combined in a specific process aimed at producing exchangeable goods for an income. The filter of the entrepreneur's mind was critical for turning mere assets into actual productive, income-generating capital. The technical nature of assets was inconsequential. Capital could therefore include various legal claims, intangibles, and especially goodwill and 'business connections'" (p. 170).

With regard to labor, Menger (1888) says that there is no reason to deny labor "the status of capital *per se*" as long as "man's labor power is considered to be a kind of asset or part of his active assets" (p. 3). Menger (1888) appears to leave the door wide open to embracing human capital under his definition of capital. He, moreover, warns for the classical Smithian distinction between capital and labor. Menger (1888) writes: "Products of labor are mostly economic goods and tend to have a market value, yet not because labor or products of labor have been used on them, but because economic actors tend to only use labor or products of labor on such goods that are likely to gain the status of economic goods and [therefore] a market value" (p. 17). Smith (1776) and the classical economists were all wrong about the economic value of labor. Labor does not *give* value to goods, but labor tends only to be used by entrepreneurs on goods that are expected to *create* economic value. In a rather clear analogy, Menger (1888) compares the Smithian confusion of labor with the idea that having a train ticket causes someone's desire to travel, when it is really the other way around. The desire to travel *causes* an individual to *acquire* a train ticket. Labor does not give value in of itself, but it tends to be used on the production of goods that are expected to be valuable in the market place (that is, consumers are willing to pay for such goods).

Hence, in stark contrast to Böhm-Bawerk (1899), Menger (1888) reestablishes consumer sovereignty as the ultimate determinant of the value of capital. Menger (1888) criticizes Böhm-Bawerk (despite avoiding any direct reference to his friend and colleague) by arguing that Böhm-Bawerk's definition of capital (as "produced means of production") is an error: "In whatever way the term "means of production" is understood, the identification of capital with means of production has been shown to be terminological arbitrariness and a factual error." (Menger, 1888, p. 9). Not only did Menger correct his *own* work on capital, he also took direct aim at Böhm-Bawerk's theory of capital.

2.3.3 A Foretaste of the Menger/Böhm-Bawerk Divide

Curiously, even though Menger had not reached these conclusions in his *Principles of Economics*, an earlier disagreement with Böhm-Bawerk in 1881 on the treatment of goodwill gives us a foretaste of their future divide on capital. Böhm-Bawerk, as Endres (1997) explains, argued that "present material producers' goods give concrete 'assurance' of future returns, (…) whereas the basis upon which goodwill is capitalized is *prima facie* more tenuous." Goodwill, according to Böhm-Bawerk (1881), is imaginary and must be "[r]uled out of the list of genuine [capital] goods" (p. 127). He called the advantages derived from goodwill "too indefinite, too difficult to observe and, above all, (…) too tenuous and incomplete (p. 125).

Mises (1940) also took notice of the discussion on goodwill and concludes that, even if such goodwill is a product of complete human imagination on the part of consumers, it still counts as much as any other material good as capital. Menger (1871) concludes that goodwill should not only count toward someone's individual capital, but is also to be counted on a societal level. Goodwill, according to Menger (1871), reflects almost indispensable trading connections in organized markets (and is, therefore, a "relationship good"). In modern-day terms, we could say that goodwill partially reflects the effort to provide valuable information to (prospective) buyers and thus brings advantages to society at large.

Endres (1997) agrees that this difference of opinion regarding the nature of goodwill was a foretaste of more fundamental differences between Menger's (subjectivist) capital theory and Böhm-Bawerk's (material) capital theory. In his own words: "The divergence between Menger and Böhm-Bawerk on the goods character of rights and relations extends correspondingly to their competing concepts of capital. Böhm-Bawerk developed a capital concept that stressed materiality." (Endres, 1997, p. 162).

On another note, Kirzner (1996) criticized Böhm-Bawerk for losing sight of the "Mengerian legacy." However, Kirzner does not revolve his critique on Böhm-Bawerk around Menger's financial capital (in fact, Kirzner mentions Menger's notion of capital goods of different orders as one of his key contributions), but instead concentrates his critique on the fact that Böhm-Bawerk was "unable to avoid permitting the physical productivity of roundaboutness (or of time, or of waiting) to play an independently explanatory role in accounting for the phenomenon of market interest." (p. 4).

Hence, Kirzner (1996) criticizes Böhm-Bawerk for not (completely) paying heed to the pure time preference theory of interest by referring to the, in a physical or material sense, higher productivity of more roundabout production methods in order to explain the market rate of interest. Kirzner (1996) then attempts to demonstrate that Böhm-Bawerk, as a consequence of his physical rather than subjective theory of interest, has essentially put economics on a wrong footing. In short, according to Kirzner (1996), the Mengerian insight consists of the fact that, ultimately, the subjective valuations of consumers are what determine and explain both capital and interest. There is no need for any reference to the superior physical or material productivity of more roundabout production methods. Böhm-Bawerk deviates from the Mengerian approach by trying to explain capital and interest with something other than consumer valuations, that is, physical productivity.

2.3.4 A Brief Summary of Menger versus Böhm-Bawerk

In sum, Menger (1888) was very critical of his Austrian colleague on multiple accounts and at the same time made several important contributions to capital theory. According to Menger (1888):

(1) The entrepreneur determines what is capital and what is not

The entrepreneur determines what goods (that is, products and services) are valuable in production and thus become capital. In other words, what is capital and what is not, depends on the *subject* (or, more specifically, the subjective appreciation of a physical asset by a profit-seeking entrepreneur). Hence, anything can become capital, including intangible assets, rights or claims. Therefore, the *object* or the objective nature of the asset is of no importance. The classical distinction between labor (wages), land (rent) and capital (interest) became untenable in the view of Menger (1888) and was thus fiercely opposed. Nevertheless, this divide between "original" factors of production and "produced" factors of production permeated Böhm-Bawerk's (1888) work on capital theory.

(2) Capital is a "fund" in terms of money of underlying productive assets

Capital, according to Menger (1888), is essentially the financial *net worth* of an entrepreneur, who has a combination of productive assets (higher-order goods) under his command. It is essentially an outcome of the entrepreneurial process of "economic calculation" or, in more modern terms, "capital accounting." On the other hand, Böhm-Bawerk criticized Clark for his view of capital as a "fund," greatly emphasizing that capital is a "complex" of heterogenous and material "produced means of production."

(3) The subjective valuations of consumers determine the composition of capital, never the physical or material "superiority" of certain production methods

According to Böhm-Bawerk (1888), longer production processes are more "roundabout" and therefore physically more productive than shorter production processes. Hence, investment (capital budgeting) decisions are partly driven by the greater physical productivity of one (more roundabout) production process over another. Menger (1888; 1871), in contrast, emphasizes that the profitability of production processes ultimately and exclusively depend on the subjective valuations of consumers: that is, on subjective time preferences and marginal utility of consumers. Means of production ("productive assets") are ultimately worth what they are worth because consumers are willing to pay for their product (so-called first-order goods) at different points in time. Means of production are not worth less or more due to some inherent physical superiority in their productivity. Capital is a tool of economic calculation, according to Menger (1888), which helps entrepreneurs to project and account for profits and losses, which are ultimately derived from consumers' spending decisions, rather than a physical composition of "produced" means of production, separate from labor and other original factors of production, or a "subsistence fund" in which wages payable are supposed to be saved up and which *physically* determines and bounds what period of production is optimal from a material notion of production.

At this point it is important to notice that Menger (1888) effectively understood and pointed out one of the proposed key contributions of our work, which we call *q*. In one of his footnotes, Menger (1888) writes:

> "Practitioners in the field of business simply call their possessions of goods "assets" as long as they are devoted to income generation; when stressing a contrast to their

consumption goods (that is, assets in the broader definition of the term!), they call them "active" or "productive assets"; the monetary values representing these goods, however, they call "capital" (especially with regard to the yield of productive assets expressed in monetary values). Stocks of resources, a factory, a warehouse etc. are "assets" - they are, however, not capital *per se,* but only with regard to the monetary values they represent." (p. 41)

Here, Menger (1888) recognizes that capital is the monetary value of underlying "productive asset(s)." Put differently, Menger (1888) understands that there are two sides to the same capital coin. Essentially, capital is a financial claim (hence *net worth*) on some or various underlying "productive" asset(s). The financial claim (*capital*) has a monetary value, but the (individual) underlying productive assets are also bought and sold on the market. The former represents a "price," the latter a "replacement cost/value." *Capital* is capital to a specific entrepreneur, who usually has various or a combination of "productive assets" under his control. We can either buy capital (as a financial claim on underlying productive assets) or buy the productive assets themselves. This is what our later discussion on *q* revolves around.

We will follow more or less the same terminology in the course of our literature review. That is, "productive assets" refer to the underlying assets which are represented in terms of money as "capital."

2.4 Wicksell's Natural Rate of Interest

One of the more important lessons of Knut Wicksell (1898) has been largely forgotten: the inseparable link between capital and money. In fact, Wicksell is often remembered for his narrowly understood "natural rate" theory of interest, rather than his broader elucidation on money, financial intermediation and prices. We will discuss both more elaborately below.

2.4.1 Wicksell's Connection between Money and Capital

Especially interesting is Wicksell's critique on the quantity theory of money (Wicksell, 1898). Wicksell (1898) argues that the line that is normally drawn in theory between money and credit is not as sharp in actual practice: "In actual fact the border line between money in this sense and true instruments of credit (ordinary book credit, bills, cheques, etc.) is extremely vague; and over a wide range one can be substituted for the other (…)" (p. 42). This results in the quantity theory of money being valid in theory, according to Wicksell, **as**

long as all its assumptions are met, despite the fact that this is practically never the case. Then Wicksell (1898) argues, quite remarkably, that all money is credit. Even metallic or commodity money, says Wicksell (1898), completely derives its value from the *belief* of a holder that he or she is able to obtain a certain amount of goods with it. It boils down to a "question of degree" (p. 49). He concludes by making a sharp observation: even at the time, the value of silver fell below the value of the notes that were backed by silver[36].

After discarding earlier versions of the quantity version of money, Wicksell shows that the demand for money is important, citing Mill for his work on the "velocity of money", which is in fact the corollary of the average holding period of a monetary unit or portfolio demand of money (see Mises, 1949). It is not just quantity or supply, but rather the demand for money that matters.

This demand, however, largely depends on the circumstances on capital markets. And as Wicksell (1898) argues, banks can impact supply and demand by either raising or lowering the discount rate. Indeed, Wicksell provides an apt summary of the law of reflux, which was formulated by John Fullarton (1845) almost fifty years earlier:

> "If a bank provides credit on too liberal a scale it is in direct danger of its notes or cheques becoming concentrated in the hands of the other banks and being presented by them for *redemption*; or, at best, it might have to pay a higher rate of interest on its current account with the other banks than the rate that it receives." (p. 85)

In fact, Wicksell cites John Stuart Mill (1848), who does an even greater job of showing how banks, or financial intermediaries more generally, are very much limited in their ability to expand the supply of notes or deposits by the general demand of the public for such notes and deposits:

> "Even if we suppose, as we may do, that bankers create an artificial increase of the demand for loans by offering them below the market rate of interest, the notes they issue will not remain in circulation; for when the borrower, having completed the transaction for which he availed himself of them, has paid them away, the creditor or

[36] As we will later discuss the (monetary) theory of financial backing, we should appreciate that Wicksell is a referring to issuers of in-silver-redeemable notes that were not 100% backed by physical silver but rather by a combination of silver and financial assets.

> dealer who receives them, having no demand for the immediate use of an extra quantity of notes, sends them into deposit. In this case, therefore, there can be no addition, at the discretion of bankers, to the general circulating medium: **any increase of their issues either comes back to them**, or remains idle in the hands of the public, and no rise takes place in prices." (J.S. Mill, 1848, book iii, chap.xxiv, §2) [emphasis mine]

At a later point, Wicksell (1898) repeats the same argument in slightly different words: "The 'supply of money' is thus furnished by the demand itself." The supply of money is endogenous, determined by the portfolio demand for monetary assets.

This is, of course, completely valid *as long as* holders can actually reflux bills or deposits to their respective issuing banks (either by asking for redemption in base money or by spending it so that it ends up at another bank that then demands redemption in base money). And we should note, as Wicksell (1898) does, that the reflux of notes can take substantially longer and is less direct than the reflux of bank deposits. In this sense, any over-issuance of notes can have a briefly felt impact on prices.

Hence, Wicksell (1898) attempts to explain the demand for "money" (in his mind, generally speaking, consisting of bank notes and deposits) with the rate of interest. The Wicksellian "natural" rate of interest is then the rate at which the price level remains constant. As a consequence, economists should aim to keep both the "natural" and the "market" rate of interest in line with each other. However, this Wicksellian concept of interest has been largely linked to the "Austrian" theory of interest and the business cycle, for instance in Mises (1949). The differences between both concepts is, however, astounding. A major flaw in Wicksell's work is that he turned cause and effect upside down. The banking system and the conditions in money markets determine the rate of interest, and the rate of interest influences the demand for "money" (again broadly defined as bank notes and deposits). Whenever the banking system deviates from the "natural rate of interest," an increase or decline in the general price level ensues. If the banking system maintains interest rates below their "natural level," inflation occurs. Conversely, if the banking system maintains interest rates above their "natural level," deflation occurs.

However, Wicksell's theory gives a surprising large degree of autonomy to the financial sector in fixing interest rates. According to Wicksell (1898), "(…) banks, or rather the aggregate of banks taken as a whole, can (…) lend any desired amount of money for any desired period of time at any desired rate of interest, no matter how low, without affecting their solvency, even though their deposits may be falling due all the time." (p. 111). What Wicksell is essentially saying, is that banks can lend freely without any limits, since their solvency is never at risk. On the flipside, liquidity is of no concern, since any change in the demand for bank money is accounted for in the value of bank money (that is, inflation or deflation). We will later see that banks, and financial intermediaries more generally, do not have such a large degree of autonomy in fixing interest rates.

2.4.2 Wicksell's Theory of Capital and Interest

As a next step in Wicksell's logical analysis, he attempts to explain the nature of capital and the origin of interest, citing the "brilliant work" of both Jevons and Böhm-Bawerk. Wicksell (1898) comes to the following conclusions:

1) **Capital as a wage-rent fund**

   It is obvious that Wicksell (1898) is still swayed by the classical economists' distinction between capital, labor and land. According to Wicksell (1898), the distinction between "circulating" (liquid) and "fixed" capital is of little importance. However, Wicksell (1898) does not offer a very fruitful alternative when he introduces the terms "free" and "invested" capital, despite the fact that the terms *liquid* and *illiquid* (or the less modern notion of circulating/mobile versus fixed capital) are actually often-used terms in finance. Capital theory should be mostly concerned with "circulating capital." While criticizing Jevons (1871) for thinking of capital as a "wage fund," Wicksell (1899) – remarkably – offers not much more than a slightly modified alternative in the form of a "wage *and* rent" fund, grounded in his emphasis on "circulating capital" (which is supposedly the scope of capital theory, since circulating capital is responsible for payments toward fixed capital). Fixed capital is largely set aside.

   First, he cites Jevons (1871): "The single and all-important function of capital is to enable the labourer to await the result of any long-lasting work, (…) to put an interval between the beginning and the end of an enterprise" (p. 223). We have seen this same erroneous idea in Böhm-Bawerk (1888). However, according to Wicksell (1898), the idea of capital as a "subsistence fund" out of which workers ought to be paid is fallacious: “The rewards of the other factors of production, and particularly of land**, must [also] be taken into account**. (…) In its essence it is not merely a “wages fund” but also a “wages and rent fund.” (p. 123) [emphasis mine]. Wicksell (1898) persists in his idea of capital as a subsistence fund, albeit not for mere labor, but also for the owners of the factors of production. Hence, besides the interest that accrues to capitalists, “the total amount of consumption goods produced yearly, monthly, or weekly can be regarded, on the assumption of a stationary state, as a *fund* for the payment of wages and rents. This fund represents the (real) demand for labour and land.” (p. 125).

Capital is, essentially, a fund that cranks out a certain amount of consumption which is needed to cover the current consumption of both workers *and* land owners (and, likewise, owners of “rent-earning goods,” which is something we will discuss below), instead of merely the consumption of workers. Then, the maximum average period of production that capital is able to “sustain,” given its size, is what determines in what lines of production we invest. It is again, in the footsteps of Böhm-Bawerk (1888), the world upside-down: capital is a “liquid” subsistence fund that determines for which production period we are able to cover our current societal consumption, and given its size, we are able to pick more roundabout and physically productive processes that optimize the productivity of laborers. We can express Wicksell’s (1898) according to his own K/T ratio. K is the total amount of capital and T the total period of production: dividing the two gives us the amount of capital that “becomes free” in a single year by sake of being converted into consumers’ goods, and thus represents the consumption of laborers and rent-receivers.

Then, Wicksell (1898) does something completely inadmissible from a theoretical point of view: in a completely arbitrary act (which Wicksell recognizes as being completely arbitrary), he argues that *durable* capital goods (e.g., houses) should not be considered capital but rather something more like *land*, as “rent-earning goods” (p. 126). His main reason for doing so is that a durable capital good such as a house does not require the co-operation of labor and other factors of production (Wicksell, 1898). He is mistaken, however, as we will later discuss in greater detail, since a consumer can only consume the services of shelter by actually *living in it* (e.g., Vargo & Lusch, 2004). This point gets even clearer if we take a moment to reflect upon the examples mentioned by Wicksell himself, such as railroads, machines and canals, which he refers to as “rent goods,” separate from “capital goods.” Owning durable capital goods provides its owner an opportunity to earn *rents*. They are therefore more like quasi-land than capital goods, which equal the “means of sustenance.” Capital then, according to Wicksell (1898), is the number of available workers multiplied by the average wage rate plus the amount of available land (or other “rent-earning goods”) multiplied by their corresponding rent times T (average period of investment). This gives an aggregate K, or capital, which equals the total sum of future means of sustenance. Dividing K by T, we get the average means of subsistence for a single year: “The total quantity of consumption goods is then the same thing as the quantity of liquid real capital in its free form” (Wicksell, 1898, p. 137). In sum, savings leads to an increase in “liquid” real capital, which allows for an expansion of production and an increase in the average period of production (or investment period). It may or not may lead to an increase in the amount of fixed capital, that is, “rent-earning” goods that are economically similar to land. Since, in many cases, longer (more roundabout) periods of production are more productive, such an increase leads to a greater physical abundance of consumption goods.

We have attempted to refute this subsistence theory of capital earlier, yet various comments can be made at this point:

(a) Not all savings are in “liquid” capital; savings can just as much be directly invested in “rent-earning goods”;

(b) Increasing liquid capital to expand the average period of production (*T*) makes little sense since saving equals postponing consumption; if society is a “net saver” than owners of rent-earning goods would also *reinvest* their rents; hence, Wicksell (1898) ends up in some circular reasoning in which it is far from clear *why* an increase in liquid capital (the “means of subsistence”) would be necessary in the first place;

(c) In line with our previous point, as Stigler (1941) pointed out, Wicksell (1898) commits a capital sin by “(…) making ownership a criterion of capital, when in fact this aspect is completely irrelevant” (p. 197); Wicksell seems to think that capital is the fund that makes possible the payment of rents to owners of land, quasi-land and labor (wages) without seeming to understand that if the receivers of such rents *save*, rather than *consume* their rents, the means of subsistence is increased, and so on *ad infinitum*. The excessive focus on ownership (of land and quasi-land) seems to confuse Wicksell's understanding;

(d) Wicksell's (1898) capital theory is the world upside-down: an increase in liquid savings, increases the means of sustenance, and therefore the ability to pay owners of land and quasi-land and wage-earners; this increases the average period of production (T) that the total means of sustenance is thus able to cover. Capital is equal to the total means of sustenance; capital is a "wage-rent" fund. Nevertheless, it is the other way around: capital is rather the result, valued in terms of a common denominator (money), of the (expected future) *value* productivity of productive assets, among which labor (human capital) and other means of production have a prominent role. Wicksell (1898) was, unfortunately, just as Böhm-Bawerk (1888) swayed by the classical legacy of land, labor and capital (a result of combining land and labor), in which all durable capital goods are rather land and capital is a mere subsistence fund to pay the owners of labor and land (and quasi-land). An increase in productivity is therefore reflected in capital as an *outcome*, rather than the other way around ("liquid capital" as a *cause* of higher productivity);

(e) And, most importantly, as Fisher (1930) showed, simply expanding a period of production does not create economic value *per se*[37].

2) Human capital and intangible assets should not be considered capital

Without providing any explicit argument on why human capital and investments in human capital should not be considered capital, Wicksell (1898) excludes human capital from his theory of capital: "[A] very significant amount of capital is often invested in the development of human capacities, but these capacities cannot be included under the conception of capital, neither in the narrow nor in the wide sense of the word. They must be regarded rather as the basis for a particular kind of labour (skilled labour)." (p. 128). The great emphasis that Wicksell puts on the materiality of capital is surprising: some objects are simply "imbued" with the inherent properties of

[37] We have discussed this point in greater detail on p. 84.

capital goods in “the real sense” (p. 128). Some sources of income, such as patents, legal monopolies, goodwill, et cetera should, as a consequence, be excluded from a fruitful capital theory. However, if we take the entrepreneurial view of capital, it would be hard to exclude both human capital and intangible assets from our concept of capital (Menger, 1888).

3) When production comes “full circle”, capital becomes free again

As Wicksell (1898) argues, every production process involves capital, labor and land. While it is difficult to determine the exact contributions of each of the factors, the production process clearly has an “average period of investment” with a clear finish line. The process comes full circle when the finished product is manufactured, and the annual flow of finished products constitutes the “annual” wages-and-rent fund. *K*, the sum of all circulating capital, is then the aggregate wages-and-rent fund. This fund should not just last “for *one* year, but a number of years equal to the average period of investment of the capital.” (p. 130). This allows a reconciliation of Jevons’ wage fund (which in reality is a wage-rent fund) with Böhm-Bawerk’s average period of production and subsistence fund. However, as we have discussed above and earlier, the concept of capital as a wage fund (or, more broadly, a subsistence fund) and the concept of an “average” period of production are completely fallacious.

4) The origin of the rate of interest is the supply of “liquid” capital

Wicksell (1898) states that an increase in savings leads to an increase in the “available amount of liquid capital” (p. 132). This increase results in raising wages and rents, but decreasing returns on capital. However, durable goods and land (Wicksell’s “rent-earning” goods) are an exception to this rule; they temporarily enjoy a higher return. Temporarily, since entrepreneurs compete and subsequently push these higher returns down, largely by investing in new “rent-earning” goods that are undervalued given their lower replacement cost to market price. Wicksell is all but clear on this point, however, since he never explains exactly how he relates this dynamic to his other ideas (for instance, if entrepreneurs borrow liquid capital to invest in durable “rent-earning goods” in an attempt to push their prices down, then that capital is no longer liquid and actually, according to Wicksell’s own theory, requires additional liquid capital to service its future “rent” payments. Considerations about these durable “rent-earning” goods aside, Wicksell argues that a decline in the rate of interest makes (again, according to Wicksell, "for technical reasons") production processes with longer periods of production more profitable and processes with shorter periods of production less profitable. This is very much in line with what Mises (e.g., 1949) would argue. Yet, Wicksell (1898) then writes the following: "The final result of these changes is *that the average period of investment of the aggregate of (circulating) capital is lengthened; and the portion of the wages-and-rent fund which becomes available in any one year is consequently diminished*" (p. 133) [emphasis mine]. Remember that (circulating) capital in a Wickselian world are liquid assets (finished consumption goods) that are used to purchase labor and services of "rent-earning" goods. It does not make much sense to speak of "lengthening" liquid assets; a nominal dollar today is a nominal dollar tomorrow. Rather than the interest rate "lengthening" circulating capital, it should "lengthen" to a far greater degree projects with high amounts of fixed capital (Huerta de Soto, 2006), that is, savings are invested in increasingly capital-intensive projects with long durations. Whereas Wicksell thinks he is "reconciling" Jevons' and Böhm-Bawerk's theories, he is actually confusing them or, perhaps, combining the poorest contributions of both.

5) **The "contractual" (market) rate of interest is different from the "uncontrolled" (natural) rate of interest; Wicksell's inflation premium**

Interestingly, Wicksell does recognize the stabilizing and equilibrating role of the entrepreneur. The entrepreneur, according to Wicksell (1898), arbitrages the market rate of interest down to its "natural" level. As a consequence, Wicksell implicitly admits that the rate of interest is a result of a society's time preferences (and represents a mere allowance for time), whereas entrepreneurs use savings to invest and arbitrage away profits, which appears to contradict his earlier exposition. Hence, the return on capital is different from the rate of interest, something Wicksell (1898) at earlier points in his seminal work seems to deny[38]. However, there is one exception to this rule: entrepreneurs have an incentive to demand savings and arbitrage the market rate of interest down, with a proper allowance for changes in the purchasing power of money.

[38] Wicksell's (1898) work can be aptly described as, at times, plain confusing and contradictory.

Wicksell (1898) perhaps would have reached a more coherent analysis if he were familiar with modern-day finance. He, in effect, argues that both labor and capital goods require outlays, that is, negative cash flows. The inflows, or positive cash flows, are netted against the negative cash flows. The net result is both a pure entrepreneurial profit (return on capital) and a compensation for time (rate of interest), besides a possible wage for the entrepreneur if it was not explicitly included in the negative cash flows as labor cost. He then considers examples such as when rents or wages fall (for instance due to population growth), which lower the negative cash flows and effectively raise the rate of return on capital. An expansion of activity, even if this adjustment takes time, ensues (Wicksell, 1898). This is our way of reconciling Wicksell's work with modern-day finance theory and salvaging whatever we can from his work on capital theory.

Interestingly, however, Wicksell (1898) does recognize the fact that long-term investments (put differently, the prices of "productive assets" with longer durations) are more susceptible to changes in interest rates. Long-term investments are affected disproportionally by changes in interest rates. We will see in the following sections that this recognition is valuable, especially with regard to the debates on "shiftability," which was supposed to stimulate long-term investment.

Circling back to Wicksell's work on money and how it relates to the capital structure, it is worth mentioning that Wicksell (1898) unfortunately seems, as many authors do, to single out the central bank. For all practical purposes, there is little difference between a saver making a deposit at a bank and the bank investing in a mutual fund or a saver making a deposit at a bank and the bank holding a central bank deposit. If a central bank's holdings consist of 100% gold while commercial banks hold central bank deposits and savers hold commercial bank deposits, it simply means that savers are indirectly invested in gold. There is no magical money multiplier; the central bank has no exceptional role whatsoever. If banks issue demand deposits which are held by the public, and banks hold central bank deposits, while the central bank is invested in illiquid long-dated debts, then the public is *de facto* invested in illiquid long-dated debt[39].

2.5 Irving Fisher's Early Discovery of the Reswitching Phenomenon

Irving Fisher's classic work came off the back of the first round of capital controversies. In 1907, Fisher published his seminal work, *The Rate of Interest*, in which Fisher, among other things, reviews the work of Böhm-Bawerk and Clark. With regard to Böhm-Bawerk's concept of roundaboutness, Fisher (1907) was the first to note the inconsistency of having two separate cash flow schedules which at various discount rates would "switch" from being less to more and back to less profitable in a counterintuitive way. This inconsistency would, of course, culminate in the third round on capital controversies, where this "capital reswitching" (or "capital reversing") would become ostensibly the most important point of discussion. Fisher (1907) can be credited with being one of the first to observe this apparent anomaly, which runs counter to the "Böhm-Bawerkian" idea that one alternative will be chosen when interest rates are high and some other alternative when interest rates are low. Fisher (1907) in his example assumed two income streams:

1. $5 after ten years, $100 after one hundred years
2. $15 after twenty-five years

In Fisher's (1907) own words:

> "[It] is not true that one of the alternatives will be chosen if the rate of interest is high, and the other if the rate of interest is low (…). The application of labor which issued in the $5 and $100 would, oddly enough, be the most economical if the rate of interest were either very high or very low, whereas the other alternative would be chosen in case the interest were at a more moderate rate" (Fisher, 1907, p. 352)

[39] This describes, in fact, very accurately the situation of the U.S. banking system in the run-up to the Great Depression. Commercial banks would on the one hand issue demand deposits and on the other hand hold callable debt from unregulated banking trusts, such as the Knickerbrocker Trust. The failure of such banking trusts, which lacked access to a clearinghouse and depended entirely on the benevolence of their commercial bank counterpart, was the source of the demise of the U.S. banking system at the time.

In Fisher's example, any rate of interest below four percent would result in a higher present value of income stream #1. From roughly four to seven percent, the present value of income stream #2 would be higher. However, at any interest rate above seven percent the present value of income stream #1 would become higher again. As Osborne (2014) remarks: "This phenomenon that the present value of an income stream can exceed the present value of an alternative income stream at both low and high rates of interest, but be lower at intermediate rates, eventually became known as reswitching" (p. 94).

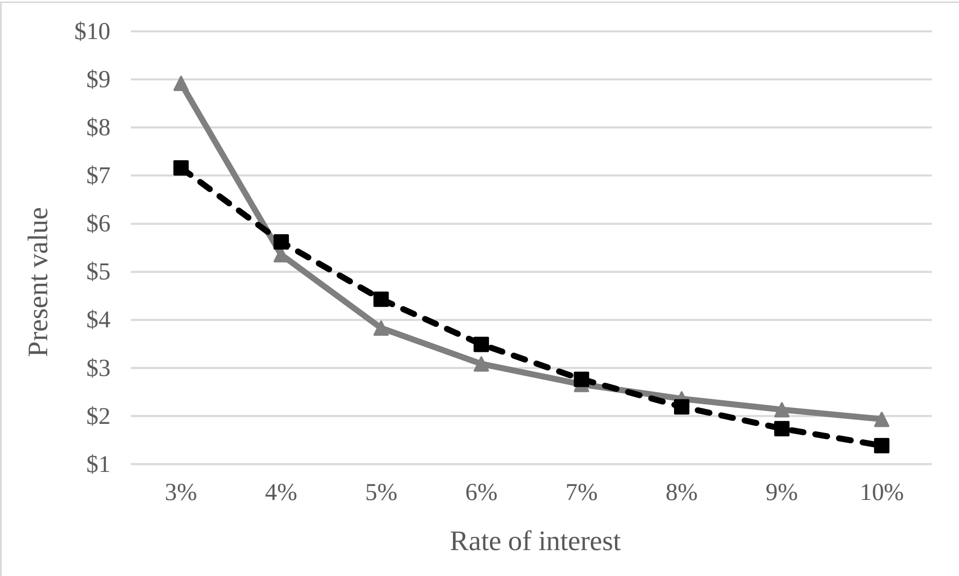


Figure 3: Fisher's (1907) discovery of reswitching, which eventually became front and center in future controversies on capital. Grey, straight line represents income stream 1; black, dashed line represents income stream 2.

The consequences of Fisher's discovery were largely overlooked at the time. Even Fisher himself did not make much of his discovery. Nevertheless, reswitching would become one of the most debated issues in the third round of capital controversies, especially in the 1960s.

Curiously, it was Fisher (1930) himself who noted the necessity of using *multiple* rates of interest, criticizing the idea of one *single* uniform rate of interest for *all* maturities (which Fisher effectively does in the above-mentioned example). Instead, every separate period of time should have its own separate rate of interest: "[T]he necessity of positing a theoretically separate rate of interest for each separate period of time, or to put the same thing in more practical terms, to recognize the divergence between the rate for short terms and long terms. This divergence is not merely due to an imperfect market and therefore theoretically subject to annihilation by arbitrage transactions, as Böhm-Bawerk, for instance, seemed to think.

They are definitely and normally distinct and due to the endless variety in the conformations of income streams. **No amount of mere price arbitrage could erase these differences**." (p. 104). In effect, Fisher (1930) defended a term structure of interest rates, whereas Böhm-Bawerk (1888) defended a single uniform rate of interest[40].

2.6 The Shiftability Debate: Can the Banking System Turn Circulating Capital in Fixed Capital Without Repercussions?

2.6.1 The Intimate Relationship between Money, Liquidity and ("Circulating") Capital

Historically, money has often been analyzed *aside* from capital. It goes frequently unrecognized that financial assets (in particular, banks' demand deposits) have a counterpart that directly influences an economy's capital structure. Money is, in essence, a transmitter of savings into investment. This becomes all the clearer, of course, in modern-day financial asset money (such as bank money). In another essay, titled *Geld* (later translated into *On the Origins of Money*), Menger (1909) shows to be aware of this intimate relationship between money and capital. He writes:

> "In fact, one could say that, beyond its [i.e., money's] function as a means of exchange (as an intermediary in the goods market) and its use as a preferred means for hoarding and accumulating capital, there exists no other function of money that (…) has such a great importance for the entire economy as **its function as intermediary between capital markets and money or monetary markets**." (p. 55)

The resemblances with Böhm-Bawerk (1888) abound:

> "[…] an economically advanced people does not hoard, but puts out what it saves—in the purchase of valuable paper, in deposits in a bank or savings-bank, in loan securities, etc. In these ways the amount saved becomes part of productive credit; it increases the purchasing power of producers for productive purposes; it is thus the cause of an extra demand for means of production or intermediate products; and this, in the last resort, induces those who have the regulation of undertakings to invest the productive powers at their disposal in these intermediate products." (pp. 115-116)

[40] We will discuss the (importance of the) term structure of interest rates further below. See p. 213.

What Böhm-Bawerk explains, is that savings is a uniform phenomenon not different from hoarding. The big question is hoarding what: by hoarding financial assets instead of physical monetary assets savings are, to a great extent, directly put to work. He therefore makes a case against saving in nonfinancial assets. However, it should be clear that Böhm-Bawerk does not seem to realize that even savings in physical, nonfinancial monetary assets are transmitted as "productive powers" to "those who have the regulation of undertakings", albeit less efficiently.

The principle focus of this work will be on money as an intermediary in capital markets and its intimate relationship with capital theory. In fact, capital theory is and should a *de facto* integration of money in economic science. The neglect of the monetary debates on shiftability (which we will discuss further below) in the 1920s by economists is, therefore, unfortunate. Indeed, one of its key contributors, Harold Moulton (1921) criticizes economics for ignoring the role of money. Economics is, says Moulton (1921), divided into the subjects of:

(1) Production
(2) Exchange
(3) Distribution

The first concerns combining the factors of production, labor, land and capital (the notion we have so far extensively criticized). Once the goods are produced, we are faced with the laws of value since we must exchange them. While referring to exchange, economists suddenly notice that money is used as unit of account and medium of exchange and that there exist various "types" of money (deposits, currency, et cetera). Last, once the goods are produced and exchanged, we are now faced with the question of how the bounty is divided[41]: who receives what? All in all, "most writers leave the impression that money is largely divorced from the productive process." (Moulton, 1921, p. 376). Yet, "every act of the productive process itself revolves about the use of money" (ibid., p. 379). When one views the role that:

> "(…) money, [as] a common denominator of value, or pecuniary unit of calculation, plays in the organization of productive activity, it becomes readily apparent that money in this capacity also plays an extremely important role in the field of

[41] This was the source of the first controversy on capital. Böhm-Bawerk and Clark, as well as their earlier predecessors, were mainly concerned with (income) distribution.

production. (…) The price-and-profit system based on the pecuniary unit serves as a guide in the directing of labor and capital from place to place and from industry to industry. Accounting systems based on the dollar unit also make possible governmental supervision and control of industrial affairs. All this finds no place, however, in the traditional discussion of the role of money in economic activity, despite the pretension that the analysis is disclosing the way in which society has organized itself for the production of wealth." (Moulton, 1921, p. 379)

Moulton (1921) espouses a very Mengerian and Misesian view of capital. Money is very intimately related to capital, in the sense that capital is simply a tool of economic calculation, of profit and losses expressed in terms of money, that guide production in a myriad of ways. Discussing money and banking under the pretext that money's exclusive role is to facilitate the exchange of goods that have been produced by land, labor and capital, with no role for money whatsoever in the process of production, is academic "in the objectionable sense of [the] term," according to Moultin (1921).

Menger (1888), introducing the idea of capital as a finance or subjective concept (as *net worth*) rather than a material concept, does distinguish between *circulating capital* and *fixed capital* (working capital and fixed capital in modern-day terminology). Yet his distinction relates more to the idea that some assets (which are capital when and if expressed in terms of money) are either less or more liquid than other assets. However, an asset's liquidity is often priced in by the market, in a "liquidity premium." As such, the inherent liquidity of an asset is a part of its value and use and appraised in terms of money.

This liquidity premium, in the case of banking, takes a very specific shape. One of the reasons why bank deposits are so *liquid*, is the fact that banks offer ways to easily exchange bank deposits among depositors (even of other banks). Hence, they offer specific services that *contribute* to the liquidity of this specific financial asset (that is, a demand deposit). The 17^{th}-century Bank of Amsterdam was located in the same building as the city hall, on the westside of the Dam Square, a stone's throw away from the market places where a majority of exchanges took place (including physical spot markets, futures markets and the stock market). Money transfers between accounts could be conveniently and swiftly handled at its office. Now, such payment services carry a certain cost, which is partly responsible for the liquidity premium on (transferable or checkable) financial assets. Competition between issuers assures arbitrage and caps such a specific liquidity premium for building and maintaining the systems required for payment and transaction services. Just as (both monetary and nonmonetary) financial assets, productive assets can have a liquidity premium. Some productive assets are more easily converted into something else, made liquid, than other productive assets. It should be beyond doubt that the optionality of a containership is very different from an account receivable (within an X number of days). Hence, whenever there are markets, the ability to exchange one asset into another is more or less guaranteed. However, the price at which such exchange can be done is highly variable. Therefore, any liquidity premium refers to the loss involved when liquidating an asset. What is generally called "circulating capital" can be exchanged into other assets at a relatively low cost and minimal, if any, losses. This is purely due to the fact that the defining factor of circulating capital is *time*. Per definition, productive assets are considered "circulating capital" if they are converted (or convertible) into cash within twelve months. The fact that they are made liquid within twelve months, assures greater liquidity even before the end of these twelve (or less) months, resulting in some type of a liquidity premium over less liquid assets.

This is one of the first instances where we can see economists grapple with money on the one hand and capital on the other hand. Is money part of capital? Does money "represent" capital? If capital means the principal of a money loan, and interest the income from a money loan, does that mean that capital in an economic sense is the complex of production goods and interest the income such production goods are able to yield? Or is there more to it?

Curiously, the treatment of various economists of this era (e.g., Menger, 1909; Wicksell, 1898) of money to circulating versus fixed capital had few repercussions, both theoretically and practically. It is all the more surprising that, in the very same period, bankers were debating this exact issue completely apart from the economists. The discussion on circulating and fixed capital, and the role of banking institutions in the composition of capital, became front and center in the debates leading up to the founding of the Federal Reserve, with however completely different actors involved. Indeed, Moulton (1921), one of the participants in the debate on shiftability, visualized the division of fixed and working capital and their relation to "money" and capital markets in the following scheme:

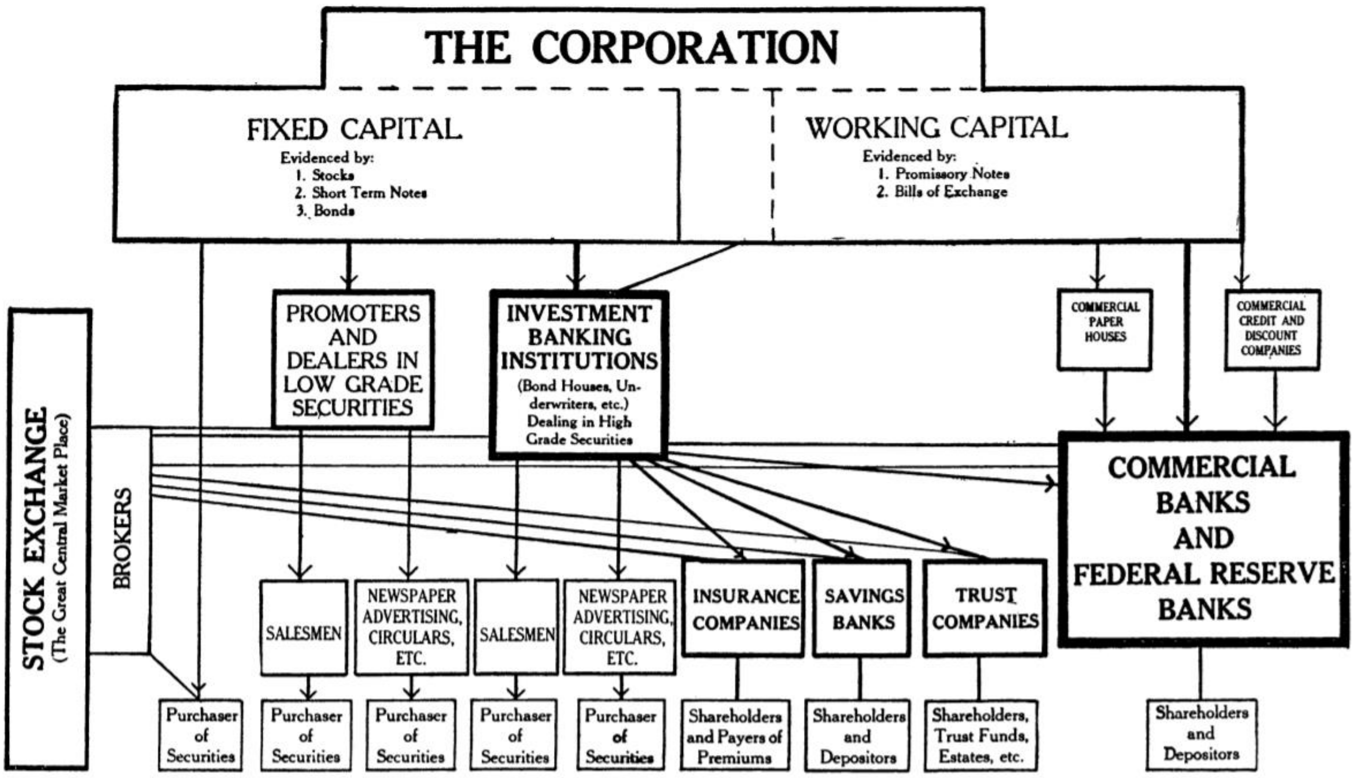


Figure 4: Moulton's (1921) visualization of the intimate relationship between (fixed and circulating) capital and capital markets[42].

Carl Menger, however, in his essay on the origins of money (1892), lays out a theory to explain the evolutionary origin of money. Money, according to Menger (1892), is the most liquid good (Menger uses, for the lack of a better term, the term "saleable" instead of "liquid"). It is important to note at this point that Menger's definition of liquidity is closely related to a popular modern-day definition of liquidity: the narrowness of the bid-ask spread[43] (Mancini, Ranaldo, & Wrampelmeyer, 2013). On the contrary, when we look at

[42] Further below we will discover that one of the main problems of "shiftability" (which even continues into our modern era at even greater extremes) is that the banking system increasingly finances (long-term) fixed capital with short-term credit (mainly zero-maturity and short-duration demand deposits).

[43] In practice, the "bid-ask spread" definition of liquidity has many problems: spoofing, to start, are fake orders that will be cancelled or withdrawn before they are executed, giving an illusion of market depth. Moreover, the bid-ask spread gives an *ex ante* indication of how illiquid an asset (or set of assets) is, yet what ultimately counts is the *ex post* loss incurred as a result of the liquidation.

Menger's *Geld* (1909), the concept of liquidity was skewed more toward the modern-day notion of "price impact," the degree to which a liquidation impacts the market price of the asset (Pastor & Stambaugh, 2003). Liquidity, then, is "the ability to trade large quantities quickly, at low cost, and without moving the price" (ibid, p. 644) or the time required to liquidate assets "without significantly changing [its] market value (…)" (SEC, 2016, p. 439). The latter two definitions are practically different sides of the same coin.

According to Menger (1892), goods are more or less saleable when "according to the greater or less facility with which they can be disposed of at a market at any convenient time at current purchasing prices, or with less or more diminution of the same" (ibid., p. 25). However, as Zarlenga (1994) notes, Menger gives unwarranted importance to "liquidity" (defined in this case as the size of the bid-ask spread) relative to "volatility" (defined in this case as its price stability over time and much more related to the modern notion of liquidity as "price impact" or the loss involved in liquidating an asset), apparently assuming that a tight spread is by definition accompanied by low volatility. Nevertheless, Zarlenga (1994) provides various hypothetical examples. He compares, for instance, a commodity *C* with a wider bid-ask spread but lower volatility and a commodity *G* with a narrower bid-ask spread but higher price volatility.

Moreover, the moment we reassess Carl Menger's (1892) theory on the origin of money, he seems to omit the fact that his theory does not explain the origin of money as *unit of account*. The unit of account is the monetary yardstick that people use to post prices, record debts and practice economic calculation (e.g., capital accounting, that is, profit and loss accounting). As Nick Szabo (2016) notes:

> "Prior to the rise of efficient competitive markets, prices for goods were often specified by custom or law rather than negotiated. This served to conserve transaction costs in a high transaction cost culture where exchange relationships resembled bilateral monopolies more closely than they resembled spot markets. Bargaining costs were high, and indeed bargaining failure often resulted in violence and destruction rather than merely in no deal. This made focal points of negotiation, such as customary prices and customary compensation amounts for specific injuries, a quite valuable and ubiquitous part of most Neolithic and earlier cultures. When specified by law, these rules setting prices were often intermingled with laws specifying legal penalties and used the same set of units: in the Mesopotamian and Anatolian law

> codes prior to coinage, most commonly weights of silver and volumes of barley." (para. 4)

Szabo (2016) seems to adopt the same language as Menger (1871) and Böhm-Bawerk (1888), although Böhm-Bawerk uses the term "isolated exchange" instead of Menger's and Szabo's "bilateral monopoly." The conclusion should be that it would be sheer impossible for a good to become a unit of account or a general measure of all other goods that are bought and sold, just as a result of its use in on the spot isolated exchanges. No prices would be ever documented or observable to external third-parties that did not directly witness a transaction. And, moreover, due to the sheer quantity of goods exchanged, it seems unlikely that mere spot exchanges would have resulted in a unit of account (Rallo, 2016).

Some type of forward exchange (or credit) would be necessary to arrive at a unit of account and make ancient accounting possible, much earlier than the first instances of double-entry bookkeeping in 15th century Italy. In fact, archeological findings show that approximately 15,000 years ago records were found that listed stocks of goods and transactions that originated in a temple in Mesopotamia. Later, the ancient Egyptians and Babylonians kept records of tax payments. The Mesopotamian shekel, which emerged at a later phase of increasing sophistication, became the world's first unit of account. Shekels were coins that represented a claim on a certain quantity of barley that was warehoused.

Moreover, as Szabo (2016) and Rallo (2016) argue, it is likely that grains such as barley and wheat as well as primitive forms of credit (IOUs) were used *within* communities with high levels of trust or effective compliance mechanisms, while other goods such as cattle and metals such as silver were used for trade *across* communities, where exchange relationships were often times characterized by low trust or even coercion by physical dominance in the form of taxes or tributes. Goods such as silver (and later gold) were generally expressed in terms of grain or cattle, which could be directly related to the productive effort and harvests of each individual or household. Tax "rates" were often recorded in both *grain* and pieces of silver, with both having a fixed rate of exchange. Both were, in many cases, expressed in terms of weight.

There is a good case for the unit of account emerging from law or taxation, either by governments or bodies that act as *de facto* governments, such as churches and temples. Legal penalties (for homicide, for instance) and taxes were largely determined relative to the production of an ordinary man. In very early instances, taxes actually depended on the produce.

So, what was the purpose of our little detour to Menger's theory on the evolutionary origins of money? What was the reason we took a more roundabout approach to money in this subsection?

We can appreciate from our insight that the *least* volatile good (or one of the least) would emerge as the preferred yardstick of prices (exchange ratios). However, a good does not become less volatile because of monetary demand. A good was already less volatile than others because of the nature of its (nonperishable) supply and steady demand[44] and is, as a result, preferred as yardstick. Hence, Menger's (1892) emphasis on liquidity as "marketability" is only part of the story and, moreover, perhaps the least important part. Liquidity should *also* refer to our notion of "price impact" (or loss) at liquidation, which we have discussed above, rather than to marketability with zero reference to price. Any asset is, in fact, "marketable," but at a cost. The degree of this cost determines whether an asset (or bundle of assets) is liquid or illiquid. We will for instance see that when banks' demand deposits are backed by or invested in short-term credit, this cost tends to be negligible, since such credit is soon repaid in full. As a result, we have vindicated Wicksell (1892) *somewhat*: to Wicksell, assets are liquid if they are converted into (or convertible into) consumers' goods in less than twelve months. Hence, while capital has little to do with Wicksell's (1892) and Böhm-Bawerk's (1888) earlier extensively criticized subsistence fund, liquidity does. As we will see below, liquidity ultimately refers to the degree in which production is aligned with consumption over time.

2.6.2 Ludwig von Mises on Money and Banks as Intermediaries of Savings

Ludwig von Mises, in his first major contribution *The Theory of Money and Credit* (1912), attempts to separate money, as a tangible good (historically, gold), from credit and the intermediation of savings. On banks, Mises (1912) writes: "Banking is negotiation between granters of credit and grantees of credit. Only those who lend the money of others are bankers; those who merely lend their own capital are capitalists, but not bankers" (p. 262).

[44] This is, consequently, one of the main problems with Bitcoin: real adoption is impaired by its price volatility (and thus the potential losses associated with liquidating positions).

He implicitly rejects the view that hoarding money (that is, gold) is a form of savings and a form of transmitting savings to production at large (or, in other words, converting savings into productive investment). To him, only if savers decide to hold claims on money, can savings be converted into investment. Yet the money business is a completely separate branch from the credit business. This view seems untenable.

Mises (1912) writes: "A person who has a thousand loaves of broad at his immediate disposal will not dare to issue more than a thousand tickets each of which gives its holder the right to demand at any time the delivery of a loaf of bread" (p. 267). He ignores, however, that a person would actually do so, *if* he expects the delivery of other loafs of bread in the near future and *if* he expects that all holders show up the very same day to claim their breads. Expanding our analogy to money and banking, a banker tries to estimate the expected rate of withdrawals and/or adverse compensations to other banks is, as to make sure that he can meet any future redemptions. Mises, in this case, ignores the temporal element of issuing claims bearable on demand (as demand deposits are) and the businessman's experience as to at what rate claims are redeemed.

At a later point in his life, in his chef-d'œuvre *Human Action* (1949), Ludwig von Mises defines money as the "commonly accepted medium of exchange." He then proceeds to explain the origin of money, very much line with Menger's (1892) theory of the evolutionary origin of money[45]. However, Mises merely explains the origin of gold as medium of exchange. In a world *without* financial intermediation, this might be of great relevance. Nevertheless, in a world characterized by an increasingly greater degree of financial intermediation, there exist many other media of exchange that should be considered. Mises (1912) called these media of exchange "money substitutes." In other words, Mises, probably unaware of his own definition, operationalizes money not as commonly accepted medium of exchange, but rather as the "ultimate extinguisher of debt." Mises says A, but uses B.

Hence, Mises (1949) falls into a contradiction. Laymen commonly accept commercial bank deposits as media of exchange. In popular language, we even call these demand deposits "money". However, Mises operationalizes money not as demand deposits, but rather as the instrument or good that is no longer "redeemable" or "convertible" into something else. In other words, Mises first defines money as the commonly accepted medium of exchange, but

[45] We have discussed Menger's theory on capital and money earlier, see p. 99.

when he begins defining “money substitutes,” he redefines money unwittingly as “the irredeemable instrument.”

Take Mises’ hypothetical view on modern “money substitutes” such as shares in money market funds (MMFs). A MMF is a mutual fund that issues shares with a nominal value of $1, which the fund attempts to maintain stable (that is, at “par”) over time. The MMF issues shares when an investor deposits funds and uses those funds to buy liquid, short-maturity bonds (certificates of deposits or CDs, short-term liquid corporate bonds, etcetera). Any interest is paid out to shareholders in the form of new shares, which are exactly worth $1 each. A MMF fund manager attempts to avoid any loss of principal (limiting himself to only creditworthy counterparties), any holdings of illiquid assets (which could potentially impair asset values if shareholders begin redeeming shares, that is, if a liquidity run occurs), while he simultaneously tries to maximize yields[46]. Not only are MMFs still completely ignored in Austrian theory. A quick search in the *Review of Austrian Economics* is illustrative. The term “money market fund” is not even once mentioned in any of its articles. Other (Austrian) papers on money market funds are scarce (e.g., Haymond, 2000).

Figure 5: An exemplary balance sheet of a typical money market (mutual) fund (MMF)

Assets	**Liabilities + equity**
Short-term securities (corporate bonds, savings deposits, CDs)	Equity

Now, if money market fund shares can be redeemed into demand deposits of a commercial bank, and these demand deposits can be redeemed into central bank currency, then paper central bank currency, according to Mises (1912), is “money.” Money market fund shares would represent “money substitutes.” In a similar fashion, other liquid credit instruments, like commercial paper and in some cases US Treasuries, would be considered “money substitutes” by Mises (1912) as well. However, bank deposits are just as “commonly accepted” as central bank currency, and in some cases even preferred over physical currency

[46] Negative interest rates have recently destroyed a great part of the MMF industry in Europe, since it is impossible to maintain a stable net asset value (NAV) and pay dividends (which can be considered *de facto* interest payments) when the underlying assets have negative yields.

(bills). Mises's position is therefore contradictory; he *defines* money as a commonly accepted medium of exchange but *operationalizes* money as ultimate extinguisher of debt.

Friedrich Hayek, in his often-cited work *The Denationalization of Money*, recognizes this error, which was not only committed by Ludwig von Mises, but by many other economists. To quote Hayek (1976):

> [A]lthough we usually assume there is a sharp line of distinction between what is money and what is not (…), there is no such clear difference. **What we find is rather a continuum in which objects of various degrees of liquidity**, or with values which can fluctuate independently of each other, shade into each other in the degree to which they function as money. (p. 56) [emphasis mine]

Here, Hayek (1976) moves to a definition of money that does not clearly separate money from credit, but rather argues that any economic good – even loans, bonds and securities – has a degree of liquidity. The most liquid goods tend to be used as media of exchange.

Does Hayek refer to money as being able to extinguish or settle debts? Does Hayek further develop his definition of money? Unfortunately, he does not. He chooses to leave this very important question aside and directly delves into the theory of currency competition:

> "There is, however, as we have just pointed out, **no need for a very sharp distinction between what is and what is not money** [emphasis of the author]. The reader will do best if he remains aware that we have to deal with a range of objects of varying degrees of acceptability which imperceptibly shade at the lower end into objects that are clearly not money." (p. 58) [emphasis mine]

There exist various modern-day authors who have taken Mises' separation between money and credit to an extreme, such as Shostak (2000). Shostak (2000) argues, for example, that a money transaction (money is a "claim" according to Shostak) should be contrasted with "(…) a credit transaction, in which the lender of money relinquishes his claim over the money for the duration of the loan" (p. 72). What Shostak (2000) misses, however, is that money holdings (as part of a portfolio) are simply one form of savings. The portfolio demand for money is thus a subset of the broader portfolio demand for (financial) assets. Moreover, he misunderstands the practical difference between legal maturity and duration. While demand deposits might have a zero maturity, they might have a duration of 12 months[47], since deposit holders do not actually use *all* of their cash balances.

A most troubling and glaring contradiction comes to light when Shostak (2000) asks: “Now, if any mixture of liquidity is accepted, why not include retail good inventories?” (p. 69). As a matter of fact, *claims* on retail good inventories have circulated as means of payment, that is, in the form of bills of exchange. And it is not just bills of exchange; US Treasuries also change hands to settle debts, for instance in the case of derivatives clearing[48]. In sum, the very fact that people hold zero-maturity cash balances of some kind as part of their savings (that is, portfolio) is only different to other types of savings (for instance, in fixed maturity deposits) in degree, not in kind, until the money is actually spent.

2.6.3 An Overview of the Shiftability Debate

The direct motive for the debate on *shiftability*, was the onset of the Federal Reserve Act, which was responsible for the creation of the central bank. What powers and tasks should this new central bank be endowed with? Hence, two lines of thought emerged: one camp was in favor of supporting the banking system so that banks can invest in long-term investment projects (the “shiftability” school), whereas the opposing camp was in favor of limiting the Federal Reserve to the mere discounting of short-term commercial credit, which would force banks to invest in short-term liquid credit (the “liquidity” school).

Ironically, after the Great Depression took its toll on the global economy, the Fed’s powers were even expanded further, as a large market for long-dated US Treasuries was necessary for the American war efforts: as a result, *shiftability* won out over *liquidity*, despite the fact that the *shiftability* theory was one of the main culprits or causes of the Great Depression of the 1930s[49].

The basic idea behind *shiftability* was that capital markets went through a marvelous phase of development. As a result, banks could invest short-term (demand) deposits into illiquid

---

[47] In fact, estimating the duration of demand deposits using a common measure of duration gives an estimate of approximately 1.5 years.

[48] Although instead of US Treasuries changing hands, what in fact changes hands are an IOUs of the clearinghouse, that is, the central counterparty (CCP), to a portfolio of collateral, consisting of cash, bonds, etcetera.

[49] We will discuss the Great Depression in detail later, see p. 418.

assets, which could always be made liquid on a secondary market with sufficiently deep pockets. As Grant's (2010) explains the shiftability-view: "It was not strictly necessary that a banking asset be "liquid" [for example, a short-dated commercial IOU]. All would be well if an asset were "shiftable," i.e., salable in the continually functioning, deep and liquid capital markets of the day." (p. 4). Shiftability was desirable since allowing banks to invest in long-term debt could lower long-term interest rates that, in turn, encourage investment (CapEx). Glock (2017) does an excellent job in summarizing the case that the defenders of shiftability made:

> "[F]rom 1913 to 1935 economists in the institutionalist tradition emphasized three new and related ideas about the importance of long-term credit to monetary management and the control of business cycles. They first noted how long-term interest rates seemed to have a substantial influence on the level of fixed capital spending, especially in regard to cyclical construction spending. Second, they argued that central bank monetary operations in the short-term money market exerted an influence on borrowing in all money markets, short and long term, and therefore could determine total fixed investment in the economy. Finally, these theorists created a new theory of the appropriate assets of commercial and central banks; **instead of emphasizing the inherent liquidity of short-term "real bills" or commercial debts, that were paid down at a fixed date, they demonstrated how long-term debts could be made liquid by being made "shiftable" to another part of the market or a central bank. These theorists argued that increased "shiftability" (…) would lower interest rates on long-term debts and (…) spur fixed investment**." (pp. 3-4) [emphasis mine]

Indeed, Mitchell (1923) defends the "soundness" of the shiftability theory by referring to the fact that it is the "(…) growth of the banking system, rather than the self-liquidating character of the bank's paper, that protects the bank's reserves." (p. 335). Mitchell (1923) refers to the fact that certain long-term non-commercial loans and investments have become very liquid paper thanks to "(…) the development of stock exchanges and other such institutions" (ibid). Long-term government bonds, for instance, had a very "secure legal foundation" (ibid). Whenever a bank is confronted with its illiquidity, it can simply "shift" its assets to stronger third-parties and obtain the liquidity it needs. Commercial banks, by investing a larger part of their assets in long-term investments and projects, could lower permanently long-term interest rates, which could give a lasting stimulus to the American economy.

Nevertheless, what Mitchell (1923) and other proponents of shiftability theory fail to understand, is that *individual* illiquidity is not the same as *systemic* (or system-wide) illiquidity. Moreover, what Mitchell (1923) also failed to grasp, was that systemic illiquidity rears its ugly head not immediately, but only after a substantial period of time and accumulation of illiquid assets. In other words, when banks engage (systematically) in *maturity mismatching* (by borrowing on the short term and investing in the long term), relying on the liquidity of fellow bankers when their assets turn sour, they arbitrage the term structure of interest rates and reduce (relative to short-term rates) long-term interest rates. Investments are undertaken that are more *capital-intensive* and have longer *durations* due to the lower (relative) interest rates. If all banks become illiquid simultaneously, there will be a *systemic* lack of liquidity and hence no assets can be shifted to “stronger” banks. An illiquidity crisis will, inevitably, hit banks at precisely the same time due to the fact that their yield curve arbitrage lowered long-term interest rates below the levels at which they should have been.

According to Moulton (1918) and Watkins (1919) this could be avoided, if some banking regulation would simply impose a fixed percentage of assets (loans and securities) to the consumer sector and to the capital goods sector that match the consumption-savings pattern of households. If banks would follow the *shiftability* theory, “a disproportionate production of capital equipment” could be a result (Watkins, 1919, p. 588). This may lead to “an industrial situation that calls for drastic readjustment” (ibid). Moreover, Watkins (1919) was early to recognize that banks do not actually *create* credit, but merely *intermediate* (bank) savings and credit. Therefore, shiftability could not lead to additional credit, but it could lead to a greater part of credit diverted to long-term illiquid investments. However, Watkins (1919), as well as other opponents of the *shiftability* theory, were unable to develop this insight into a more broader and precisely formulated theory of the (endogenous) business cycle. Yet Watkins (1919) intuitively understood the relation between *intertemporal* consumption patterns and the capital structure and how a misalignment could wreak havoc.

Although Harold Moulton (1918) was a proponent of the *shiftability* theory, he was also one of the first to criticize the excessive focus of economists on banking as deposit-taking institutions and their effects on the supply of money, rather than banking as intermediaries of scarce savings that, in large part, influence the *formation* of capital (specifically: in what kind of assets savings end up being invested). Analyzing the literature up to that time, Moulton (1918) showed that there were three approaches to banking in economics:

1. The analysis of banks by merely describing their "functions";
2. The analysis of (bank) credit as media of exchange;
3. The analysis of banking as a transfer of capital from those who do not wish to use it to those who do.

However, according to Moulton (1918), analyzing the alleged function of a bank is a misnomer: "Providing safety-deposit vaults, accepting cash deposits, making loans, collecting checks, etc., are not functions of banks any more than receiving freight and issuing bills of lading are the economic functions of railways." (p. 495). He concludes that analyzing banking from this point of view is a waste of time.

The second approach relates to the analysis of banking, the supply of money and its effects on the price level, which has been the focus of economists thus far. Yet this is also an unfruitful approach: "[T]he emphasis upon money and its relation to prices has led to a confinement of the discussion of banking to commercial banking only-this for the reason that commercial banking alone provides, in bank notes and deposit currency, media that are acceptable in exchanging goods. In consequence, the whole discussion of banking, with the exception of a brief section devoted to financial panics, has usually centered around the maintenance of the parity of bank currency with gold; the function of bank currency in exchanging goods; and the relation of bank credit to prices." (Moulton, 1921, p. 381). In other words, the idea that commercial banking – money – is directly involved with production, remains unappreciated. This approach is thus far too narrow for a serious treatment of money and capital.

The third approach, to Moulton (1918), is an approach in which capital markets serve as intermediaries between borrowers and lenders or, more specifically, savings (savers/capitalists[50]) and investment (entrepreneurs). Yet Moulton (1918) is highly critical

[50] We, and others, have also used the term *resource providers* to signify the same meaning. However, resource providers may be ambiguous, since some might think of either (raw material) producers or *financial entrepreneurs*/*intermediaries*. To a certain extent, the term *capitalist(s)* suffers from the same ambiguity.

of the fruits that this approach had yielded up till that time. Banks, according to a too narrow conception of money and banking as "intermediation," merely "move" *capital goods* from one sector of the economy to another: "[I]t should be emphasized that these statements do not intimate any connection between commercial banks and the formation of capital goods; the view is that they merely transfer existing capital goods from one party to another." (p. 495). Moulton coincides with Mises (1949), who both argue that what matters is not the material capital goods, but capital in a broader sense (*net worth* to Menger). More importantly, Milton (1918; 1921) supplements Mises (1949), in the sense that he shows that financial intermediaries not only *move* capital, but are rather a decisive factor in the *formation of* capital, as financial intermediaries exercise considerable influence as to where and at what terms capital is allocated. This is where Moulton's (1918) third "conventional" approach to banking fails: "Some of the standard treatises point out that commercial banks act as intermediaries between borrowers and lenders, thus directing capital into the most productive channels. There is no suggestion, however, that commercial banking is in any way related to capital *formation*." (p. 508). Hence, Moulton (1918) rightly emphasizes that commercial banks and, more broadly, financial intermediaries are much more intimately related to commerce than often is appreciated.

However, Moulton (1918) commits a mistake when he argues that consumption precedes production, since "without consumption no one will go to the trouble of producing" (p. 599). Watkins (1919) points out Moulton's error. Moulton argues that consumption has to increase to make capital formation possible. Entrepreneurs would not invest capital without (expecting) increasing consumption. Moulton seems to ignore the fact that entrepreneurs invest capital to respond to *future* consumption, not merely *present* consumption. The very act of saving is, in fact, postponing present consumption (Mises, 1949). As such, banks transfer capital *intertemporally*. They take savings that depositors do not wish to consume *today*, and invest them according where risk-adjusted returns are highest, so as to allow depositors to consume *tomorrow*. As this insight was not fully appreciated by Moulton (1918; 1921), he ended up defending the notion of shiftability, where troubled banks can count on "stronger" banks, to which assets can be *shifted* in exchange for cash. Moulton (1918; 1921), however, did not recognize that depositors actually express time preferences and that, thus, *shiftability* will be unsustainable since episodes of bank runs and banks' "balance sheet weakness" will tend to be clustered at specific points in time. Inherently illiquid assets cannot be made inherently liquid.

In the debates about banking *shiftability*, banks were quick to define liquidity as "marketability with price stability." This idea is very much akin to the modern-day bid/ask-spread and "market depth." (e.g., Mancini *et al*., 2013). In present-day terms, "market liquidity" is the term used for the broader salability of securities. However, one of the problems with such a definition "is that it defines the quality of the moment and not an inherent quality of the asset. It indicates merely a degree of probability (…)" (Berle & Pederson, 1934, p. 23). It depends on market conditions and is variable over time (the very fact is that, during crises, liquidity on secondary markets largely disappears). Despite all these obstacles, this definition is inherent to all views that support shiftability. Hence, we are able to observe an important relationship between the way proponents of shiftability defined "liquidity" in the 20th century and the way banking institutions and financial regulators define "liquidity" in the 21st century, after the famous 2008 "liquidity crunch."

Indeed, centuries later the lack of a clear definition of liquidity continues to haunt the financial sector. The Security Exchange Commission (SEC) released "Rule 22e-4" that requires every open-ended mutual fund to establish a liquidity scheme, indicating liquidity risk. The objections from the industry were almost entirely geared toward the regulator's definition of liquidity. Hence, the importance of clarifying the concept of liquidity, both from an economic as finance perspective, is paramount.

All in all, there is an undeniable relationship between *liquidity* and *duration* (weighted maturity). We could in fact say that they are two sides of the same coin as mere antonyms. From the perspective of an economic system as a whole, the longer the duration of a society's productive assets[51], the less liquid they are, and the greater the problems of liquidating such assets (ultimately to final consumption). Hence, savers (resource providers) pick *durations* of a certain kind (which is, ultimately, what the theory of time preference pretends) according to their relative *liquidity* preferences, precisely the point many defenders of shiftability (e.g., Moulton, 1921) missed.

To sum up, the debate on *shiftability* was, in fact, on what we nowadays call *maturity mismatching* of commercial banks. According to the proponents of the shiftability theory,

[51] Here we see some positive contribution by the Böhm-Bawerkian and Hayekian legacy of the "production period," albeit formulated in (1) a financial rather than a material sense and (2) an income rather than a consumption sense.

banks' maturity mismatching is no issue, as long as a bank's long-term investments have a liquid secondary market. According to its opponents, the notion of liquid secondary markets that underlies the *shiftability* theory is entirely misguided, since in a crisis there are generally not enough strong banks and an excess of “weak” banks. Consequently, in a crisis illiquid assets cannot be shifted from weak to strong banks. However, what opponents of the shiftability theory did not consider, was the notion that maturity mismatching might lead to an endogenous business cycle, that is, shiftability plants the seed of its own destruction. Hence, not only were opponents of shiftability (the so-called “liquidity” school) right about the dangers of maturity mismatching in the sense that it makes commercial banks and the banking system *more fragile* to outside shocks, but shiftability also plants the seeds of its own destruction, in the sense that maturity mismatching lowers long-term interest rates that set in motion the very process that inevitably ends in a financial or economic crisis[52]. The latter, unfortunately, was never discussed or recognized by the “liquidity school,” mostly because the economists and bankers debating shiftability were not simultaneously engaged in the economic debates on capital theory and the business cycle. One of the motives is, we could speculate, timing. The bulk of economic debates on the business cycle began in the late 1920s and 1930s. Most of the debates on *shiftability* were pre-Great Depression and coincided largely with the founding of the Federal Reserve in 1913. Another possible factor that explains why both debates were rather isolated from each other is language. Most pre-1930s work on the business cycle was in German, especially since Böhm-Bawerk would write in German and his main adversaries and advocates (J.B. Clark, Wicksell, et cetera) could understand German. In contrast, the principal participants in the debates on shiftability were Americans, communicating in English, such as the *new institutionalists* (Glock, 2017).

[52] We will discuss the mechanics behind maturity mismatching and how it is a lead cause of recessions later. See p. 409.

Chapter 3: The Second Round (1925 – 1942): Keynes versus Hayek and the Kaldor-Knight Controversy

An important thing to note, at this point, is that the origins of the debates on capital theory have been surprisingly different. In this second round, which we can historically pinpoint from roughly 1925 to 1942, most of the research and discussions on capital theory were a consequence of the debates on the *business cycle*, sparked by the onset of the Great Depression. The first round of controversies was mostly concerned with the *distribution of income*, especially with regard to how wages compare to profit and interest. The third round, which is nowadays known as the famous "Cambridge Controversies" (with the groups divided between Cambridge, U.S. and Cambridge, U.K.), was mostly a product of discussions regarding *economic growth models*, mainly used for economic forecasting.

3.1 The Second Round in Historical Context

One of the key actors in this second round of debates on capital theory is, without a doubt, Nicholas Kaldor. Kaldor was a student under Hayek at Cambridge and even (co-)translated one of his books, *Monetary Theory and the Trade Cycle*, into English (Huerta de Soto, 2006). His first contribution was his review of the Knight/Hayek controversy on capital, with a brief reference to the earlier and preceding Clark/Böhm-Bawerk controversy (Kaldor, 1937). In this essay, Kaldor (1937) responded to Knight's critiques on Hayek.

However, Kaldor ended up largely dissenting from Hayek on capital. In a rather curious chain of events, when Kaldor (1937) published his response to Knight, he began noticing his own differences of insight with and the apparent flaws in Hayek's capital theory, which sparked two other essays by Kaldor on capital and the trade cycle, in effect criticizing Hayek (e.g., Kaldor, 1939).

This second controversy on capital theory, which was mainly a controversy between Hayek, Knight and later Kaldor, took place against the backdrop of the successful publication of J.M. Keynes' *General Theory*. Even though Keynes paid surprisingly little attention to these debates on capital theory, Hayek did debate Keynes on the business cycle. One of Hayek's most important instruments of choice was his capital theory. Indeed, in many instances Hayek loathed Keynes for his lack of an exhaustive and explicitly articulated capital theory. Irving Fisher also published an important work on capital and interest in more or less the same period. Like Keynes, Fisher also lost a fortune on the stock market after famously stating that the stock market had reached “a permanently high plateau” (Fisher, 1929). Fisher would never recover from his losses, whereas Keynes did.

The historical context of this round of correspondence is of vital importance as well. On the one hand, the Federal Reserve System was established in 1913. In the run-up to the founding of the Fed, the U.S. banking system suffered from many panics and bank runs, which sparked a debate between mostly bankers (and government officials) on “shiftability” and “liquidity.” These controversies culminated in the founding of the Fed in 1913, but the debates continued far into the future (e.g., Palyi, 1936; Mitchell, 1923; Moulton, 1921). These monetary debates are highly relevant, as we have already seen the role of money as intermediary of capital (or, more appropriately put, transmitter of savings into capital). It is unfortunate that the debates on capital theory and the trade cycle, mostly among economists, were completely ignored in the debates, mostly among bankers, on shiftability and liquidity and vice versa[53].

On the other hand, the world economy was struck by its most severe crisis on record, the Great Depression. The Great Depression began in 1929 and led J.M. Keynes to lose a fortune on the stock market. Keynes then published his *General Theory of Employment, Interest and Money* in 1936 in response to the Great Depression and the alleged failure of neoclassical economics to account for short-run disequilibrium as opposed to explaining long-run equilibrium. Prior to the publication of the *General Theory*, the discussions that took place in the beginnings of the 1930s mostly involved exchanges between Keynes-Hayek and Sraffa-Hayek (Piero Sraffa was a disciple of Keynes who was invited by the latter to partake in the debate) and coincided roughly with the publication of Hayek’s *Prices and Production* (1931). As we have mentioned earlier, the origin of the debates on capital theory in this second round should be traced back to the debates on the business cycle, which were of vital urgency given the Great Depression. Keynes’ work ended up having a larger impact on economic science than Hayek’s work. As White (2012) notes, “Keynes’s theory quickly caught on among younger economists and completely eclipsed Hayek’s theory” (p. 150).

At the end of this round of ‘controversies on capital’, Ludwig von Mises finally published his work on capital theory, which was embodied in his magnum opus *Human Action* (1949). That Mises (1949) took a very Mengerian approach to capital, despite being swayed by the alleged

[53] We have discussed the debates, at the beginning of the 20th century, around shiftability on p. 121.

double nature of capital, was a pleasant surprise. However, much of it came too late to correct for the mistakes of others, including Böhm-Bawerk. Moreover, Mises does not directly address others in his seminal work *Human Action* and never traces errors back to their corresponding originators[54]. Mises' contributions to capital, and his attempt to rescue Menger's contributions from oblivion, should be looked to as a critique on the Böhm-Bawerk line of thought. However, Mises' work on capital was largely ignored, even when the debates entered a third round which we will discuss further below. Moreover, his ability to effectively engage in debate would be severely hampered by the onset of the Second World War anyway. A year after publishing the predecessor of *Human Action*, his German *Nationalökonomie* (1940), the Second World War began. *Human Action* was published after WWII, at a time the interest in the debates on capital theory largely faded. Similarly, Ludwig Lachmann (1941) joined the controversies on capital rather late and failed to have any impact on mainstream economics. Finally, one of Böhm-Bawerk's disciples, Richard von Strigl (1934), wrote a treatise titled *Capital and Production*. Yet, Von Strigl's impact on the debate was largely indirect, as Friedrich Hayek and Fritz Machlup were students of him. His work, however, fell almost entirely into oblivion and he died an early death in the 1940s, unable to participate in this round of debates on capital theory (Hülsmann, 2000).

3.2 Irving Fisher's List of Impressive Contributions to Capital and Interest (but Less Impressive Contributions to Money)

3.2.1 Fisher's Positive Contributions to Capital and Interest

Although Irving Fisher was already involved in the first round of debates (arguing back and forth with, among others, Eugen von Böhm-Bawerk), his 1930 treatise *The Theory of Interest* marked his entire contribution to capital and interest. Irving Fisher (1930), in contrast to a long list of earlier and later economists, was actually one of the first to distinguish the two sides that result in a market rate of interest in attempt to explain both the demand and supply side to interest rates.

Fisher made numerous other impressive contributions. Fisher's distinction between income and "capital value" is highly informative and as close to irrefutable as a pure theory gets.

[54] There is one exception to this rule: as you can read on p. 83, Ludwig von Mises directly criticizes Böhm-Bawerk for his theory on the "period of production."

Fisher (1930) depicted the relationship between capital goods, income and capital in the following scheme:

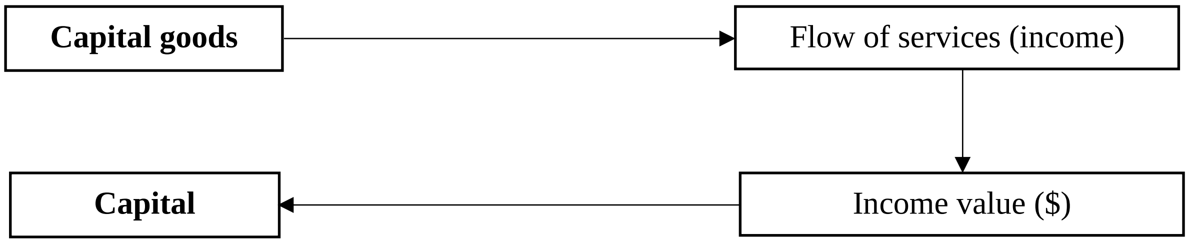


Figure 6: Fisher's (1930) scheme of capital

The classic distinction between labor, land and capital is completely irrelevant. Indeed, as Fisher (1930) explains, the price of a bond depends on: (1) its cash flows and (2) the interest rate at which these cash flows are discounted. Moreover, this principle applies not only to bonds, but to anything within the scope of capital: "It applies in any market to all property and wealth – stocks, land (which has a discounted capital value just as truly as any other capital), buildings, machinery, or anything whatsoever. Risk aside, each has a market value dependent solely on the same two factors, the benefits, or returns, expected by the investor and the market rate of interest by which those benefits are discounted" (Fisher, 1930, pp. 17-18).

Moreover, Fisher (1930) by showing that capital is the present value of income ("income capitalized"), shows where ultimately income originates from by his double-bookkeeping analysis. He concludes that "when we take the sum total of all income items for society, including psychic as well as physical items, this double entry results in cancelling out everything except the psychic items of enjoyment (…)" (p. 20). What Fisher (1930) in effect does, is expressing the 'Mengerian' insight by deriving the value of capital from, ultimately, consumers and their acts of consumption (what Fisher calls "psychic items of enjoyment"). In Mehrling's (2000) words:

> "Fisher's accounting system presents a unified picture of the economy as a stock of wealth moving through time, throwing off a flow of services as it goes. In Fisher's formulation all wealth is capital, not just machines and buildings, but also land and even human beings. Indeed for Fisher human beings are the most important form of capital because the most versatile. Thus, at the highest level of abstraction, there is no

distinction between the traditional categories of labor, capital, and land. All produce a stream of income (services) so all are capital, and their income discounted back to the present is their capital value. Similarly, at the highest level of abstraction, there is no distinction between the traditional categories of wages, profit, and rent. All are incomes thrown off by capital, hence all are forms of the more general category of interest, which is the rate at which income flows from wealth." (p. 5)

The degree of subjectivity inherent in Fisher's theory on capital is astonishing. In his writings, Fisher (1930) describes his notion of consumption in the following words:

> "The *total income* of a real person is his *enjoyment income* only provided we include the credits and debits of his own body. The physical music, or vibration which pass from his piano to his ear are, strictly speaking, only interactions to be credited to his piano and debited to his bodily ear." (p. 23)

On a funny side note (not often seen in academic writings on capital), Fisher (1930) remarks: "As a business man said to me, his pleasure yacht is capital and gives him dividends every Saturday afternoon." (p. 22). This notion of ownership of a (more or less durable) consumer good *does not imply* consumption, but rather that the actual enjoyment or consumption of its services *truly is* consumption, is an insight that has been overlooked in economics. Fisher (1930) provides another example of this insight: "A dwelling renders income to the owner who dwells in it himself just as truly as when he lets it to another. In the first case, his income is shelter; in the second, his income is rent payments in money" (p. 22). This is the only truly subjectivist concept of consumption. Capital, hence, is not only in the hands of business owners. Indeed, it is in the hands of virtually every consumer.

Moreover, Fisher (1930) provides us with another extremely valuable insight, especially in light of the later Cambridge controversies. Fisher attempts to dissect capital gains from income: "Capital gains, as already implied, are merely capitalization of future income. They are never present income." (p. 25). An increase in capital does not equal income, according to Fisher, in a strict sense. Even when it comes to interest on a savings account, argues Fisher, there is no actual "income" involved, but only an increase in capital. "Income may be *invested* and thus transformed into capital; or capital may be *spent* and so transformed into income. In the first case, as we have seen, capital accumulates; in the second case, he is living beyond his money income" (Fisher, 1930, p. 27). These ideas are beautifully embodied by the later work of Melchior Palyi (1936) on liquidity: apparently, some businesses and banks confuse capital gains for income. However, any business or bank that relies on capital gains is relying on a temporarily illusion. If present and future cash inflows (that is, income) are completely aligned with the maturities of liabilities (that is, present and future cash outflows), then an economy is inherently *liquid*. However, if present and future cash inflows are not aligned with the maturities of liabilities, for instance because banks or businesses are counting on capital gains as additions to present and future income, then an economy is inherently *illiquid*. Fisher's insight is revealing: if a 10-year US Treasury with a 4% coupon yield appreciates 10%, our annual income is still 4%, not 10%. Of course, we can sell the bond, in which case it *becomes* income, but this depends on the availability of a prospective buyer.

Additionally, Fisher (1930) criticizes the concept of a "supply and demand" for capital, which in modern times gained prominence especially in neoclassical circles (e.g., Solow, 1963). Capital, according to Fisher (1930), is "merely the translation of future expected income into present cash value, whatever supply and demand we have to deal with are rather the supply and demand of future income." (p. 32). This idea is directly followed up by a critique on the traditional fallacy that interest is a subcategory of income, next to rent and wages.

Curiously, Fisher (1930) then proceeds to apply this concept of capital as the present value of future income even to human labor (even though it is not customary in practice), showing the fallacies that underlie the traditional dichotomy between capital and labor: "The simple fact is that any or all income may be capitalized, including that credited to human beings, thus giving the resultant economic value of a man" (p. 34). In fact, Knight (1938) makes a similar point on human capital, as he describes the same idea by referring how the underlying principle still applies yet, ever since the end of slavery, is no longer customary.

In order to explain the rate of interest, Fisher (1930) abstracts away from risk, limiting himself to the explanation of a rate of interest under the assumption that all exchanges over time are risk-free, as well as that the unit of account holds a stable purchasing power over time.

First, Fisher (1930) attacks, very much in line with Böhm-Bawerk (1888), the "naïve" productivity explanations of interest:

> “The statement that "capital produces income" is true only in the physical sense; it is not true in the value sense. That is to say, *capital value does not* produce income value. On the contrary, income value produces capital value. It is not because the orchard is worth \$100,000 that the annual crop will be worth \$5000, but it is because the annual crop is worth \$5000 net that the orchard will be worth \$100,000, if the rate of interest is 5 per cent. The \$100,000 is the discounted value of the expected income of \$5000 net per annum; and in the process of discounting, a rate of interest of 5 per cent is already implied. In general, it is not because a man has \$100,000 worth of property that he will get \$5000 a year, but it is because he will get that \$5000 a year that his property is worth \$100,000—if the pre-existing rate of interest remains unchanged.” (Fisher, 1930, p. 26)

Here we see Fisher (1930) clearly espousing the ‘Mengerian insight’ that capital does not beget income for being capital. In fact, what gives capital value, is that it is able to generate future income. If we sum up all these future income streams and discount them to the present, we have something we can call “capital.” In stark contrast to, for instance, neo-Ricardian economic thought, income creates capital instead of the other way around. There are market incentives for entrepreneurs to bid up the price of capital as to leave no margin for any surplus income. Hence, the naïve productivity theories of capital are to be discarded completely.

In contrary to economists who try to explain the rate of interest exclusively through a theory of pure time preference (e.g., Mises, 1949; Kirzner, 1996), Fisher (1930) describes a two-sided theory of interest. Surely, time preference (or “impatience”) is an important factor, but it is in itself does a sufficient explanation. In fact, time preference only is one side of the equation, that is, the supply side, on intertemporal markets, but for the equation to be completed another side is needed, that is, the demand side. The demand side, thus, ought to be explained through, what Fisher (1930) calls, “investment opportunity.” This oversight in modern Austrian theories (e.g., Mises, 1949; Kirzner, 1996; Huerta de Soto, 2008; Gunning, 2005; Garrison, 1990) of interest is grave. Indeed, Fisher (1930) criticizes this one-sided explanation of interest aptly: “Some economists, however, still seem to cling to the idea that there can be no *objective* determinant of the rate of interest. If subjective impatience, or time preference, is a true principle, they conclude that because of that fact all productivity principles must be false.” (p. 64). And a few sentences later: “If, then, I am asked to which school I belong—subjective or objective, time preference or productivity—I answer 'To both.'" (ibid). Curiously, Fisher (1930) was often criticized for defending exclusively the subjective time preference school. While Fisher (1930) might be partly to blame, critics of Fisher's theory of interest (e.g., Brown, 1913) seem to have completely missed part of Fisher's explanation of interest. As Tobin (1985) claims, "A revised and enlarged version was published in 1930 as *The Theory of Interest*. One motivation for the revision was that Fisher's many critics apparently did not understand the 1906 version. They typically concentrated on the "impatience" side of Fisher's theory of intertemporal allocation and missed the "opportunities" side." (p. 686). However, uniting the "impatience" side with the "opportunities" side was precisely one of Fisher's key contributions to the theory of interest. Fisher (1930) himself noted that both "schools" (subjective time preference versus objective productivity explanation of interest) criticized each other for not taking into account the other (supply or demand) *side* of the equation and thinking they refuted the other side by pointing out such ommission:

> "[A]ny attempt to solve the problem of the rate of interest exclusively as one of productivity or exclusively as one of psychology is necessarily futile. The fact that there are still two schools, the productivity school and the psychological school, constantly crossing swords on this subject is a scandal in economic science and a reflection on the inadequate methods employed by these would-be destroyers of each other. Each sees half of the truth and wrongly infers that it disproves the existence of the other half. The illusion of their apparent Incompatibility is solely due to the failure to formulate the problem literally and to count the formulas thus formulated." (p. 104).

With regard to time preference, Fisher (1930) remarks that: "(…) all time preference resolves itself in the end into the preference for comparatively early *income* over comparatively remote, or deferred, *income*" (p. 28). What is important to note, at this point, is that Fisher (1930) calls time preference a "psychological concept." Although he may be partially right, is was Mises (1949) who later showed that time preference is a logical concept inherent to all human action, since any opinion to the contrary would involve admitting that people would postpone consumption indefinitely. Since people never postpone consumption indefinitely, time preference is inherent to all human action over time. There is always some time preference implied in human decision making.

Fisher (1930) also shows the relationship between spot and futures prices, as well as interest rates between different markets. Take his following visual description of the interaction between interest and prices (p. 29):

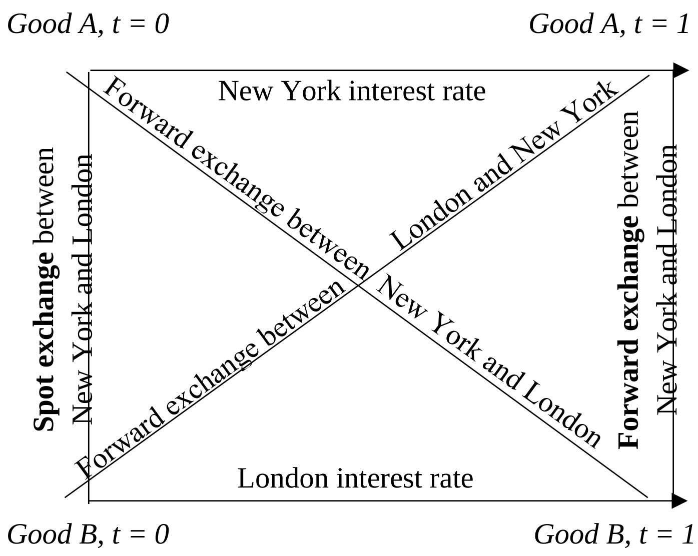


Figure 7: Fisher's (1930) interest rates between different periods compared to exchange rates between different places

Fisher (1930) shows here various important facts in one single chart. He shows that:

(1) There are forward exchanges between good A and B, where a present good A (or B) is exchanged for a future good B (or A). This is an apt description of what occurs on futures markets (or forward contracts) for physical goods.
(2) There are exchanges where a present good A is exchange for a future good A. This could be, for instance, a credit, or an exchange between $100 today against $100 in a year.
(3) There are exchanges where simply spot goods are exchanged, for instance a present good A for a future good B.
(4) There are exchanges where promises (forward exchanges) are swapped against each other.

Hence, as Fisher (1930) summarizes, there exists a "time to time factor" and a "place to place" factor and there are important interactions between spot and futures markets.

In addition to this impressive treatment of the theory of interest, Fisher (1930) dives into risk, an element largely ignored by his peers. He shows that he assumes that exchanges (especially forward exchanges or, in other words, credit) are riskless. However, he readily admits that they are not. Risk, and trust, can therefore influence the time preference of an individual.

The idea of capital as being the present value of future income, is taken even further by showing how changes in the rate of interest affects the present value of the various options an investor is confronted with when he either invests or decided what use to give to his production resource (Fisher uses the example of land):

Options	**Present value at**		
	5%	**4.5%**	**4%**
For forestry	$8,820	$9,920	*$11,300*
For farming	$9,000	*$10,000*	$11,250
For mining	*$9,110*	$9,280	$9,450

Figure 8: Three "investment" options open to an owner of land at three different rates of interest from Fisher (1930)

A higher rate of interest makes income streams with shorter-term returns more profitable (attractive), whereas a lower rate of interest makes income streams with more distant returns more profitable (Fisher, 1930). In a market economy, individuals have the incentive to choose the particular income stream that has the highest present worth according to the rate of interest.

Returning to the other side of the interest equation, Fisher (1930) cites Böhm-Bawerk (1888) as being "profoundly right": "The statement of how the productivity of capital works into and together with the other two grounds of the higher valuation of present goods, I consider one of the most difficult points in the theory of interest, and, at the same time, the one which must decide the fate of that theory" (p. 277). This is what Fisher consequently intended to achieve.

Time preference does not suffice for a complete explanation of the interest phenomenon. Rather, what lacks, is the demand side in intertemporal exchanges. The demand for present goods does not depend on physical productivity, but rather *investment opportunity* (Fisher, 1930).

What Fisher effectively attempts, is to separate a rate of profit (or rate of return) from the rate of interest. In doing so, he evokes a modern-day finance concept where investors compare returns on invested capital (ROICs) against weighted-average costs of capital (WACC). As Fisher explains in his own words, an investor tries to pick the project that yields the highest surplus income over the rate of interest: “Out of all possible options open to a person that particular one is selected, the comparison of which with any other option affords a rate of return over cost equal to or greater than the rate of interest.” (p. 57).

Nevertheless, it remains unclear to what Fisher (1930) actually attributes the existence of this rate of return (return over cost in Fisher’s own terms). He seems to imply some kind of superior arrangement or way of combining goods, or innovation, but never explicitly explains the origin of such returns; he merely assumes their existence in various examples (“The range of man's investment opportunity widens as his knowledge extends and his utilization of the forces and materials of Nature grows. **With each advance in knowledge come new opportunities to invest**. The rate of return over cost rises. With the investments come distortions of the investors' income streams.” (p. 112) [emphasis mine]. Any “important discovery,” according to Fisher (1930), then leads to an increase in the rate of interest as investment demand rises. This overlooks the fact, however, that in many cases “important discoveries” are made possible *because* of low interest rates. That is, lower rates of interest allow greater opportunities to innovation or make “important discoveries” in the first place (Mises, 1949). Indeed, investments in human capital, discoveries (innovation) or, more broadly, knowledge, behave very much like any other long-term investment. The discovery and implementation of any new applied knowledge require generous amounts of time, which would easily turn into unprofitable and frivolous endeavors at higher rates of interest.

In sum, while Fisher as one of few economists explains interest as a result of supply (time preference) and demand (rate of profits), he fails to identify the true origins of the latter.

Another important contribution of Fisher (1930) is his explanation of how supply and demand on intertemporal markets “balance” each other out. If demand for present goods rises, there exists a countertendency of diminishing returns (diminishing rates of profit). If demand for present goods falls, rates of profit go up. If the supply of present goods rises, rates of profit go down. If the supply of present goods falls, rates of profits go up. Fisher (1930) summarizes such countertendencies, or negative feedback mechanisms, on capital markets in the following way:

> "If the pendulum swings too far toward the investment extreme and away from the spending extreme, it is brought back by the strengthening of impatience and the weakening of investment opportunity. Impatience is strengthened by growing wants, and opportunity is weakened because of the diminishing returns. If the pendulum swings too far toward the spending extreme and away from the investment extreme it is brought back by the weakening of impatience and the strengthening of opportunity for reasons opposite to those stated above." (p. 62)

In various other contributions, even citing Carl Menger, Fisher (1930) shows how the price of any financial asset possesses a liquidity premium: "A man who keeps an average cash balance of $100, rather than put his money in a savings bank to yield him $5 a year, does so because of its liquidity." (p. 73). Moreover, Fisher (1930) notes that capital cannot be a mere input to a material production process. As markets function by arbitraging away profits, capital is revalued downward as well. As a consequence, capital cannot be a reflection of a society's wealth, contrary to what many economists have argued, including Mises (1949). Capital is a *reflection* of the productive process, not an input to. A high "capital value" might indicate maladjustments, rather than capital accumulation, which rather *incentives* capital accumulation and investment.

3.2.2 Fisher's Negative Contributions to Monetary Theory

Fisher's contributions to monetary theory were less impressive. He mostly adhered to a fallacious quantity theory of money, unable to bridge capital and monetary theory. While his attempt to achieve a stable unit of account is praiseworthy, his theoretical work on money shows many gaps. Indeed, Fisher (1911) is largely remembered for his famous MV = PT equation, a product of the quantity theory of money. In this equation, M stands for "money supply," V stands for "velocity," P stands for "prices" and T stands for "transactions." Any increase in M or V would generally lead to an increase in P, that is, the general price level. As Huerta de Soto (2006) correctly criticizes Fisher's equation of exchange:

> "[T]he great majority of the theory's defenders have accepted the mechanistic equation of exchange which, at best, **merely represents a tautology: that the income and expenditure involved in all transactions must be equal**. Furthermore they attempt to supply a comprehensive explanation of economic phenomena by adding up the prices of goods and services exchanged in different time periods and assuming the value of the monetary unit is determined by, among other factors, the

> "velocity" of circulation of money. They fail to realize that the value of money originates with humans' subjective desire to maintain certain cash balances, and to focus exclusively on aggregate concepts and averages (…). Nonetheless economic agents' demand for money comprises both the cash balances they retain at all times, as well as the additional amounts they demand when they make a transaction. Thus money performs its function in both cases and always has an owner; in other words, it is included in the cash balance of an economic agent, regardless of whether the agent plans to increase or decrease the balance at any point in the future. According to Mises, another crucial defect of the equation of exchange is that it conceals the effects variations in the quantity of money have on relative prices and the fact that new money reaches the economic system at very specific points, distorting the productive structure (…)." (p. 531) [emphasis mine]

Indeed, the "MV" part of Fisher's (1911) equation should be an elaboration of the nature of the financial system (the financial titles to), whereas the "PT" part of the equation should be a reflection of the underlying productive assets and the consumption. Fisher's very own equation seems completely irreconcilable with his capital and subjective consumption theory.

Unfortunately, Fisher seems oblivious with regard to this logical relationship between the financial system (money and banking, in a broad sense) and capital. It is as if two Fisher's existed. At the very least, Fisher leaves a lasting impression that he completely separated both studies in his own mind.

In sum, Fisher made various impressive contributions to the theory of capital, whilst espousing at the same time a naïve and mechanical version of the *quantity* theory of money, with no apparent direct link between the two (except for inflation and inflation premia in rates of interest).

3.3 The Knight-Kaldor-Hayek Controversy: Was Frank Knight Right about Capital Theory?

3.3.1 An Introduction to Knight's Capital Theory

Frank Knight began developing an interest in capital theory in the 1930s. He set out to refute both the Böhm-Bawerkian theory of capital and the time-preference theory of interest, which began with a response to Irving Fisher in 1931 (Stigler, 1985).

Frank Knight (1931) actually defended a very modern *net present value* approach to investment decisions. He takes the present value of the cash inflows of a project and the present value of the cash outflows of a project, and concludes that an investor, if he were to behave "economically" as Knight would say, chooses the project with the highest *net* present value.

As Northrup Buechner (1976) argues, Frank Knight would consider capital as the *only* factor of production. What Knight (1931) meant, in effect, was that both labor and land were to be seen as part of capital, rather than separate factors. Indeed, as Knight (1931) himself puts it, "The form of the capital item, like the nature of the income yielded (…) **is immaterial to the theory of capital and interest**." (p. 262). It follows then that *capital*, as Knight (1934) explains, is rather a tool of economic calculation: "(…) the phenomenon of capital accounting inevitably arises, giving expression and precise form to the economic rationale of all activities which involve saleable productive resources. (…) The operation of capital accounting converts every saleable productive resource into a pure quantity of "capital." The theory of capital accounting (…) is virtually the sum and substance of the theory of "interest" as a *rate* of return on *capital*." (p. 258). Knight (1931) then sums up what he calls the *essentials* of the theory of capital and interest:

(1) The "amount" of capital is simply the net present value of future yields discounted to the present at a *uniform* rate of interest;
(2) The "amount" of capital bears an important relationship to its "construction cost" (replacement cost), which is equal to expressing our fundamental "WACC = ROIC" truth that forms the basis of our later *q*. Further down, as Knight (1931) repeats his position, he says that: "(…) in so far as [entrepreneurs] have correct knowledge and direct their capital creation in accord with it, **cost will be equal to capitalization value, and the tendency of the two to be equal is the basis of the general theory of capital in a system in which it is being created under economic conditions**." (p. 278) [emphasis mine].
(3) According to Knight (1931), the *construction cost* would in equilibrium simply equal the *present value of the anticipated yield*, as long as the rate of discount of the outlays to create the new item does not deviate from the rate of discount at which future yields are discounted.
(4) The rate of discount (or rate of interest) is the maximum possible under any given *technical* circumstances (remember, Knight adheres to a productivity theory of interest)

and *economic* circumstances (the expected differential between present and future prices of products). This presupposes: (a) freedom of choice between "investment opportunities" and (b) a free market for "capital items" (the underlying productive assets) and "capital" (claims of various quantities and durations).

Then Knight (1931) compares his *net present value* theory to the earlier attempts of Jevons (1871) and later Böhm-Bawerk (1888) and Wicksell (1898) to define capital as a "subsistence fund" or "wage fund" that allows to maintain workers[55]. In essence, Knight (1931) tries to reconcile his own capital theory with Böhm-Bawerk's and Wicksell's. Without adding much of an argument, he comes to the conclusion that both views are similar. Any *increase* in the duration of capital must take into account wage expenses and present consumption. The "production cycle" must be matched with the "consumption cycle." A longer production cycle requires "a larger subsistence fund" (p. 262). Yet he never makes clear whether this is the defining *feature* of capital. As we will see later, it is far from beining important to or defining for our notion of capital. Since savings (that is, a curtailment of present consumption) is a prerequisite of capital formation, a "larger" subsistence fund is an automatic (secondary) result of the very act of savings, instead of a "larger" subsistence fund allowing for a "longer production cycle" as if it were a (primary) *cause*. As we have seen before, this is the world upside-down, akin to Böhm-Bawerk's (1888) and Wicksell's (1898) digressions.

Whereas Knight's (1931) theory of capital is noteworthy, Knight's (1931) theory of interest is fraught with flaws. Knight (1931) essentially explains interest as a result of the "productivity of capital." For some reason, there is always some excess return, which is, unless other returns, not arbitraged under his assumption of "perfect competition." What Knight (1931) assumes as perfect competition is in fact not perfect competition. Under perfect competition any excess return, if due to mere productivity of capital, would be arbitraged away (Mises, 1949). Yet, according to Knight (1931), "(…) it is obvious that the rate of interest, or productivity of capital, could never reach zero, since there is rarely any ultimate limit, even in an individual industry, to the possibility of increasing output by further investment." (pp. 283-284). While equilibrium exists in commodity markets, in the case of capital and interest there is "no equilibrium price which has any meaning." (p. 284). However, with long-term

[55] Knight (1931) would describe Böhm-Bawerk's attempt as a popularization of the subsistence fund theory "in a muddled form" (p. 262).

increases in capital (marginal return on capital), the rate of interest tends to go down, albeit never reaching zero. While Knight (1931) constantly assumes a state of equilibrium in espousing his capital theory, he suddenly reverses course and decides that such as an assumption is “fallacious” and “inadmissible.” Capital begets productivity and therefore future income, which in turn makes it easier to save and thus creates a positive spiral which leads to a constant fall in the rate of interest. Moreover, there is no valid argument to back the claim that more investment, according to Knight (1931), leads to investment in more *durable* goods or a *lengthening* of the production period, directly referring to Hayek’s work. The glaring contradictions in Knight’s work are sometimes unexplainable. If Knight defines capital as the net present value of future income, discounted at a rate of interest, then why does he not grasp that by discounting future income at a lower rate makes income further out in the future *relatively* more attractive than at a higher rate? *Duration* is of great importance, which is something that Knight (1931) completely misses, since he explicitly rejects *any* role for time in the theory of capital and interest.

On the other hand, as the entrepreneur (and later the financial entrepreneur) become focal points in our discussion on capital, Knight (1921) also came to a rather novel approach to entrepreneurship. According to Knight (1921), an entrepreneur is someone who “bears uncertainty” since the future is fundamentally unknowable. One of the consequences of this view is that entrepreneurs can never have a coordinating role and profits or returns are largely the result of some type of “random walk” (sheer, unpredictable luck). As Howden (2009) writes: “[T]he Knightian view of uncertainty excluded the entrepreneur from having any sort of coordinating effect. The “fog” of the future is so thick that it is purely unmanageable.” (p. 105). However, this directly contradicts Knight’s (1931) equilibrium assumption in his capital theory, which *requires* coordinating entrepreneurs. He explicitly mentions in his elaboration of his capital theory that he assumes that “the investor behaves ‘economically’” (p. 266). Hence, Knight (1931) would be unable to reconcile his capital theory with his theory of the entrepreneur (and, subsequently, his theory on risk and uncertainty)[56].

[56] We will discuss Knightian risk to greater length on p. 201.

3.3.2 The Knight-Kaldor-Hayek Debate on Time Preference, Interest and the Period of Production

As we have seen in the previous paragraph, Knight (1934) was well ahead of his time by using modern-day finance theory to demonstrate the nature of capital. Specifically, he shows how the value of capital is equal to the present value of all the future income that capital is able to yield: “The amount of capital "in" any item, i.e., the theoretical sale value of any productive instrument or property item at any moment, is determined mathematically as a "present worth" by discounting its future yield (assumed to be known) back through time to the moment of valuation, at a uniform rate” (ibid, p. 260). In contrast, Hayek (1941) emphasized capital as a complex of heterogenous capital goods (“nonpermanent higher-order goods”) of different nature used in different “stages of production,” either closer or further removed from final consumption. Hayek (1941) maintained the same fallacious production triad of land (permanent resources), labor and capital (nonpermanent resources) as Böhm-Bawerk and the classical economists, but discarded the Böhm-Bawerkian idea of capital as a “subsistence fund,” which Hayek called “misleading.” Kaldor, who joined the debate at a later point, agreed with Hayek: capital is a factor of production different from other factors of production such as labor.

Knight (1934) especially took aim at Hayek’s defense of the Böhm-Bawerkian concept of the “period of production.” According to Knight (1934), consumption and production are simultaneous. Therefore, the “period of production” is irrelevant. Nevertheless, there is a major flaw in Knight’s own line of thinking, as he *completely* rules out any role for time. As Knight (1934) himself writes:

> “It should now be clear why the ‘length of the production process’ has nothing to do with the case, and in fact [has] no real meaning. As long as capital is maintained by replacing the capital goods, if their life is limited, by others of any form with equal earning capacity in imputed income, **the durability or service-life of the good is a mere technical detail.**” (p. 267) [emphasis mine]

In fact, Knight (1934) states that “[t]he primary reality is income, a rate of service-value through time, and in a society which is not planning for the end of all things, **all property income is perpetual”** (p. 268) [emphasis mine]. As such, Knight (1934) follows in the footsteps of Clark (1899), who also rejected the role of time in capital (remember that abstinence was, according to Clark, only necessary to “give birth” to capital, after which it was “permanent” and “perpetual”). Knight (1935) would later repeat his criticism of Hayek’s “period of production” in another rebuttal: “[The] quantity of investment is (…) unconnected with any production period. (…) It is extremely difficult to give any intelligible meaning to a “period of production,” and it certainly has no meaning of the sort assumed in the Böhm-Bawerk-Hayek theory of capital.” (Cohen, 2003, p. 469).

While there is some truth to Knight’s statement (especially in reference to Böhm-Bawerk), he fails to grasp the fact that not every dollar of “perpetual income” is worth the same in present dollars. That is, as interest rates for Knight do not reflect *intertemporal* consumption and investment preferences, he fails to understand the role of time in capital investment (the rate of interest is simply the rate of profit for Knight, produced by the quasi-automatic ‘productivity of capital’). Considering his reference to “*durability*,” we can assume that a more durable good requires a larger present outlay (investment) than a less durable good (that is, this good or combination of capital goods is more “capital-intensive”), yet is able to yield more services (and thus income) before another outlay is necessary to the replace the good. In brief, this implies that a more “durable” good has higher negative cash flows in the short run, but higher positive cash flows in the longer run than a less durable good (“more durable,” in this sense, simply equals a *higher duration*). As a consequence, an increase in the rate of discount would disfavor or deter investment in the more durable good compared to the less durable good (Cachanosky & Lewin, 2014). Put differently, an asset or an investment with a *longer duration*, **is more sensitive to changes in interest rates**. This is exactly the view that Hayek defends (Cohen, 2003). The period of production is not a “technical datum” and, moreover, the “backward-looking interpretation of the ‘period of production’ will always lead to absurd conclusions.” (Hayek, 1934, p. 227). Hayek (1934) breaks away from the Böhm-Bawerkian physical conception of the “period of production” (and “roundaboutness”) and adopts a forward-looking notion akin to financial *duration* (Cachanosky & Lewin, Roundaboutness Is Not a Mysterious Concept: A Financial Application to Capital Theory, 2014). Yet this is, at the same time, precisely the fact that Knight (1934) completely overlooks. To Knight (1935), consumption and production are simply simultaneous. Therefore, the production period of consumption is zero. We can only speculate as to the reasons why Knight is unable to grasp the nature of *duration* and capital: for starters, Knight seems to be unable to distinguish between cash flows and accounting profits (and expenditures). Yet his omission is rather disconcerting, especially since Fisher (1930), for example, largely avoids such inaccuracies. Nevertheless, as Cohen (2006) notes, Hayek

(1941) ends up agreeing largely with Knight (1935) on the notion of the (average) period of production: "All attempts to reduce the complex structure of waiting periods (…) are bound to fail, because the different waiting periods cannot be reduced to a common denominator in purely technical terms." (p. 143). We will see later that they can, by using financial *maturities* or *durations* (weighted-maturities).

Moreover, as Machlup (1935) pointed out, Knight appears to make the mistake of confusing the assumptions of a stationary or equilibrium model with real world conditions and is therefore unable the appreciate the role of time. That is, Knight seems to confuse the assumption of capital maintaining its value in society since, on an aggregate level, society does not engage in capital consumption. In fact, Knight actually went as far as calling his notion of perpetual capital as the "*fact* of perpetuity" (Knight, 1935, p. 15). Instead of using equilibrium to *understand* the world, Knight (1935) appears to fall in the trap of using equilibrium to *define* the world.

Knight (1931) also condemns Fisher's theory of interest. Fisher (1930) attempted to show that wealth (that is, for our purposes, capital) is merely anticipated real income in consumption satisfactions. Knight (1931) calls this a "half-truth", since it seems to him "indisputable (…) that people desire wealth for many reasons, of which the guaranty of the future delivery of groceries or other consumable services is sometimes the main and sometimes a quite minor consideration." However, according to Kirzner (1960) this would involve abandoning the "Mengerian" subjectivist legacy, which refers to the fact that *all* economic phenomena must be traced back to the subjective valuations of consumers, a position Fisher (1930) clearly defends. Mises (1949) did not try to refute or respond to Knight directly, but did write a rebuttal regarding this same point. Mises (1949) argued that the nature or motives of human action are irrelevant and should be left to other human sciences, such as psychology. The fact of the matter is that, one way or another, individuals do (either consciously or unconsciously) express time preferences: "the removal of future uneasiness are directed by the categories of *sooner* and *later*" (p. 483). Mises does not refer to "groceries" or "consumable services", but to the fact that human beings spread consumption over time and that no human being would postpone *all* his want satisfaction (that is, consumption according to its subjectivist meaning) to an indefinite future (Gunning, 2005). An absence of time preference, would imply that an individual would be "willing to postpone all his satisfaction to the indefinite future" (Gunning, 2005, p, 4).

Or as Gunning (2008) puts it: "[Mises] is making the point that *human actors are not indifferent concerning when they will consume goods*." (p. 7). As Gunning (2008) correctly asserts, the time preference theory of interest is often misunderstood because writers consider the consumption of present and discounted future *physical* or *material* goods. However, for Mises, a good in t=1 is different from t=2, even if its physical shape remains the same. Mises provides the example of ice in winter compared to ice in summer. It refers to the same physical object (ice) but to a different subjective context. Although from a physical point of view the object is the same, they are two different economic goods since they are means to a given, subjective end. "[T]he actors of the market economy exhibit time preference in the sense that they prefer *some* goods in the present and *some* goods in the future." (Gunning, 2008, p. 10).

In sum, for Knight (1931) the *source* of interest cannot be traced back to the decisions or preferences of consumers, but can (and should) be traced back to the physical productivity of capital (although with diminishing marginal returns). Yet this theory is rather unsatisfactory. Moreover, time has no role in Knight's capital theory. Hayek (1941) attacked both ideas. Yet Hayek (1941) himself was largely unable to avoid the same crude mistakes as one of his predecessors, Böhm-Bawerk (1888). Hayek (1941) ends up defending a material theory of capital, in essence distancing himself from Menger's (1888) subjective revolution.

3.3.3 Kaldor on Hayek's Theory of Capital

The debate between Hayek and his disciple Kaldor is important, since Kaldor's critiques of his mentor's work on capital is seen by many economists not just as a refutation of Hayek's capital theory, but as a refutation of the entire "Austrian" theory of capital. As Desai (1991) explains: "Kaldor publicly broke with Hayek in 1938, in his article 'Capital Intensity and the Trade Cycle' (CITC). By the time he wrote his 1942 review article of Hayek's Profits, Interest and Investment (PII), 'Professor Hayek and the Concertina Effect' (CE), the rupture was complete. Between 1938 and 1942, Kaldor repudiated not only Hayek but the entire apparatus of Austrian capital theory." (p. 54). Having established the importance of this debate, let us examine Kaldor's objections to Hayek's capital theory.

Desai (1991) argued that "Nicky Kaldor was the only economist in the 1930s who had the talent to bridge the gap between Hayek and Keynes", but later says Kaldor "killed capital theory" (p. 55). Kaldor was, indeed, one of Robbins' and Hayek's most promising pupils. He was well known with the work of Hayek, since he co-translated Hayek's *Paradox of Savings* and *Monetary Theory and the Trade Cycle* to English. In the 1930s, Kaldor (1937) went on to write an impressive review of the debate between Knight and Hayek on capital theory, which was published in *Econometrica*. Kaldor (1937) summarized Knight's critique on Hayek's work according the following three points:

1. It is impossible to distinguish between permanent and nonpermanent resources (or "original" and "produced" means of production);
2. It is irrelevant and, in many cases, impossible to distinguish – on an analytical and a physical level – between outlays to 'maintain' or 'replace' assets;
3. There is no relationship between the "production period" and the supply of capital (or "quantity of capital" as Kaldor (1937) calls it).

With respect to Knight's first argument, Kaldor (1937) argues that Knight criticizes the notion of "permanent resources" (that is, almost exclusively land) since "no type of natural resources truly possesses 'indestructible powers'; the best that can be expected is that the flow of services can be kept upon permanently by continued maintenance" (p. 204). Moreover, Knight criticizes the notion of "permanent resources" with the fact that the economic value of a permanent resource such as land would be exactly zero if not supplanted by the services of "nonpermanent resources". That is to say, in absence of the nonpermanent resource (for instance, machines and human labor), the permanent resource is no longer permanent in any economic sense. However, Knight (1931) later does attribute permanency (or perpetuity, as he prefers) to capital (but not to capital goods), because after an initial investment (say the beginning of times), the capital is kept intact since the depreciation of the capital good is accounted for. The physical shape of capital (that is, the capital good) could change for a variety of reasons, but the value of the capital (as a fund) is not altered. Howden (2009) agrees: "Hayek's mistake in so viewing a difference between permanent and non-permanent resources was in his confusion between physical and value productivity. For while it may be true that some resources retain their physical productivity over extended periods of time, value productivity is a completely different story. Value is never derived from some innate, absolute and time-invariant quality (…), instead value is a shifting flux that varies as consumers' needs change. (…) For what permanence would a diamond mine have if tomorrow the world's demand for diamonds was eliminated?" (p. 411). Indeed, in *Capital, Time and the Interest Rate*, Knight (1934) writes: "Even as regards the very first infinitesimal increment of capital to be recognized at the beginning of economic life on earth, it would be merely fanciful to assume that it was produced under economic conditions by pre-existing 'primary factors'" (p. 262).

Hence, we see Kaldor committing the same mistakes as Hayek when it comes to arbitrarily separating factors of production on the basis of physical characteristics. Kaldor (1937) ends up distinguishing capital as "producible resource" from labor and land as "non-producible resources." Moreover, Kaldor (1937) ends up using the Böhm-Bawerkian explanation of "roundaboutness" as a synonym of mere "capital intensity" which is inversely related to changes in interest rates, at least in his exchange with Knight[57]. To Kaldor (and to Knight) the flipside of roundaboutness is simply the notion of diminishing returns on capital. Kaldor´s (1937) work at this stage is a preview of what he would deliver later in the third round of capital controversies, siding with Cambridge, U.K. This material "capital intensity" leads, according to Kaldor, to diminishing returns on capital as an input. In his machine/slave/bread production function (where machines and slaves are capital inputs, even though later he "liberates" the slaves and turns them into wage-earning laborers), he shows that diminishing returns occur as long as one of the inputs is fixed and, moreover, that an "optimum" proportion between the inputs exist that maximize output (Y). As Kaldor is not concerned with value and prices, but with physical production (and his production function), he cannot come to grips with the notion of diminishing returns: "[T]he existence of diminishing returns always presupposes the existence of some 'fixed factor' as their cause (…)" (Kaldor, 1938, p. 175). In a production function with two inputs (starting with, say, one laborer and one capital good), you will out of logical necessity get diminishing physical output if one of the two inputs is fixed. That is, if there is only one laborer, every additional capital good will lead to a smaller increment in output.

However, Knight (1944) was trying to defend the view that capital, as a "fund," was the *only* factor of production, yet at the same time tried to defend that capital has diminishing returns. Therefore, Kaldor (1937) thought that he had refuted Knight by showing that any increase in capital investment has no diminishing returns since such investment is able to increase the quantity of *every* production factor. The *rate of interest*, which simply equals the *return on capital* in Knight's work, is as a logical consequence *independent* of any increase in capital investment/capital/savings. Kaldor, thus, found a way to trick Knight into a logical

[57] Kaldor would later abandon the idea of interest rates influencing capital intensity or the *roundaboutness* of production (Cohen, 2006).

contradiction: hence, merely the given level of technical knowledge could determine the rate of interest, instead of the supply of capital.

The solution that Knight (1944) worked out, boiled down to human capital. For Knight, “knowledge” became the fixed factor (Cohen, 2006). Diminishing returns would occur *until* some increase in knowledge occurs that upsets the current state of an economy[58]. Hence, in the long run, diminishing returns would be periodically offset (returns would be raised) by such exogenous “knowledge” shocks. This would stroke with Knight’s observation that long-term interest rates were (and are) actually quite stable over time, directly contradicting any theory on diminishing marginal returns on capital.

Since Kaldor was committed, in light of Hayek, to a material theory of production and Knight (e.g., 1944) appears plain confused by his own capital theory, none of the involved recognized that the “fixed factor” (or, really, factor*s*) are in reality the input and output prices that entrepreneurs on one side incur and on the other side earn. If we were to be in price disequilibrium, that is, there exist pure profits since input prices do not reflect the prices of final goods (Kirzner, 1960), yet no further changes would occur, *diminishing returns* would be the unavoidable result to the extent that entrepreneurs arbitrage away the price differential (that is, arbitrage away the profit). A first entrepreneur would bid up the input prices and sell (thus bidding down) output prices, reducing the profit margin. If a second entrepreneur steps in, and bids up input prices while bidding down output prices, the profit margin shrinks further. This process continues until the whole differential is arbitraged away. The last entrepreneurial effort earns a quasi-zero return, while the first entrepreneurial effort earns a higher return. Nevertheless, equilibrium analysis helps us here to uncover how the law of diminishing returns works. The “objective” and “fixed” factors that explain diminishing returns are price differentials or profit opportunities[59], as opposed to some fixed physical input that yields some optimal physical output at some given level of a variable physical

[58] This Knightian notion comes very close to the Schumpeterian entrepreneur, who upsets the production structure by introducing a new innovation or invention (using new knowledge).

[59] As we will see later, there is a relationship between the demand for savings and profit opportunities, which tends to encourage increases in savings and the resources of profit-seeking entrepreneurs when returns are high, and tends to discourage savings and restrict the supply of resources of profit-seeking entrepreneurs when returns are low. We coin the term “Fisher’s pendulum of returns” to signify this dynamic, which applies to other areas as well. For more, see p. 295.

input. Prices, as objective *data*, thus serve as fixed factors. In the real world, where variables change constantly, such reality is not easy to grasp, especially when you look for months straight to a production function with material inputs and output.

With regard to the second point, Knight (1934) explains that there is a difference between physically replacing capital goods and maintaining (the value of) capital. In capital accounting, Knight argues, allowances are made for maintaining and, eventually, the replacement of productive assets. To the economist, it is completely irrelevant whether or not the replacement is in anyway the physical equivalent of (in practice, in a majority of times obsolete or "worn out" assets are replaced with different assets) the previous asset. What matters, is the fact that the "income-earning capacity in society as a whole is maintained or increased" (Knight, 1934, p. 264). From this perspective, all capital is "perpetual." Only ownership can change. It is important to note that Knight, in this case, seems to ignore the variety of cases in which capital was destroyed, either by war, natural disasters or failing government policies. Indeed, since Knight (1934) follows a flawed theory of interest, as we have repeated earlier, he fails to account for *losses* precisely as a result of a decline in the discount rate. Imagine I possess a capital good, or combination of capital goods, C, which has a present worth of $1,000 dollars at a rate of interest (or discount) of 10%. Now, the market interest rate rises, which leads to a decline in the present worth of my capital. In effect, I can sell the asset(s) only at a loss, or incur a future loss at the time of replacement (assuming the price of the asset remains equal), which reduces my capital and subsequently my ability to effectively replace the asset. Obviously, the assets have not been destroyed; but they are less valued. Even on an aggregate level, any increase in the rate of discount, leads to a decline in capital (capital consumption), albeit more capital-intensive projects with longer durations will be affected to a greater extent. It is such relative marginal changes that truly matter.

Third, according to Kaldor (1937), Knight argues that when the rate of interest declines, production methods that are already in place will most certainly not become more roundabout in a material sense. In addition, again according to Kaldor (1937), the marginal investment in (new) production methods, is not by any means per definition "more roundabout". Mises (1949) adds to Knight's argument: "More often it [higher productivity] consists in the fact that they produce products which could not be produced at all in shorter periods of production. *These processes are not roundabout processes*." (p. 470) [emphasis mine]. In other words, Mises (1949) argues that a decline in the rate of interest makes it possible to achieve certain goals that were not feasible at a higher rate of interest, and does not necessarily consists of a greater quantity of physical output with an equal quantity of input, as Böhm-Bawerk's concept of "roundaboutness" might imply. According to Cohen (2006), Kaldor did not "(…) side squarely either with Knight or Hayek. Kaldor agrees with Knight in rejecting Hayek's production periods and roundaboutness for analysing business cycles" (p. 143), since they could be hardly of use for a "dynamic problem."

In other words, Knight's third argument against Hayek is based on the idea that an increase in savings, and thus the supply of capital, leads to a lower market rate of interest, which induces investment in assets (projects) with lower returns that were not profitable at higher interest rates. In essence, Knight is expressing the fact that there exists a tendency in a market economy for the return on capital (ROIC) to equal or tend toward the average cost of capital (WACC). Spitznagel (2011) would later label this mean-reverting relationship (ROIC = WACC) the "dao of corporate finance." Böhm-Bawerk would most certainly not disagree. Nevertheless, a semantic confusion seems to kidnap the debate: Böhm-Bawerk confused concepts when he began adding the "lengthening of the period of production" to the concept of "roundabout production." The key takeaway is that whenever the average cost of capital (WACC) decreases, projects and assets with a lower ROIC become feasible. Nevertheless, this does not imply that this marginal investment at a lower interest rate will be invested in more (materially) roundabout production methods.

Parting from his mentor, Kaldor (1939) seemed to disagree with Hayek on the concept of the "period of production" or "amount of waiting." Hayek's theory of the trade cycle revolves around an unwarranted "lengthening" of the period of production. Kaldor (1939), however, in response to Knight's criticism, appears to reject the construct of a "period of production" in favor of another concept, "It will be better, therefore, to drop the expressions "investment period", "period of production", or 'amount of waiting' altogether and substitute some less ambitious term, such as (…) the 'degree of capital intensity'" (p. 42). Indeed, as Cohen (2006) mentions, Kaldor accepts "Knight's criticisms of period of production concepts," yet "Kaldor defends roundaboutness on multiple grounds: as correlated with the quantity of capital and capital intensity, as equivalent to the law of diminishing returns, and as useful for providing explanations of both the process by which equilibrium is achieved and the interrelationships between distributive shares and capital intensity." (p. 144). Stigler (1985) argues that it is "fair to claim victory for Knight over his adversaries (including Hayek, Machlup, Lange, and Kaldor) on this score: the period of production concept, which had never been fertile in real applications of capital theory, has virtually vanished from the literature" (p. 8). This, however, is a mistake. Knight's criticism on the Böhm-Bawerkian concept of a technical period of production was spot-on. His mistake, however, was not to replace it with something better. The fatal flaw of Böhm-Bawerk's "period of production" is, in essence, its reliance on capital as a combination of labor and original factors of production and its backward-looking perspective regarding the value of capital (e.g., Rallo, 2014). With regard to Böhm-Bawerk's backward-looking view of capital, Mises (1949) writes:

> "[The period of production and the duration of serviceableness] are essential elements present in every act of reasoning that precedes and directs action. **It is necessary to stress this point because Böhm-Bawerk, to whom economics owes the discovery of the role played by the period of production, failed to comprehend the difference**.
>
> Acting man does not look at his condition with the eyes of a historian. He is not concerned with how the present situation originated. His only concern is to make the best use of the means available today for the best possible removal of future uneasiness. **The past does not count for him**. He has at his disposal a definite quantity of material factors of production. He does not ask whether these factors are nature-given or the product of production processes accomplished in the past. It does not matter for him how great a quantity of nature given, i.e., original material factors of production and labor, was expended in their production and how much time these processes of production have absorbed. He values the available means exclusively from the aspect of the services they can render him in his endeavors to make future conditions more satisfactory. **The period of production and the duration of serviceableness are for him categories in planning future action, not concepts of academic retrospection and historical research**. They play a role in so far as the actor has to choose between periods of production of different length and between the production of more durable and less durable goods." (p. 477) *[emphasis mine]*

Its modern-day equivalent, "*duration*", and the role of interest is, as Mises (1949) appreciates, a forward-looking concept, related to the timing of the expected cash flows that a capital good or a combination of capital goods is able to yield (e.g., Cachanosky & Lewin, 2014). That is the correct explanation of "roundaboutness" or "period(s) of production." Hence, with regard to the third point, Knight and later Kaldor were mistaken that no relationship between the "production period" and the supply of capital exists (we will see later below that when time preferences – the supply of savings at different intervals – partially explain the rate of interest, and the rate of interest influences the average maturity of the productive assets); in addition, Hayek's concession was unnecessary.

As a result, as we have briefly mentioned above, the third point goes hand in hand with Knight's biggest mistake: his inadequate treatment of interest. His theory of interest, as we have shown, is faulty. As a result, Knight falls in the trap of circular logic. He refers to capital as the present value of all future yields (with which we agree), but then explains the rate at which these future yields are discounted as the return on capital (with which we disagree). The return on capital equals the outlays and inflows that a given good or combination of goods is able to yield, expressed as a percentage. But in order to bring these cash flows to the present, one would, according to Knight, actually need that return on capital! Put differently, the value of capital depends on the interest rate, but the interest rate depends entirely on the value of capital. This contradiction is responsible for even the modern-day confusion surrounding capital. This circular logic formed, in fact, the basis of the third round of controversies on capital theory, which we will discuss further below (see p. 256). Knight's neoclassical heirs effectively abandoned Knight's key contributions (that is, labor, technology and land are not separate categories from capital, capital as a value fund) and preserved Knight's errors (that is, interest as a rate of return on capital and the rate of interest as a prerequisite to determine the value of capital). Even though Knight would no longer partake in the third round of controversies on capital, Kaldor would, and the debate would hinge on much of the same points that were already raised in the Knight-Kaldor-Hayek exchange.

3.3.4 A Critique of the Hayekian Triangle

The Hayekian triangle, depicted below, is in fact a continuation of the (mistakenly) alleged Mengerian but truly Böhm-Bawerkian legacy. Interestingly, the Hayekian triangle is based on Böhm-Bawerk's rings is depicted as a structure of heterogeneous capital goods arranged in various orders further or closer removed from final consumption (1931). In effect, Hayek took the worst of Böhm-Bawerk and moved further astray. As White (2016) notes: "Hayek in 1927 lamented the way in which Böhm-Bawerk's approach had excluded productivity explanations of interest." (p. xviii). The Hayekian triangle was one of the way to reintroduce, even partially, the combined notions of productivity and time.

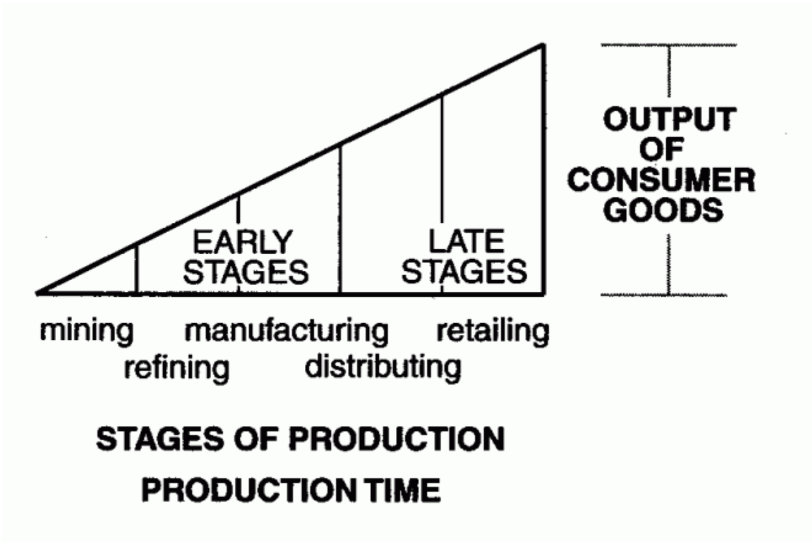


Figure 9: The Hayekian triangle representing a heterogeneous structure of higher-order goods (Hayek, 1931)

Curiously, with lower rates of interest, the Hayekian triangle would both flatten and lengthen, as pictured below. This flattening of the triangle is because present consumption (and stages near final consumption) falls, whereas there is an increase in investment represented by increases in early stages (such as mining, refining, et cetera). The lengthening of the triangle is due to the fact that stages even further from consumption that were not viable at a higher rate of interest have become viable. Hence, new early stages are added that are even further removed (in time) from final consumption.

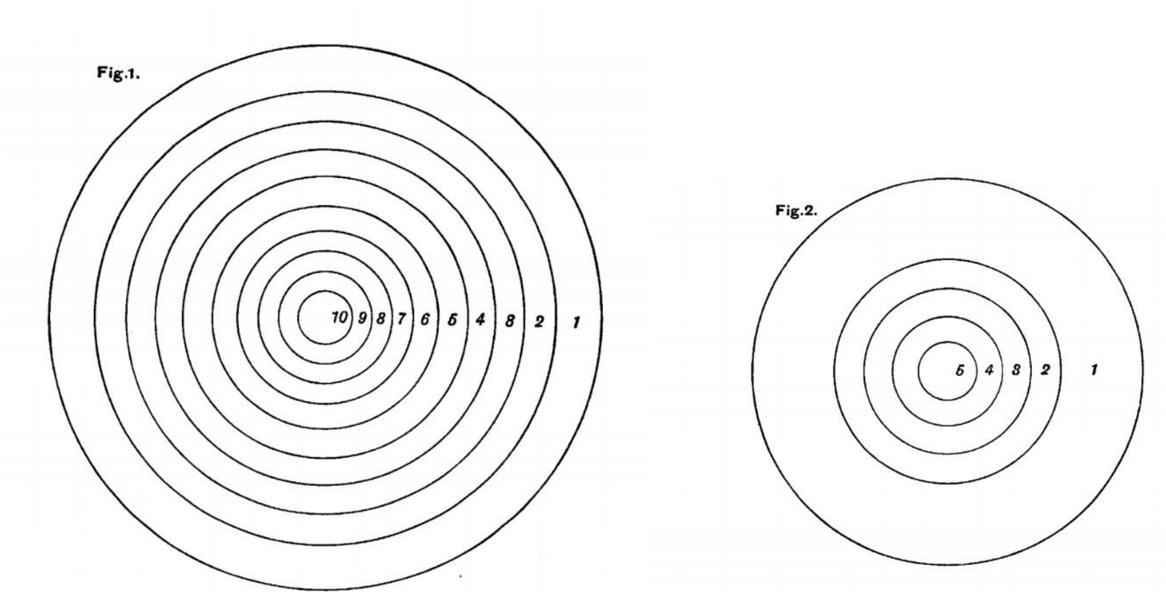


Figure 10: The predecessor of the Hayekian triangle: Böhm-Bawerk's bull's-eye model or rings to represent Böhm-Bawerk's physical capital structure. The final consumption product would equal the most outer ring (ring number one). The left circles would represent a more roundabout capital structure, the circles on the righthand a less roundabout capital structure.

That this model is completely wrong and misleading, should be clear after a short example. Let us assume that The Walt Disney Company is deciding whether or not to build a Disney cruise line. A Disney cruise line is highly capital intensive, as it costs as much as building an entire theme park, and is characterized by a relatively long *duration*. A cruise ship, once built, has an average life span of anywhere from thirty to fifty years. In fact, Disney´s first cruise ship, *Disney Magic*, is about to enter its 20th year of service. Building a cruise ship requires an investment anywhere from $350 million (the cost of *Disney Magic*) to about $900 million dollar (the cost of more recent Disney cruise ships) and takes approximately two years. Nevertheless, the ship, once built, is catering exclusively to final consumers. That is, it would be placed somewhere in the late stages of Hayek's triangle or the inner circles of Böhm-Bawerk's triangles.

On the other hand, an oil well would require a fraction of that initial investment, perhaps even 1/100. Moreover, an oil well has four stages that cover its full lifespan: early drilling and production stage, primary recovery stage, secondary recovery stage and a tertiary recovery stage. We assume that our oil well lasts (economically) for twenty years. However, just like the cruise ship, it takes two years to take an oil well from planning to production, or two years before the first barrel recovered is sold.

According to the Hayekian triangle, oil extraction, especially when the recovered oil is used in other early production stages (Garrison, 1990), such an investment is relatively further removed from final consumption than Disney's cruise ship. Yet, first of all, our oil well is less capital-intensive in the sense that the initial investment requires less outlays. Second, our oil well has a shorter maturity *and* duration than our cruise ship. Hence, a lower rate of interest would make, assuming all other things equal, an investment in Disney's cruise ship more profitable than an investment in our oil well. Nonetheless, if we were asked to squeeze in both projects according to the above visualized Hayekian triangle, we would put the Walt Disney cruise ship closer to final consumption than the oil well. Moreover, consumers will pay Walt Disney directly for its cruise ship services, whereas many intermediate stages of production are required before consumers are able to purchase the oil. In fact, our cruise ship might even consume the oil from our oil well! Hence, lower interest rates do *not* lead *per se* to a marginal increase in investment projects *further* removed from final consumption. Lower interest rates *do* lead to a marginal increase in investment projects that have longer durations and are more capital-intensive. The Hayekian triangle, if based on "proximity" to consumption (e.g., Hayek, 1931; Garrison, 1990), should be rejected.

Cachanosky & Lewin (2018) argue that Hayek's triangle is a mere pedagogical tool, which enabled Hayek to more effectively transmit and communicate the key insights of his business cycle theory. The triangle was apparently easier to grasp. Yet, Cachanosky & Lewin (2018) acknowledge that: "(…) the pedagogical benefit of the stages of production also brought some new theoretical problems. The first one is that there is no objective counterpart in the real world of the stages of production used in the Hayek-Garrison framework." (p. 6). Indeed, as we have concluded from our earlier example, there exists no correct way to categorize production stages according to proximity to (final) consumption. Many stages, according to Hayek's definitions, overlap.

Yet the idea of the triangle being an effective way to communicate Hayek's (endogenous) equilibrium business cycle theory should be questioned. As we have seen above, the business cycle theory does not depend on "production stages," since the concept of *duration* (of investment projects) applies to any business in *any* production stage, including late-stage production near the consumer level. Could a visual description possibly be called a pedagogical tool when it mistakes rather than simplifies an underlying concept?

Moreover, Cachanosky & Lewin (2018) argue that Hayek refers to mining and manufacturing (as "late stages") in *Prices and Production* (1931), which actually was a (re)compilation of a series of lectures, which again, entailed an oversimplification of capital theory so his audience would better understand Hayek's business cycle theory. Again, we disagree. In *The Pure Theory of Capital* (1941), Hayek mentions, for instance, that "(…) stage *a* would be **twice as far from consumption** as stage *c*" (p. 289) [emphasis mine]. Now, Hayek (1941) seems to grasp the underlying components of his theory, but at the same time confuses the issue by referring to "production periods" and "production stages" in the context of proximity to final consumption. What Hayek (1941) attempts to make clear, and at times explicitly explains, is that some assets are more durable than others: they have *longer* durations. For instance, a house with an expected lifespan of 10 years has a shorter *duration* than a house with an expected lifespan of 30 years. Capital goods are (or capital is) "converted" into consumer goods to the degree that they yield present income. Since capital is simply the present value of a future income streams: a more durable good requires more time to *materialize* the total sum of its expected income streams (income, in this sense can even be psychic; a house yields income as in psychic pleasure over time as we live in it and, moreover, there exists a direct opportunity cost – in this case the cost of renting). This income can thus be reinvested at certain maturities (effectively increasing or decreasing the *average* duration of the structure of production) or consumed (decreasing the *average* duration of the structure of production to a greater extent than if income were to be reinvested). In other words, Hayek (1941) sometimes seems to confuse the *timing* of cash flows with *real* consumption, yet the two really do not coincide since the recipient of the income has always the choice to reinvest *or* consume his income.

At any rate, what Hayek ostensibly has done, is turn the *net present value* approach to capital goods (including durable consumer goods) into a visual aid in which every "stage" simply represents a period with a diminishing present value over time as later periods are further removed from the present (in time to income, not to consumption). Hence, we arrive at the conceptualization of Hayek's triangle into a crude representation of something somewhat akin to Macaulay duration (Cachanosky & Lewin, 2014).

The fact that Hayek (1941), despite several confusions, adheres to this "income" concept (rather than the "consumption" concept) becomes clear when he attempts to "measure" and chart the roundaboutness of the U.S. economy. Hayek uses the inverse of the depreciation rate of business assets to calculate their lifespan in years and then charts each "maturity class" according to its contribution to annual output:

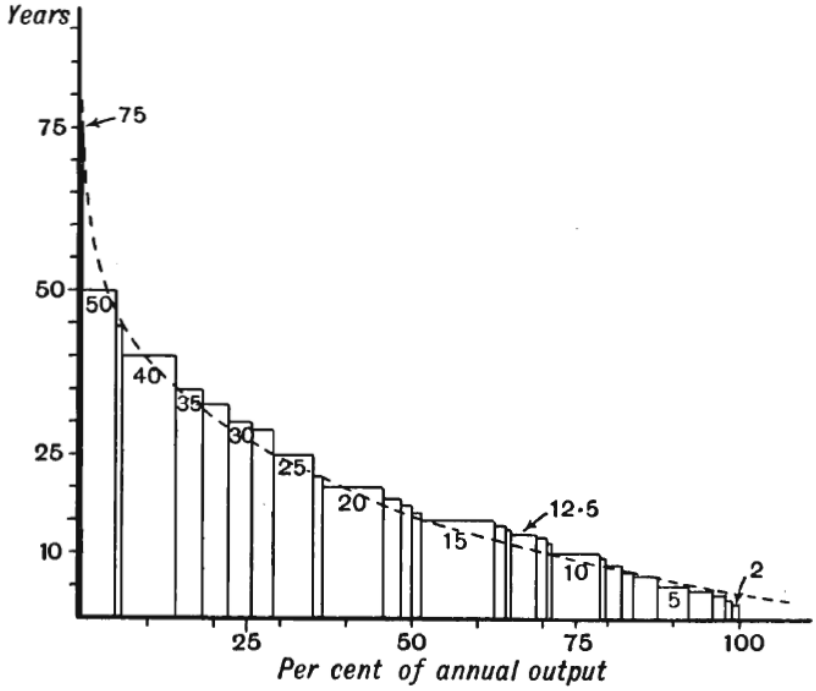


Figure 11: Hayek's attempt to chart an observed 'capital structure' from empirical data[60]. However, Hayek (1941) mistakes duration *for "production stages."*

Nevertheless, after doing so, Hayek (1941) objects that "the services rendered by the durable producers' goods may still be many stages removed consumption" (p. 133), repeating the same fallacy we have tried to refute above. This only serves to prove our point. Moreover, Hayek (1941) mentions that the data do not include durable consumer goods, which is a direct contradiction to his earlier statement but in terms of our analysis a valid point.

Let us take similar data, but a slightly different approach. I have taken long-term (10-year) government bond yields (US Treasury yields) from Shiller[61], adjusted for annual CPI inflation (source: St Louis Fed). Next, I have collected data from the U.S. Bureau of

---

[60] Data from Solomon Fabricant (1938), available at the National Bureau of Economic Research. Of course, since nowadays depreciation rates are often nonlinear (many assets are depreciated or amortized at an accelerated rate).

[61] I have used the annual series "long term stock, bond, interest rate and consumption data since 1871" provided by Robert Shiller. Shiller uses this series in his book *Market Volatility* (1989).

Economic Analysis (BEA) on the average maturity of fixed assets and consumer durable goods[62].

First, I simply ran correlations between the real 10-year bond rate and the average age of fixed assets and consumer durable goods, which gave me a negative correlation of *-0.32*. In other words, there exists a negative, historical correlation between interest rates and asset maturities: the lower the rate of interest, the longer the average maturity of fixed assets and consumer durables. Moreover, the negative correlation remains but diminishes when I include time lags. At a one-year lag (between change in bond rate and average age), correlation equals *-0.25*. At a two-year lag, correlation drops to *-0.22*. At a three-year lag, correlation reaches *-0.18*. Consequently, I plotted the data (the line is a trendline that represents the negative correlation):

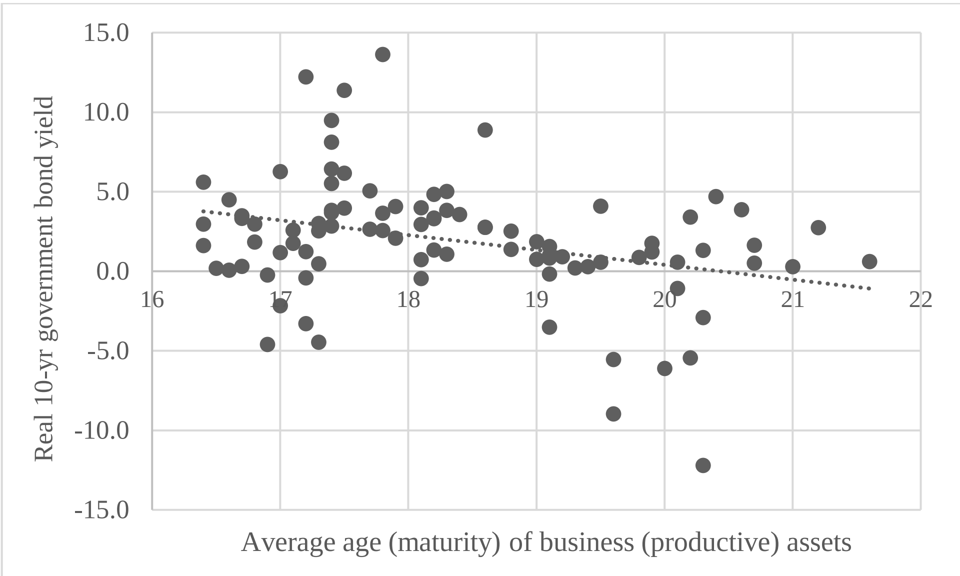


Figure 12: The average age (maturity) of fixed assets and consumer durable goods plotted against the 10-year bond rate adjusted for inflation.

We can observe that, in effect, data backs the statement that a decrease in interest rates leads to a lengthening of *maturity* (average age of assets). This backs the theoretical relationship between long-term interest rates and financial "roundaboutness" and mimics Hayek's own attempt to use data to illustrate the relationship between asset maturity and interest rates.

[62] More specifically, I have used the data series "Current-Cost Average Age at Yearend of Fixed Assets and Consumer Durable Goods" provided by the U.S. Bureau of Economic Analysis (BEA).

Now, I repeated the same procedure, but this time by comparing (real) interest rates with the average age of fixed assets, of durable consumer goods, of private nonresidential assets, and of private residential assets, each in isolation. Surprisingly, there was no marked observable correlation between interest rates and consumer durables (negative correlation of *-0.04*). Yet, the correlation was higher in the other cases: fixed assets *-0.30*, private nonresidential assets *-0.41*, and private residential assets *-0.45*. I have plotted, again, interest rates against the average age, for starters against the average age of consumer durables:

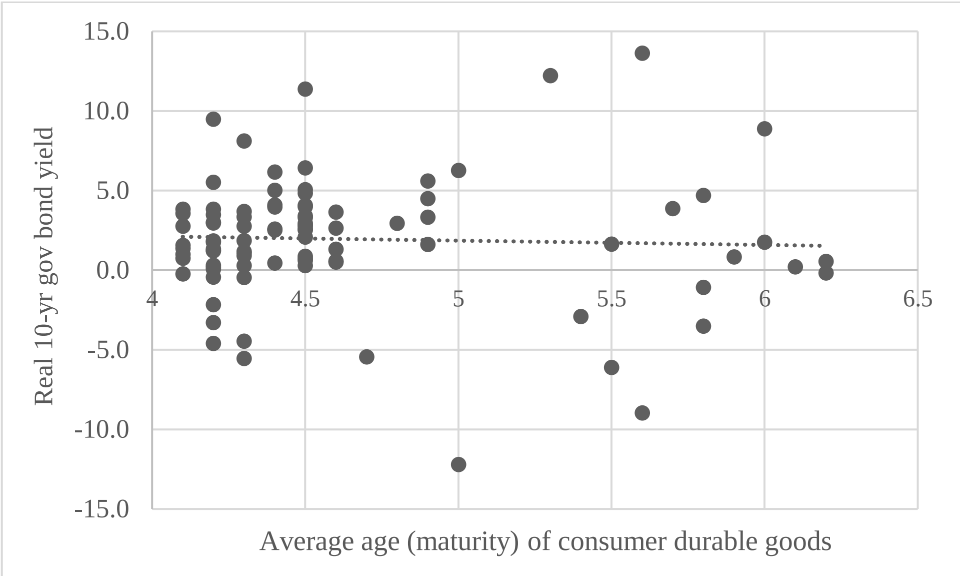


Figure 13: The average age (maturity) of consumer durable goods plotted against the 10-year bond rate adjusted for inflation. The data is all over the place; yet the range of average maturities is much less wide than in the other cases: observations range from a bit over four years to a bit more than six years.

Second, I plotted the (real) 10-year interest rate to the average maturity of private nonresidential assets:

Figure 14: The average age (maturity) of private nonresidential fixed assets, which includes intellectual property, structures and equipment, plotted against the 10-year bond rate adjusted for inflation.

Last but not least, I plotted the 10-year rate against residential fixed assets (effectively residential real estate), where the strongest negative correlation can be observed:

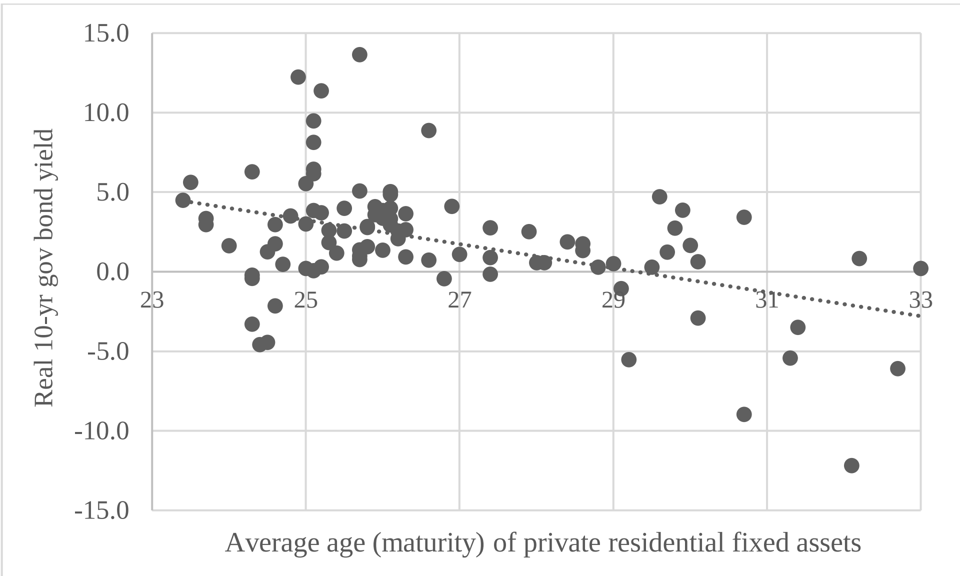


Figure 15: The average age (maturity) of private residential fixed assets, which basically reflects the residential housing market, plotted against the 10-year bond rate adjusted for inflation.

Another important argument against the generally used version of the Hayekian triangle, is the fact that many goods and services are used in different stages of production. The mere existence of goods and services that are used in *any* of the stages, would refute the Hayekian triangle and theory of capital in a heartbeat. Take accounting services, for instance; any business needs an accountant or accountant software, whether their activities would classify them closer or further removed from final consumption. Computers are also a capital good in, frankly, any stage. Both in early and late stages, roughly the same computers are purchased, even for entertainment purposes. Even our oil example shows that oil is an input to *any* stage, turning the Hayekian triangle into an arbitrary pastime of grayish, quibbling economists.

A final objection to the Hayekian triangle might be summarized by the fact that the vertical "output" axis suggests some type of physical quantity of consumer goods. An increase in savings, leads to an increase of the capital stock and, if invested correctly, to an increase in *valuable* production (an increase in *economic value*!). However, this does not mean that, by definition, an increase in savings leads to an increase of *physical* output. Yet, more than a real objection to Hayek's theory of production (e.g., 1931), this is a critique mostly directed to using the triangle as a visual aid to communicate capital theory.

In sum, the Hayekian triangle as it is used should be retired. However, there are other interpretations of the Hayekian triangle. Cachanosky & Lewin (2014), for instance, show that the Hayekian triangle can be interpreted as a financial representation of a cash flow schedule of a single investment project. Any capital investment has a maturity (the length of the triangle), whereas the vertical height represents the present value of the net cash flows, which goes down over time due to the time value of money. However, in this case, the triangle would be limited to simple bond investments that yield a steady stream of payments over its lifetime and remains inapplicable to any other investment outside that narrow domain. Perhaps, outside the realm of interest rates, the Hayekian triangle could show changes in consumption and savings. The Hayekian triangle could illustrate marginal changes in consumption and savings and their effect on the *returns* on capital (and hence the aggregate market value of the firms along the triangle). That is, any reduction in consumption would lead, *ceteris paribus*, to lower profits of retailers and thus their respective returns on capital. Any increase in savings, *ceteris paribus*, would lead to higher profits and thus higher returns on capital in the sectors where returns are highest (and hence more economic or price maladjustments exist). Yet, this fails to produce consistent outcomes, too. Some middle-stages could have higher returns than late *and* early stages, putting a premature end to any pretension to visualize the capital structure in a triangular shape. There is no need to confuse readers and insist on using a questionable visual aid.

It is not the distance to final consumption that counts, as the Hayekian triangle suggests, but rather the *capital-intensiveness* and the *weighted-maturity* of the cash flows of any given investment. Higher interest rates do not necessarily shorten or increase the slope of the triangle, but rather disfavors new investment in more capital-intensive projects with longer maturities/durations and lower the market value of already existing projects, of which projects with longer durations are affected to a greater degree than projects with shorter durations.

3.4 Keynes versus Hayek: J.M. Keynes in his *General Theory* on Capital Theory

3.4.1 The Alleged Absence of Capital Theory in Keynes

It would not be a stretch to argue that Hayek's *The Pure Theory of Capital* was a reaction to the success of Keynes' *General Theory*. Indeed, Hayek was highly critical of Keynes' lack of attention to capital theory in his seminal work. Looking back on the past, Hayek (1983) concluded:

> "[I]n the Cambridge tradition that governed Keynes's brief study of economics, the Mill-Jevons theory of capital, later developed by Böhm-Bawerk and Wicksell was not seriously considered. By about 1930, these ideas had been largely forgotten in the English-speaking world." (p. 48)

As Huerta de Soto (2006) adds: "According to Hayek, Keynes's lack of knowledge in this area accounts for the fact that he overlooks the existence of different stages in the productive structure (as Clark had done and Knight later would) and that he ultimately fails to realize that the essential decision facing entrepreneurs is not whether to invest in consumer goods or in capital goods, but *whether to invest in production processes which will yield consumer goods in the near future or in those which will yield them in a more distant future*." (p. 561). Effectively, Keynes assumed an economy consisting of consumer and capital goods, disregarding the temporal aspect of production.

Moreover, Hayek began criticizing Keynes' capital theory even *before* the publication of the *General Theory*. Hayek (1931) noted Keynes' neglect of capital theory in his previous work, *A Treatise on Money*:

> *All this would do no harm if his analysis of this complicating moment were based on a clear and definite theory of capital and saving developed elsewhere, either by himself or by others*. But this is obviously not the case. Moreover, he makes a satisfactory analysis of the whole process of investment still more difficult for himself by another peculiarity of his analysis, namely by completely separating the process of the reproduction of the old capital from the addition of new capital, and treating the former simply as a part of current production of consumption goods, in defiance of the obvious fact that the production of the same goods, whether they are destined for the replacement of or as additions to the old stock of capital, must be determined by the same set of conditions. New savings and new investment are treated as if they were something entirely different from the reinvestment of (…) amortization of old capital, and as if it were not the same market where the prices of capital goods needed for the current production of consumption goods and of additional capital goods are determined." (p. 278) [emphasis mine]

However, Keynes did imply a, although often contradictory, capital theory. The aim of this section is therefore to briefly summarize and, wherever necessary, criticize Keynes' contribution to capital theory. Moreover, we should point out his omissions. Up till this point, we have seen different perspectives on the nature of capital. On the one hand, we have seen physical concepts of capital (e..g, Böhm-Bawerk, 1888; Hayek, 1931), in which the physical input and output of material means of production are emphasized, whereas on the other hand, we have seen value or financial concepts of capital, in which physical output is de-emphasized. Keynes (1936) favored the financial approach to capital and even called the physical or material concept of capital "insoluble and unnecessary":

> "There is, to begin with, the ambiguity whether we are concerned with the increment of physical product per unit of time due to the employment of one more physical unit of capital, or with the increment of value due to the employment of one more value unit of capital. The former involves **difficulties as to definition of the physical unit of capital, which I believe to be both insoluble and unnecessary**." (p. 138)

Keynes, however, fails to mention or cite any other authors who share his point of view.

Another interesting theory, which gained academic traction, is Keynes' concept of the marginal efficiency of capital (MEC). MEC is very much akin to our modern-day internal rate of return (IRR), that is, MEC equals the rate of discount that makes the present value of a capital good equal to zero. Or in Keynes (1936) own words: "[T]he rate of discount which would make the present value of the series of annuities given by the returns expected from the capital asset during its life just equal its supply price." (p. 135). If this rate is higher than the rate of interest (or, more specifically, the cost of capital), then marginal investments will be undertaken. Whenever MEC is lower than the rate of interest, no investments will be made. Although investment decisions based on the MEC (or IRR) are the theoretical equivalent of decisions based on the net present value (NPV) of an investment, which might be called the "Austrian" approach to capital budgeting (e.g., Menger, 1888). In either approach, however, a decline in interest rates leads to marginal investments becoming profitable that at a higher rate were still unprofitable (and therefore unfeasible).

Less well known, is perhaps the fact that Keynes changed his opinion with regard to MEC, favoring a capitalization approach (Perelman, 1989), which is more in line with Menger (1888), Knight (1934) and Fisher (1930). Keynes (1937) wrote:

> "Capital assets are capable, in general, of being newly produced. The scale on which they are produced depends, of course, on **the relation between their costs of production and the prices which they are expected to realize in the market**. Thus if the level of the rate of interest taken in conjunction about their prospective yield raises the prices of capital assets, the volume of current investment (meaning by this the value of the output of newly produced capital assets) will be increased; while if, on the other hand, these influences reduce the prices of capital assets, the volume of current investment will be diminished." (p. 117) [emphasis mine]

Yet Lachmann (1956) was critical of this view, writing: "Keynes, to be sure, did not neglect the effect of the prices of existing capital goods on new investment, but, treating in characteristic fashion all capital as homogeneous, only saw the possibility of substitution. So he held that prices of existing capital goods below reproduction cost would weaken the incentive to invest." (p. 50). Yet in reality "(…) capital is as a rule heterogeneous and complementary. Except in the case, which Keynes alone considered, **where existing and**

new capital goods happen to be substitutes, low prices of the former will have a favourable effect on the incentive to invest. **Neglect of the heterogeneity of capital thus vitiates the theory of investment**." (ibid.) [emphasis mine]. This is largely inaccurate, however. Keynes's error does not lie in the fact that he treats capital as "homogenous," but rather that Keynes does not distinguish between investment in the underlying productive assets and investment in the *financial claims* on such assets and the dynamic between the two: this is what Tobin (1976) and later, among others, Spitznagel (2011) and Wright (2004) supplemented. If the price of existing capital goods (or, better, "capital") falls below reproduction costs, investment would shift, on the margin, to buying and bidding up existing capital (and capital goods would be the reproduction cost of capital as a claim on the underlying assets)[63]. There was absolutely no need for a "heterogenous" capital theory to "fix" Keynes' above explanation of capital assets. What Lachmann (1956) moreover misses, is *scale*. Criticizing others for their view on capital as a homogenous "fund" ignores scale, since such fund could be on asset level (the monetary value of a single asset), on a firm level, industry level, et cetera. The idea of "substitutes," mentioned above by Lachmann (1956), does not refer to material substitutability, but to *value* substitutability with regard to different *scales* or units of capital.

Moreover, Keynes (1937), unlike many other authors, did recognize the effect of lower interest rates on *financial asset* prices, which are more pronounced in more capital-intensive industries. The stock and bond prices in these industries increase relatively more than in other industries. More marginal investment, then, flows into these industries due to the higher returns, if investors *expect* these returns to persist (Perelman, 1989). What Keynes misses, however, is the fact that such increases or unable to expand "aggregate economic activity." This becomes apparent in Keynes' assumption that open market operations are "successful" in lowering interest rates. However, open market operations involve the exchange of one credit for another (a central bank might, for instance, purchase short-term government debt in exchange for central bank deposit money). Hence, central banks can only lower *some* interest rate at the expense of another. If and when a banking system engages in widespread *maturity mismatching*[64], a lowering of the long-term rate comes at the expense of an increase (in real

[63] This is another example of what we termed "Fisher's pendulum of returns." See p. 295.

[64] We will show that the banking system cannot *create* credit, merely transform credit/savings. See p. 271; For a more extensive treatment of maturity mismatching, see p. 288.

terms) of the short-term rate. Hence, what is the gain of (investment in) some industries, always comes at the expense of (investment in) other industries.

Keynes' work can largely be summarized by pointing out the great lack of an explicit treatment of capital theory; the less than satisfactory Hayekian capital theory, fraught with classical and physical fallacies, was however not the appropriate answer.

3.4.2 The Sraffa-Hayek Debate

In sum, Keynes had little appetite for capital theory. When Hayek began criticizing him for a lack of capital theory, one of Keynes' disciples, Piero Sraffa, would come to the rescue of Keynes. The real debate on capital theory between Keynes and Hayek did not involve Keynes directly, but rather his Italian pupil[65]. Since Sraffa was already aware of and familiar with the weaknesses of Böhm-Bawerk's theory of capital (Kurz, 2015), he seemed up to the task. Keynes, on the contrary, could sparsely understand the German language. Consequently, Sraffa (1932) wrote a rejoinder to Hayek's *Prices and Production* (1931), titled *Dr. Hayek on Money and Capital* (1932).

Sraffa (1932) criticizes Hayek for assuming "neutral money": "The money which he contemplates is in effect used purely and simply as a medium of exchange. There are no debts, no money-contracts, no wage-agreements, no sticky prices in his suppositions. Thus he is able to neglect altogether the most obvious effects of a general fall, or rise, of prices." (p. 44). Not only does Hayek rejects the notion of a (change in the) general price level, he abstracts away from *any* notion of the value of money.

Sraffa's critique here is completely justified. Hayek (1931) tries to treat capital theory completely separate from monetary theory, which, as we have seen earlier, is a contradiction and a mistake. If money equals, among other things, bank (demand) deposits, then any act of saving would paradoxically *add* to the quantity of money. Money, in this sense, is endogenous. In other words, Hayek (1931) contradicts himself in assuming neutral money, since an increase in savings (which forms the basis of his business cycle theory, that is, investment not backed by *real* savings leads to unsustainable investment that inevitably leads

[65] This turned out to be not Sraffa's last appearance in the history of controversies on capital theory. Sraffa would be on the side of Cambridge, U.K. in later the famous Cambridge controversies. For more, see p. 223.

to a generalized crisis) already implies non-neutrality. He fails, however, to recognize the intimate relationship between savings, investment and money.

Let us for now zero in on this link between capital and money. Hayek (1931) distinguished two different forms of capital accumulation in a monetary economy:

(1) Capital accumulation through voluntary savings;
(2) Capital accumulation through *forced* savings.

Under voluntary savings, "consumers place certain sums of money in the hands of the entrepreneurs who use them for lengthening the process of production, and thus capital accumulates." (Sraffa, 1932, p. 46). The result is investment "(…) identical with the effect which would have been produced if the savings were made in kind instead of in money" (Hayek, 1931, p. 49). There is a simple shift *away* from demand to consumer goods *toward* producer goods.

However, in this case, Hayek's (1931) treatment of money is inherently contradictory, as Sraffa (1932) correctly points out and as we have briefly mentioned above. On one side, Hayek assumes for analytical purposes to hold the quantity of money equal (which, for Hayek, equals the quantity of money under a fixed demand for money, that is, a fixed velocity), while on the other side, he assumes savings in money, which would in fact alter Hayek's velocity and thus quantity of money, which he before assumed to be equal. Hayek (1931), indeed, wrote: "(…) suppose that consumers save and invest an amount of money equivalent to one fourth of their income of one period" (p. 50). As Sraffa (1932) writes: "[Hayek] is driven to assume that the money saved is hoarded for a time, thus directly contradicting his postulate that the quantity of money multiplied by its velocity is constant." (p. 46). In other words, savings usually take the form of either (a) an increase in idle cash balances (demand or callable savings deposits) at banks or (b) an increase in nonbank investment, either direct or indirect (e.g., through an investment fund). In the former the increase in savings implies an increase in the "quantity of money"; in the latter case, there is more to Hayek's notion of "savings being made in kind." However, Hayek (1931) explicitly assumes the former case. More importantly, the forced separation of money from capital leads Hayek down the wrong path and to ignore the financial system (the complex of financial intermediaries) as "transmitters" of savings, of which commercial banks form part.

Under “forced savings,” banks expand the amount of money in circulation by granting credit to producers (Hayek, 1931). The initial consequences will be equal to the case of voluntary savings: the production structure is lengthened and the demand for producer goods increases relative to the demand for consumer goods. This time, however, this change in proportions is a result of the increase in the money supply. Therefore, “this sacrifice is not voluntary, and is not made by those who will reap the benefit from the new investments. It is made by consumers in general who, because of the increased competition from the entrepreneurs who have received the additional money, are forced to forego part of what they used to consume. **It comes about not because they want to consume less, but because they get less goods for their money income**.” (Hayek, 1931, p. 57) [emphasis mine].

However, this additional supply of money first leads to higher producer prices, before trickling down to increase wages and consumer prices (i.e., inflation), as producers got their hands on the new money first and used it to bid up the prices of factors of production. Then, as consumers try to restore their previous rate of consumption preference, there is a shift back toward producer goods, or as Hayek calls it, at least partly reestablish their initial consumption preferences (that is, the proportion between consumption and savings).

However, Hayek (1931) commits another error here, since he contradicts himself by first stating that producers use the new money to bid up the prices of producer goods, but at the same time equating this to an increase in the general price level, which would increase consumer prices and therefore reduce the real “money income” of consumers. Then, as wages readjust to the new supply of money (wages, in this case, adjust with a lag after the “inflation” is observed), consumers reestablish their previous consumption-and-savings pattern, which amounts to “a transition to less capitalistic methods of production [which] necessarily takes the form of an economic crisis.” (Hayek, 1931, p. 58).

Nevertheless, the great feat of Hayek (1931) was, according to Sraffa (1932), to show that money influences *relative* prices, rather than the general price level.

Sraffa (1932) then criticizes Hayek (1931) for arguing that there exists “one single equilibrium rate of interest.” Moreover, Sraffa (1932) began his critique by paraphrasing Hayek as if he said that the only reason the market rate of interest deviates from the “natural” equilibrium rate of interest, is money. In the absence of money, then, no such problem would arise. Sraffa (1932) continued by assuming a no-money economy, where every loan is made in terms of the same commodity (for instance, wheat is lent in wheat with interest paid in wheat). In this case, every commodity would have its own "rate of interest," and since loans in "wheat" would affect the production of wheat, no possible convergence of "interest rates" between different commodities could occur. Moreover, says Sraffa (1932) that if more is invested (lent) than saved, the income from the funds that are invested can be saved after all. Hence, such a second-order effect could potentially "fix" the whole issue, as long as entrepreneurs *save* whatever income stream they generate.

3.4.3 Keynes versus Hayek: Is Hoarding Different from Saving?

What later became the "paradox of thrift" was essentially Keynes' theory on hoarding. The "paradox of thrift" refers to the idea that savings are not transmitted (by the financial system) into investment, and thus "leak," metaphorically, from the economic system (or from Keynes's circular flow). As White (2012) explains Keynes's argument: "The channel through which savings go to fund investment spending had completely disappeared, so there was no possibility of the interest rate equilibrating saving with investment, as it did in the interest theory of Böhm-Bawerk, Wicksell, Fisher, or Hayek." (p. 135).

One the more important, and often underappreciated, errors of Keynes, is the fact that his positive 'investment multiplier' out of sheer logic assumes a negative 'investment multiplier' in a closed-accounting system, except when during a crisis financial intermediaries use proceeds from new deposits to simple cover the losses on their previous investments. Nevertheless, the problem is then the loss of value (loss in capital) of the intermediary's assets, rather than some paradox of thrift. Hence, Keynes (1937) should be focused on explaining the original source of his 'paradox of thrift', which would naturally lead him to the conclusion that an entire part of his theory is left unexplained. In this sense, Keynes (1937) is a monetary crank, since he seems to assume that capital losses can be compensated for by government spending, which quite obviously it cannot. Moreover, "income" does not equal "profit" and "income" does not equal "consumption," if the receiver of the income is willing to abstain from consumption (in this case, the receiver simply adds the income to his or her portfolio, adding to the amount of financial assets). However, Keynes (1937) seemed to think that *net* investment could generate his "investment multiplier", since the new investment is not offset by any disinvestment. Nevertheless, as Keynes (1937) basically lacks a sound theory of entrepreneurship, a dynamic model of price formation and a solidly grounded (intertemporal) capital theory, Keynes is unable to recognize the true source of profits. As

Huerta de Soto (2006) notes: “Keynes’s lack of an adequate theory of capital also explains his development of a mechanistic conception of the investment multiplier, which he defines as the reciprocal of one minus the marginal propensity to consume. (…) However, the investment multiplier hinges on a purely mathematical argument which contradicts the most basic economic logic of capital theory. (...) [According] to Keynesian logic, the less people save, the more real income will grow.” (p. 559) The paradox of thrift is, hence, merely a fallacy of composition.

Keynes’s theory of hoarding and the paradox of thrift thus results from viewing “hoarding” as a “leakage” from the system, whereas this would only occur when financial intermediaries *are already impaired* to expand their lending or investments.

3.5 Machlup’s Contribution on Human Capital

Machlup (1982) summarizes the debates on capital in, perhaps, the most appropriate words so far. He argues that all economists have considered tangible assets yielding tangible services, mostly limited to land, mineral deposits, et cetera. Machlup (1982) indirectly criticizes Böhm-Bawerk for his intellectual legacy, arguing against the narrow notion of capital as “produced producers’ goods” (p. 11). The concept, therefore, excludes both intangible assets and consumer durables and is limited to mere physical goods that were not provided by nature.

Machlup (1982) then argues that, at a later point, tangible assets yielding intangible services were also beginning to be considered as part of capital. He provides the examples of automobiles, residential housing, but would most likely consider computers if he still were to live as another illustration. Lastly, he arrives at the idea of “human resources.” Machlup (1982) argues: “[W]e find that the capital concept is no longer restricted to *tangible* assets yielding tangible or intangible services, but extended to *intangible* assets yielding tangible or intangible services.” (p. 12). He then goes on to show how education and training are methods of *capital formation*, but that the expected future benefits (that is, the post-training wage increase) are not accounted for as capital income in the statistics of national product. Machlup (1982) also discusses the treatment of research and development (R&D) investments in capital accounting, which leads him to distinguish four categories of capital:

1. Investment in *tangible non-human capital*: construction, machinery and inventory;
2. Investment in *tangible human capital*: rearing children to working age;

3. Investment in *intangible non-human capital*: research and development;
4. Investment in *intangible human capital*: education, training, health and mobility of workers.

Machlup (1982) continues his analysis by discussing the importance of the depreciation of human capital (in modern-day accounting terms, we would rather use the term amortization to refer to the fact that we are depreciating an intangible asset). He writes that human capital should be depreciated, since depreciation, albeit at different rates, is due: "(…) to obsolescence of acquired skills, to the loss of the workers' physical and mental strength, to changing rules of their retirement from active service and so forth." (p. 17).

Here we find, even though Machlup (1982) makes no such reference, a disagreement with Frank Knight (1935). Of course, Knight would be in complete agreement with Machlup on the nature of labor, as affirmed by Stigler (1985): "(…) the traditional distinctions between capital and labor are vigorously – and properly – criticized" (p. 5). Yet Knight (1935) argued that human capital was the factor that could overcome the tendency of diminishing marginal returns on capital, effectively distinguishing it from nonhuman capital.

The argument of Kaldor (1938) against including labor in capital, boils down to the fact that the tangible asset yielding the intangible services (in this case, a human being providing labor services) cannot be exchanged. Indeed, before 1937, Knight (1935) did make a slight distinction between labor and other capital goods on the basis of his argument that slavery no longer exists, and therefore no "market value" for the tangible asset would emerge. However, this is bullocks. Labor services do command market prices and labor could be capitalized by the owner, that is, the worker. As with any other capital[66], such labor services have a given lifespan over which they will be able to yield income, which then can be discounted to their present values. The fact that most human beings fail to do so in a formal and articulate way according to modern-day accounting principles, does not mean that labor services are not capitalized. If only workers would recognize the economic value of their labor, they would be able to make more informed decisions on education and training (estimating how much an investment in education would raise future income and thereby adding to their "human" capital). Licenses, permits and natural resource extraction rights are more often than not

[66] Please refer to Chapter 7 for our proposed definition of capital, p. 321.

nontransferable to third parties; they cannot change ownership. But they can still be valued as capital, by discounting their future cash flows to the present.

Knight (1938) apparently agrees with this notion (Northrup Buechner, 1976), since even though the principles of capital accounting are not applied to investment in labor, this “does not mean that rational management is not applied to an important extent, still less that results are vitally different from what they would be if it were much more generally and deliberately applied” (Knight, 1938, p. 76). This case is perhaps similar to the case of farmers failing to account for the value of their land that Ludwig von Mises (1949) mentions: “Even today in the most advanced countries only a part of the farmers are familiar with the practice of sound accountancy. Many farmers acquiesce in a system of bookkeeping that neglects to pay heed to the land and its contribution to production. Their book entries do not include the money equivalent of the land and are consequently indifferent to changes in this equivalent” (p. 262). The argument is, essentially the same. Land, in this case, is similar to labor, since there are many instances of non-transferable land that does hold economic value and should be considered part of the capital stock[67].

Thus far, we have analyzed the *source* of labor. The knowledge and skills involved in a certain amount of labor should be capitalized as human capital by the owner of the labor (that is, the individual worker), even though this is certainly not a modern every-day practice. However, we should also look at the other side of the coin, that is, the *product* of labor. How is the product of labor accounted for? Knight (1934) convincingly shows that any labor product can only be imputed to a “bearer” that is separate from the workers: that is, a tangible or intangible asset. In Knight’s (1934) own words (p. 263): “[L]et us assume that a capital instrument were to be constructed by the expenditure of a thousand dollars' worth of “labour” distributed uniformly through one year and that the interest rate is 5 per cent. per annum. The

---

[67] One of these examples can be found in *Our National Forests*, by Richard H. Douai Boerker: “The Secretary of Agriculture has the authority to permit, regulate, or prohibit grazing on the National Forests. Under his direction the Forest Service allows the use of the forage crop as fully as the proper care and protection of the National Forests and the water supply permit. The grazing use of the National Forest lands is therefore only a (…) *non-transferable privilege*. This privilege is a temporary one, allowable under the law only when it does not interfere with the purposes for which the National Forests were created. It is *non-transferable* because it is based upon the possession of certain qualifications peculiar to the permittee.”

value of the instrument when finished at the end of the year will be $ I,025. This $25 of "surplus value" over and above the direct outlay cost can only be imputed to the instrument itself, i.e., to the "capital" invested in it as this capital increases cumulatively. In the kind of world in which we live and think, there must be some such "bearer" (tangible or intangible) of the accumulating investment. This bearer at any stage of construction is a productive instrument, a capital good as well as a quantity of capital, and correct accounting must impute to it its rigorously definite share in the final result."

This point, regarding the *product* of labor, is especially important given the labor theory of value and, among others, Marx' work on capital, which traces the value of all capital back to the labor invested in it (and that, as a consequence, labor produces all wealth). However, rather than labor being a source and the defining factor of capital, labor is capital. Human capital is simply the present value of future wages. Hence, if we would compare in numbers the sum of nonhuman capital with the sum of human capital (by discounting future wages to the present), our inevitable conclusion would be that human capital is by far the most important component of capital.

3.6 Hicks' Early Explorations of the Aggregate Production Function and the Birth of the Cobb-Douglas Production Function

Charles W. Cobb and Paul H. Douglas published their well-known work *A Theory of Production* in 1928. Cobb, a mathematician, barely played a role in economics afterwards. Douglas dedicated himself fully to politics, abandoning his economic studies. Their paper is completely empirical; no capital theory is discussed, although it is clear that they put into practice the theory of capital by J.B. Clark[68]. After deriving a mathematical function from their empirical work (the Cobb-Douglas production function), they compare the results of their model with the actual data, and argue that the correlation between both equals .97.

Capital and production are assumed to be something physical, supplemented by labor, to yield a certain physical output. Cobb and Douglas (1928) exclude working capital considering that "for this is the result and not a cause of the process of manufacture" (p. 140). Yet they recognize that capital "produces value," but they are "(…) **not concerned with**

[68] For an analysis of J.B. Clark's capital theory and his exchange with Eugen von Böhm-Bawerk, go to p. 71.

value but with physical production." (p. 140) [emphasis mine]. Moreover, land is excluded as well, since they "are largely composed of the unearned increment" (ibid). Hence, capital is literally defined as the physical quantity of factory buildings, machinery, tools and equipment. As such, the Cobb and Douglas (1928) production functions equals:

$$P = bL^{\alpha}C^{\beta}$$

Where[69]:

P : Physical volume of production (output)
b : Residual, independent from L and C
L : Labor input
α : The coefficient (marginal productivity) of L, which equals 3/4
C : Capital goods input
β : The coefficient (marginal productivity) of C, which equals 1/4

In their estimation, however, Cobb and Douglas (1928) defer to a "value" measure of (narrowly defined) "capital," even though they explicitly stated that they were not interested in value but physical production. The value, however, is not the replacement cost of the mentioned productive assets, but rather the historical book value. Cobb and Douglas (1928) see no harm in this fact, since by simply correcting the historical purchase prices for inflation they would eliminate any misunderstandings due to changes in the general price level. (Curiously, this approach *still* closely mimics the modern-day approach to measuring capital in adaptations to the Cobb-Douglas production function.) Next, Cobb and Douglas (1928) estimated the average number of employed laborers (in manufacturing), regretting that the data at the time did not allow to calculate the exact "labor hours," since average length of a working week had declined and overtime was difficult to account for. They supplemented the capital and labor inputs with physical production (manufacturing) output.

First, they calculated a labor-to-capital ratio (equal to the modern-day capital-labor ratio K/L), with which they found that the ratio declined over time, meaning that the amount of "capital" grew relatively to the amount of labor over time (i.e., more capital goods per worker). Second, Cobb and Douglas (1928) compare the contributions of each factor to output and the distribution of gains, concluding that there is apparently "(…) a decided tendency for

[69] In modern-day applications of the production function *C* is replaced by *K* to represent capital.

distribution to follow the laws of imputed productivity" (p. 163). However, far from having established some causal truth, what we have established is two magnitudes that sometimes move together up, move together down or of which one moves up while the other moves down and vice versa. Unfortunately, the idea that this ratio *means* something has haunted economics and top-down government for so many years that it will be difficult to get rid of. Imagine, for instance, a top-down educated workforce with the necessary skills to work in the oil and gas industry. Nevertheless, on the other side of the world, an enormous, easy and cheap to extract oil resource is found, or a new technology arises that completely eliminates the demand for oil. Now, if we follow the *exact* physical logical of the Cobb-Douglas production function, output will decline, yet labor supply and the 'capital stock' (in material terms) remain equal. The residual (*b* in Cobb-Douglas) would thus be negative and have to explain the drop in output. Yet the fundamental economic facts are obscured by using the production function: what has happened is actually a malinvestment on a massive scale and an astonishing destruction of *human capital* and nonhuman capital, not captured by the model. The losses arising out of malinvestment will be largely ignored by the production function, since the 'capital stock' remains an input *as if* it is still economically equally valuable (given its physical presence) and 'labor' remains an input *as if* the present value of the training of the labor force is irrelevant and what matters are homogenous labor-hours.

Put differently, the higher the degree of complexity of a given economic system, the worse the production function will fit to the data and the worse it will be able to forecast long-term growth. The Cobb-Douglas function has other important limitations, even if we would assume for the sake of argument that its underlying assumptions are theoretically coherent. As (Folsom & Gonzalez, 2005) explain: "[T]he Cobb-Douglas function is homogeneous, it cannot have inflection points at which the relationship between output quantity and two or more input quantities can switch among convex, linear, or concave (…)" (p. 60). Indeed, the Cobb-Douglas production function *per definition* assumes away the crude nonlinearity of the real world. Increasing some variable input (say labor hours) will always increase output quantity, either at a constant, decreasing or increasing rate. However, in the real world, many phenomena are dose-related. For instance, the input of labor given a certain amount of capital might follow an *S*-curve:

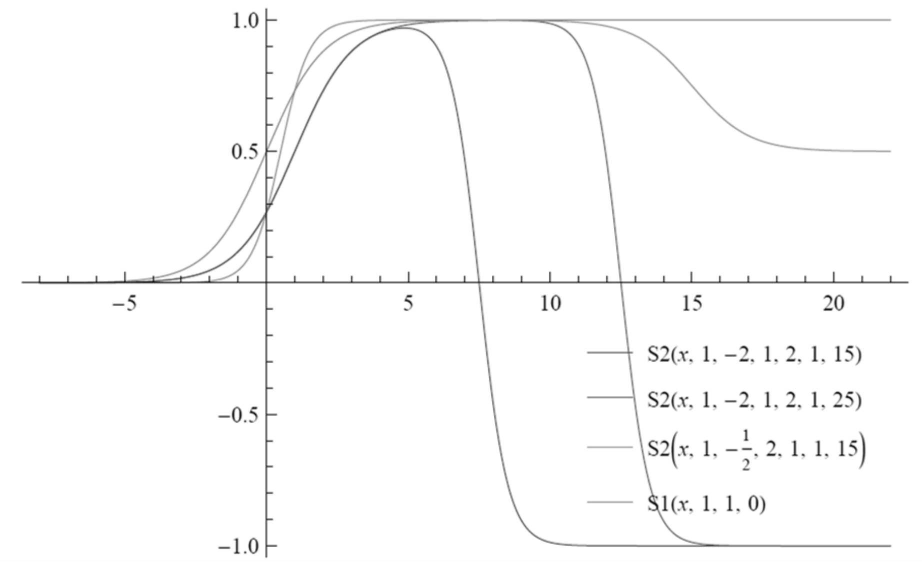


Figure 16: Four different S (dose-response) curves (Taleb, 2012).

As an example, an increase in labor hours, given a certain supply of hammers to build a cabin, will at first give convex returns (increasing rather than diminishing returns). At a certain point, the amount of labor hours (active laborers) diminishes as more and more laborers are concentrated on a few square meters trying to hit nails. We went from increasing to diminishing returns. Now we add even more manpower and increase the labor input. At this point, there are no spare hammers left, and adding labor input will actually have a net negative effect, decreasing total output (assume that workers begin to fight and use the hammers on each other or distract each other with bad jokes about econometricians) at an increasing rate for every laborer added (negative convexity). At a certain point, when all workers have clawed each other's brain out, it cannot get possible worse, so adding more labor will make us increasingly less worse off. We went from increasing returns, to diminishing returns, to increasing negative returns, to decreasing negative returns. It is this type of dynamic that a production function, including the Cobb and Douglas (1928) production function, is unable to handle, since it focuses on *physical* rather than *value* production[70].

[70] As often is the case, the disciples are worse than their masters: Cobb and Douglas (1928) recognized that they were focused on physical production, while others have taken the Cobb-Douglas production function to represent the complex dynamics of a whole economy, including its financial sector.

Moreover, Cobb-Douglas took the production function to back-test or fit past data to their simple input/output model. Yet the reason why is beyond me. According to Cobb and Douglas (1928), "(…) we shall have most interesting material on the slope of the curves of imputed productivity for a wide variety of industries and may be able to frame combined curves for a country as a whole and from this frame interesting international comparisons." (p. 165). The big question is: to what end?

Nevertheless, the Cobb-Douglas production function would later gain a prominent place in the third round of capital controversies, the famous Cambridge controversies. Before it reached its zenith, Hicks (1965) gave a first spin to the famous Cobb-Douglas paper of 1928. (Sir) John Hicks would later *Capital and Growth* (1965), which was based on Cobb-Douglas's initial work.

Before that publication, however, Hicks (1939) described a theory of capital with many 'Austrian' tenets[71], such that Hicks even considered himself a "neo-Austrian." Indeed, many of Hicks's ideas sound fundamentally 'Austrian'. Hicks (1939) would argue that capital is, in fact, a "fund." As such, he considers himself a "fundist," in contrast to the "materialists" (Kirzner, 1996). Nevertheless, he obfuscates what he means by "fundist," since he associates the "capital as a fund"-view with Böhm-Bawerk's view of capital as a subsistence fund. At any rate, Hicks ends up defending a view in which capital represents a subsistence fund spread out over time, in the present-shape of income yielding, heterogenous capital goods, apart from the production factor 'labor' (Hicks, 1973).

Additionally, on the notion of "roundaboutness" (from a similar financial point of view as us), Hicks (1973) wrote: "It is always true that a fall in the rate of interest will raise the capital value curve of any process will – will raise it throughout – while a rise in the rate of interest will lower it." (p. 19). He would call this a "Fundamental Theorem." Hicks (1973) also discussed the phenomenon of *reswitching*, which was a focal point in the later Cambridge controversies on capital (Hicks published his book *Capital and Time* after the Cambridge controversies). Reswitching would be both a problem for the 'Austrian' theory of "longer periods of production" at lower rates of interest, as well as a problem for the

[71] Hicks (1939) wrote that the "(…) core of truth of the Austrian theory needs to be discovered before we can really claim to have a satisfactory theory of capital." (p. 193). This is exactly what this treatise attempts to do: put the 'Austrian' theory of capital on solid footing.

neoclassical theory of capital-labor substitution (Lachmann, 1977). As we have mentioned before with respect to Lachmann (1956), timinig is everything. Hicks (1973) had little impact on the course of economic debate. After the Cambridge controversies, he published *Capital and Time*, which was sparsely cited and read (Burmeister, 2002). His 1939 treatise *Value and Capital* was cited more often.

Returning to Hicks (1965) initial explorations of the (Cobb-Douglas) production function, rather than applauding him for his advances in capital theory, we should note that his use of the production function is driven by an eagerness to apply theory. Even as Hicks (1973) considered himself a "neo-Austrian," Lachmann (1977) remarked that Hicks is "(…) so eager (…) to "get result," to show that feasible forms of the Traverse are at least possible (since otherwise the "steady state" remains a mere figment of imagination) that he seems ready to make any assumptions sufficiently restrictive to ensure them. We all understand that the present weakness of the neoclassical position may call for desperate measures. It is hard to see what is "Austrian" about them." (p. 262). Taleb (2012) would later call this the "best map fallacy": the idea that having the wrong map beats having no map. As a consequence, Hicks (1965) completely abandons the 'Mengerian insight' of consumer sovereignty, where capital (and the value of capital) depends on consumer preferences and decisions. Indeed, as Burmeister (2002) notes: "[Hicks] departed significantly from his earlier work by assuming that the technology of an economy consisted of a set of neo-Austrian production processes in which a time sequence of inputs $\{a_t\}$ produces a time sequence of outputs $\{b_t\}$." (p. 1).

Solow (1956) and Swan (1956) would later do what was inevitable and turn the Cobb-Douglas production function into its only possible use case: forecasting. Yet both as a forecasting tool and a policy tool it is completely unfit to its task. The on the Cobb-Douglas production function based Solow-Swan model and similar models to forecast (and pinpoint) economic growth leads to *overoptimization*. Whereas public finance in the past was a function of managing and mitigating cyclicality in government cash flows. The Solow-Swan model thus should not be replaced with other prediction models, but with heuristics and ways to appreciate where you have negative convexity to (often unexpected, unpredictable and unforecastable) change. Yet by *overoptimizing* variables (such as government budgets, industrial policy planning, infrastructure investment, military interventions, et cetera) on the basis of precise point growth estimates from flawed forecasting models, you are bound to run into trouble (Taleb, 2012).

3.7 The Untimely and Mostly Forgotten Contribution of Ludwig Lachmann

Timing is everything, especially when it comes to intellectual debates. Lachmann (1956), in that sense, arrived too late to have any influence on the debate on capital at the time. After a series of articles in the late 1930s, among which *Uncertainty and Liquidity Preference* (1937), *Investment and Costs of Production* (1938) and *On the Measurement of Capital* (1941), Lachmann published his treatise on capital theory in 1956, titled *Capital and Its Structure*. Nevertheless, Lachmann's timing was unfortunate in the sense that when he finally finished his treatise, few academics had any interest in capital theory. Lachmann's work on capital theory made little impact. Economists such as Frank Knight and Irving Fisher, who were active participants in the second round of controversies on capital, were largely on the verge of their retirement or already kicked the bucket.

Lachmann (1956) largely defended the same position as Böhm-Bawerk (1888) and Hayek (1941). Capital is a complex of heterogenous production goods. Adding up physical quantities (of different heterogenous capital goods), then, is pointless (Lachmann, 1941). Moreover, aggregating capital goods in terms of money (money value) is also pointless, according to Lachmann (1941), since it ignores the many "subjective" valuations of the future income that these heterogenous capital goods are able to yield (Lewin, 1977). In a dynamic economy, the money value of capital will fluctuate and therefore it cannot be represented as a "stock" or "fund" (Lachmann, 1977). This position, however, makes no sense from the perspective of the "subjective paradigm," as Lachmann calls it himself. It is the entrepreneurs' own estimate and there is only one entrepreneur who calls the shots over any given productive asset or combination of assets. Diverging valuations may lead to an exchange of the underlying assets[72]; yet, in an *ex post* sense, given a rate of interest and given a certain pattern of consumer spending, the "value of capital" (and the value of individual productive assets, which are the heterogeneous capital goods that Lachmann emphasizes) is far from subjective, but an objective fact in the context of price formation. The fact that price formation, as such, is ultimately driven by subjective consumer valuations, is a different matter. Prices of higher-order goods are therefore simply the outcome of subjective consumer valuations, but that does not mean that (future) prices of higher-order goods are subjective. They are an objective *datum*, as Mises (1949) would call them, in the

[72] Perhaps best represented by the modern-day phenomenon of corporate takeovers.

sense that they either materialize or not. Which of the entrepreneurs earns a profit *ex post* and thus discovers that his expectations or past valuations were correct, is in this case of less relevance. In addition, there is nothing wrong with the "value of capital" fluctuating to reflect changes in either (a) time preferences or (b) consumption preferences. If the "value" of the capital of a tea producer decreases because consumers prefer coffee, then this signals an important fact of economic life. What matters, in the end, is economic value and not the physical order of a set of capital goods. Lachmann (1977) takes the term "capital stock" too literal, that is, in a material rather than a value sense. Lastly, all the critiques we have levered in our past discussions on the idea of capital as a set of material or "produced" means of production, equally apply to Lachmann's theory of capital.

In the light of Lachmann's Hayekian theory of capital, he focused on the role of complementarity and substitution of capital goods. Lachmann (1977) intended "(…) to show that it is in the theory of capital that the concept of complementarity proves a most powerful lamp to throw light into some notoriously dark corners." (p. 198). Lachmann (1977) begins by arguing that by considering capital as a "fund," *perfect substitutability* is assumed. He then criticizes this idea. Nevertheless, a dollar worth is a dollar worth. If I control a production plant, then its "worth," as it forms part of my total capital (*net worth*), is simply the present value of its future yields. I can sell it and monetize it and reinvest the proceeds into some other investment. In this sense, nothing happened to "capital," only ownership has changed. Yet, as Lachmann (1977) rejects the idea of capital as a value fund, he assumes *imperfect substitutability*, since he now refers to the underlying material assets. Of course, one means of production, say Lionel Messi, cannot be replaced by Cristiano Ronaldo: they are not perfect substitutable. However, the monetary value of Messi (the present value of his future economic contribution) *can* be substituted, by for instance investing the proceeds of a sales into various other players or nonhuman assets. Thus, Lachmann (1977) defines capital as a structure of material goods, which are sometimes substitutes and sometimes complementary goods. Lachmann's obsession with the physical nature of production comes from his obsession with "malinvestment." If capital is a mere fund of abstract monetary value, then "malinvestment" could be simply reasoned away as irrelevant, especially if you assume it is a "stock" immune to change and fluctuations. Yet if you have a physical factor of production, for instance a hotel boulevard paired with a complete absence of tourists, then the malinvestment becomes something tangible and concrete: you can point to it. One of the consequences of this view is that capital is just a material input next to labor. If we ought to study the technical complementarity and substitutability of production goods, then why not focus as well on the technical productivity of production goods and the optimal, technical ratio between labor and nonlabor input? Lachmann (1977), as such, pretends to further the cause of subjectivism, but actually does a disservice to subjectivism with his physical theory of capital. His reference to the "importance of complementarity" in the light of "technical rigidity" (p. 200) and "coefficients of production" (p. 202) should, for example, be interpreted in this light. Lachmann (1977) later emphasizes "the existence of unemployed labour and unutilised resources" since they "provide potential complements for (…) new productive combinations." (p. 206), which again demonstrates his fixation on material production. Lachmann (1956) even flat out admits that he is more concerned with a physical theory of capital, as he explicitly states: "The theory of capital is thus primarily **a theory of the material instruments of production**." (p. 54) [emphasis mine]. Moreover, capital goods ought, according to Lachmann (1977), be considered separately from labor and "permanent" resources, as they are "more sensitive to unforeseen change." (p. 203). If we take Lachmann's capital theory to its extreme, economics has nothing to say about the real world, apart from some general statements that entrepreneurs combine complementary capital goods and that imperfect substitutability can lead to losses and "malinvestment." Lachmann (1977), in effect, devised a new but equally fallacious justification of separating factors of production and to persist in the errors that characterized Böhm-Bawerk's and Hayek's material theories of capital, which was an unfortunate classical endowment to begin with. Lachmann (1977), if he would have used the imaginary construction of equilibrium, would have come to the conclusion that any maladjustments in the structure of capital arising out of imperfect combinations of productive assets would simply be one of the reasons why a potential arbitrage profit would exist for entrepreneurs to earn. Equilibrium would, then, be such that all prices of all productive assets reflect the fact the value of capital, such that no further "rearrangements" can be made to earn a profit. Yet, the focus on "complementarity" seems just one of the many *material* underlying motives of maladjustment and causes of profit opportunities. As an example, "geography" could equally be a concern. Yet economists would not pretend to have anything useful to say about the physical location of entrepreneurial ventures, other than urging to pick the most profitable location[73]. Yet there are no regularities, or economic laws, that can be distilled from the idea that some physical locations are more "optimal" than others, just as some combinations of complementary

[73] Profitable as in risk-adjusted profitability, that is.

capital goods are more “optimal” than others. Since economic science concerns economic *value*, it suffices to conclude that any disequilibrium as to the preferences of consumers (as evidenced by their spending) is embodied by a profit opportunity. The source of a profit opportunity is no object of study for economic science, since there are million possible sources of profit and a suboptimal degree of complementarity between a given arrangement of capital goods is one of such sources.

While we have been remorseless so far with regard to Lachmann’s *general* theory of capital, Lachmann’s (1956) work on the *capital structure* in a financial sense is without a doubt a most valuable and impressive contribution, emphasizing the role of the entrepreneur as a coordinative establisher of equilibrium. The difference between Lachmann’s theory capital compared to his theory on the capital structure is the fact that the former describes one possible of many states of disequilibrium, whereas the latter describes equilibrium and thus serves as a guide to entrepreneurs to unravel all possible states of disequilibrium in the real world.

Lachmann’s (1956) contribution lies in recognizing the intimate relationship between money and capital, which we have discussed on earlier occasions. His capital structure can be divided into three parts:

Plan Structure (left) & *Control Structure* (right)

Operating assets	**Securities**
First-line assets (fixed capital)	Debt
Second-line assets (working capital)	Equity
Reserve assets (excess cash)	

Portfolio Structure (left)

Portfolio	**-**
Assets / Securities	

Figure 17: Lachman’s (1956) capital structure captured in two simplified balance sheets. We have preserved the original terms used by Lachmann.

Lachmann (1956) himself summarizes his capital structure as follows:

> “[T]here are three kinds of structure: The *Plan Structure* based on technical complementarity, the *Control Structure* based on high or low gear of the company's capital, and the *Portfolio Structure* based on people's asset preference. These three

> structures are not independent of each other. Whether a given production plan with its accoutrement of plant, equipment, raw materials, etc., is at all feasible depends *inter alia* on whether people are willing to take up the securities necessary to finance it, and this in its turn will depend on whether debentures, preference shares, or common stock are offered to them, and in what proportions." (p. 91)

The reference to "people's asset preferences," according to Lachmann, should definitely not be compared to people's consumption preferences. When we are faced with a choice in a supermarket between, by means of illustration, two brands of chocolate, we do not have to think about, as Lachmann calls it, the "managerial efficiency" of each of the two producers. We simply pick the chocolate we like most (given their respective prices) without any further consideration. When we pick securities, however, we tend to take into account different variables, such as the security and profitability of the underlying "managers" (also notice here Lachmann's focus on the manager rather than the entrepreneur).

Lachmann (1956), for instance, assumes that profits are – for some reason – higher than expected. What happens with the surplus of profits? They are either (a) reinvested in the business, (b) used to pay off debt, or (c) paid out to shareholders. Each of these options has a different impact on the capital structure. Each of these options will change the composition of the plan, control and/or portfolio structure. Similarly, when losses are incurred, at some point they are accounted for in either one of the three structures. Paradoxically, as Lachmann goes on to show, both an "(…) expansion following on success as well as reconstruction following failure cause the 'demand for money' to increase." (p. 93). Of course, Lachmann here assumes that savings equals the demand for money. However, as will become apparent, it is actually the broader concept of the demand for financial assets that matters, of which the demand for money is a subcategory. Yet the underlying, often underappreciated, connection between money and capital is explicitly mentioned by Lachmann (1956). However, at the same time, it should become clear that Lachmann missed a layer, a 'structure', in his analysis. Financial intermediaries, specifically, are the missing link between his *Control Structure* and *Portfolio Structure*. Households save money, which adds to their asset portfolios, representing their accumulated portions of deferred consumption. That money is, generally, taken by financial intermediaries – *financial entrepreneurs* – who then invest, against some kind of fee, savings into a variety of different investment opportunities, presented to them by *nonfinancial (industrial) entrepreneurs*. Hence, in this dynamic process of capital allocation, there are three important layers: the *control structure* (financial intermediaries' assets), the *funding/finance structure* (financial intermediaries' liabilities) and the *portfolio structure* (savers/capitalists' assets). Such processes, as Lachmann (1956) himself explains, involve a "(…) transmission of knowledge [that] bring the various constituents of the asset structure into consistency with each other, modifying the control structure and the composition of portfolios." (p. 95). The three structures are part of the coordinative function of markets and entrepreneurs, throughout which consumer sovereignty reigns.

Curiously, Lachmann (1956) blames Keynes's theory of liquidity preference for "assuming equilibrium," when in fact Mises (1949) showed that in equilibrium (the "stationary" state) there would be no need for cash balances or money. In fact, Lachmann (1956) rightly emphasizes that we are principally interested in how change works, rather than equilibrium in itself: "A theory of capital relationships based on the assumption of invariable success of plans is apt to lead to wrong conclusions when applied to a world of unexpected change." (p. 91). The *capital structure*, summarized by three 'structures', is Lachmann's most important contribution (Lachmann, 1956).

Despite Lachmann's (1956) neat contribution, he fails to incorporate the overarching leitmotif that connects the three structures (*plan*, *control* and *portfolio* structure): *time*. Or more specifically, as we will see further below, *duration*. Nor did Lachmann (1956) extensively analysis the role of (commercial) banking and capital markets in his analysis, something which Moulton (1921) did attempt. In fact, as Lachmann (1956) distinguishes between fixed and working capital, as well as excess cash balances, he does not recognize that the financing behind these different types of capital is very different. For instance, working capital can be partly financed by factoring (selling) accounts receivable, inventory financing (with inventory as collateral) from specialized lenders, credit lines from banks, issuing short-term debt on the bond market or even by accepting pre-orders (prepaid orders[74]). Fixed capital, however, can be financed by issuing shares (stock market), issuing long-term bonds (bond market), private lending (for example, by family-offices), bank lending, issuing asset-backed securities, or nowadays even by issuing shares on crowdfunding platforms (in this case you have an equity stake held outright by a household: the role of intermediary or *financial entrepreneur* and *saver/capitalist* converge). Of each

[74] Tesla, for instance, financed part of its production by charging down payments on pre-orders (reservations) on their Model 3.

instrument the conditions and the costs tend to differ, which will ultimately effect both portfolio decisions of savers and investment decisions of businesses[75], yet what distinguishes and unites one financial asset from and with another is, as a common thread, its *maturity* or the element of *time*.

Let us attempt to show the importance of time (maturity) in Lachmann's analysis of the capital structure. We assume no financial intermediation exists. Society's stock of savings is only allowed to be invested in one way: loans with a five-year duration. There exists one single business. At $t = 0$, we end up with two simplified balance sheets:

Figure 18: Society's consolidated balance sheet at t *= 0 in absence of financial intermediaries*

Assets	**Liabilities + equity**
5-year loan to Business X	Equity

Figure 19: Consolidated balance sheet of Business X at t *= 0 in absence of financial intermediaries*

Assets	**Liabilities + equity**
Savings	5-year loan from society

In this hypothetical and simplified world, what would be the interest rate on the 5-year loan? Let us assume that Business X expects to make a return on capital of 10%. This would, theoretically, mean that at 9.9% they are willing to borrow and start the venture, while at 10% or higher they would not be willing to borrow and start the venture. This would be the "price ceiling" in this exchange. On the other hand, society is only willing to save and invest at five years, if they will get 5% in return. That is, they value saving for five years more than consuming now at a 5% rate of interest, but not at 4.9%, given the intertemporal marginal utility of society. This would be the "price bottom" in this exchange. Any other outcome, would result in no exchange.

[75] We have covered here, for simplicity sake, more or less the private realm without considering public financing needs.

Figure 20: An isolated exchange between society and Business X

Supply of savings (from society)	**Demand for savings (from Business X)**
> 5%	< 10%

Where would in this hypothetical case the interest rate be? Anywhere between 5% and 9.9%. Notice that there are two "if's" in this example: on the one hand, the demand for savings depends on the expected return on capital of businesses and, on the other hand, the supply of savings depends on the subjective (intertemporal) valuations of consumers.

Now it is time to insert one single bank. The balance sheets shift with the introduction of a financial intermediary:

Figure 21: Society's consolidated balance sheet at t=0 in presence *of a financial intermediary*

Assets	**Liabilities + equity**
5-year loan to Bank Y	Equity

Figure 22: The financial intermediary's (Bank Y's) consolidated balance sheet at t=0

Assets	**Liabilities + equity**
5-year loan to Business X	5-year loan from society

Figure 23: Consolidated balance sheet of Business X at t = 0 in presence *of a financial intermediary*

Assets	**Liabilities + equity**
Savings	5-year loan from Bank Y

What power does our financial intermediary, Bank Y, has in this case to set and fix interest rates? The answer is straightforward: no more power than in our bank-less example:

Figure 24: An isolated exchange between society and Bank Y

Supply of savings (from society)	**Demand for savings (from Bank Y)**
> 5%	**< 9.9%**

Figure 25: An isolated exchange between society and Bank Y

Supply of savings (from Bank Y)	**Demand for savings (from Business X)**
> 5.1%	< 10%

Bank Y, in our example, is effectively arbitraging the market for time preferences, but it is limited to the actual underlying preferences of the players involved – the suppliers of capital (savers and their *portfolio structure*) and the users of capital (entrepreneurs and *their plan structure*).[76]

In real circumstances, consumers have a wider range of choices and have more options to choose from. If a saver is not convinced by the prospect of lending at five years for a rate of interest of five percent, then the alternative is not merely present consumption or a liquidation of his savings. He can either remain "liquid", that is, stay invested in liquid short-term investments, so he can wait for a better offer (the saver is here simply exercising his entrepreneurial judgment) *or* consume in the present and liquidate his savings. If entrepreneurs believe that rates of interest are excessively high, only businesses with exceptional returns on capital will be established. Hence, there can be a greater abundance (scarcity) of arbitrage opportunities by entrepreneurs and/or there can be a greater preference (diminished preference) to postpone the use of funds (that is, ultimately consumption) into the future by consumer-savers.

[76] Of course, the whole issue of economic disequilibrium and business cycles, finds its origin in one of the few powers the financial sectors actually *does* possess: introducing asset-liability (balance sheet) mismatches, principally maturity and interest rate mismatches. Currency mismatches have great practical consequences as well.

Both will even each other out. If maladjustments become increasingly numerous, profits will rise and returns on capital will increase, increasing demand for savings on the intertemporal market. If maladjustments are increasingly scarce and hard to adjust, profits and returns on capital will be low, lowering demand for savings on the intertemporal market. Consequently, if many consumers decide to save, the supply of savings goes up. If few consumers decide to save, the supply of savings is curtailed. Hence, the pure time preference theory of Mises (1949) lacks one side of the equation (pure entrepreneurial profits), very much akin to Fisher's (1930) "investment opportunities," while opposing theories of interest lack one side of the equation and simultaneously misinterpret the other (that is, there is no role for time preference or abstinence, while there *is* a role for available returns on capital, yet the returns on capital are "automatic", "natural" or the metaphorical apples of an apple tree that reflect the inherent material productivity of capital goods).

Last but not least, Lachmann (1956) summarizes the Misesian (and Hayekian) theory of the business cycle, but without any reference to his previously elaborated *capital structure*. While briefly mentioning that price distortions lead to an inevitable cycle of boom and bust, with interest rates being one of such prices, he is unable to connect the dots and relate his capital structure, time and mismatching of time preferences to the Misesian theory of the business cycle.

3.8 Risk, Probability and the Value of Capital

One of the few Schumpeter (1954) praised Knight's work on the role of risk in capital theory: "Among the rest, many authors developed Mill's (or A. Smith's) element of risk. This was done most successfully by Hawley and especially by Professor Knight. To the latter we owe, in the first place, a very useful emphasis upon the distinction between insurable risks and non-insurable uncertainty; and, in the second place, a profit theory that linked this non-insurable uncertainty on the one hand to rapid economic change—which, barring extra-economic disturbances, is the main source of this uncertainty—and on the other to differences in business ability—which are much more obviously relevant to the explanation of profits and losses in conditions of rapid economic change than they would be otherwise. He thereby achieved a synthesis that is not open to the main objection against the ordinary type of risk theories" (p. 861). However, as Taleb (2015) notes[77], "[t]here is no such thing as "Knightian risk" in the real world, but [only] gradations of computable risk." (p. 60). Such gradations of risk also appear at the level of the probability distribution itself (a "meta" or higher order of uncertainty). [At] one end of the spectrum, "Knightian risk", is not available for us mortals in the real world." (p. xiv). As a result, we could typify the Knightian theory of risk (with on the one hand computable risk and on the other hand noncomputable uncertainty) as an artificial distinction of something that cannot be inherently distinguished.

In many cases, capital can only be valued by estimating some sort of probability accounting for the riskiness of future cash flows. Entrepreneurs (which includes financial entrepreneurs or, colloquially called, investors) try to anticipate the market and outcompete fellow entrepreneurs in a game of arbitrage. For instance, a bank's loan book is basically a function of its probability. While this probability is uncertain, it can be based on historical empirical data and/or other statistical instruments: no clear distinction between risk and uncertainty exists.

However, some economists have taken great strides to argue for a conceptual divide between, on the one hand, risk and, on the other hand, uncertainty. As Stigler (1985) summarizes Knight's stance on this dichotomy, "[t]rue uncertainty is to be "radically distinguished" from calculable risks."

Yet, Frank Knight's distinction between risk (which is probabilistic and insurable) and uncertainty (which remains unknown, is not probabilistic and therefore not insurable) has major issues (later economists would refer to these as "Knightian risk" and "Knightian uncertainty"). Knightian uncertainty implies that "no meaningful probability can be assigned to possible future results [and, w]orse, it means that in many cases no known lower (or upper) bound can even be assigned to the range of outcomes." (p. 6). As Taleb and Pilpel (2004) note, an entrepreneur acting in the real world does not know *ex ante* if he is faced with, what Knight calls, risk or uncertainty: "[T]he distinction is irrelevant, actually misleading, since, outside of laboratory experiments, the operator **does not know beforehand** if he is in a situation of 'Knightian risk'" (p. 4) [emphasis mine]. In fact, Keynes (1937) would later

[77] Taleb (2015) argues that a "good heuristic is to disqualify any adult who uses the idea of "Knightian risk" as incompetent." (p. 60)

criticize this idea with regard to economic events: "(...) there is no scientific basis on which to form any calculable probability whatever. We simply do not know!" (pp. 113-114).

Mises (1949) followed very much the line of thought of Knight. He distinguished between "case probability" and "class probability," which are the rough equivalents of "Knightian uncertainty" and "Knightian risk." Class probability is insurable and can be accounted for in market prices, whereas case probability is unique and uninsurable. However, we can lever the same critique on Mises as Knight; the distinction between such assumingly knowable and unknowable becomes superfluous when we take into account that *ex ante* all risks in economics are fundamentally unknowable, since agents are unable to know *ex ante* on the basis of past data what the corresponding probability distribution and parameters are.

This unwarranted distinction between risk and uncertainty leads us down a treacherous path. "[W]e cannot tell not only whether or not X will happen, but not even give any reliable estimate of what $p(X)$ is. The cardinal sin risk managers commit is to "force" the square peg of uncertainty into the round hole of risk, by becoming convinced without justification both of the generator type and of the generator parameters." (Taleb & Pilpel, 2004, p. 7). Since the risk that Knight (1921) discusses as "priced in" in asset and firm prices, whereas true uncertainty cannot be priced and thus is to be shouldered by "uncertainty-bearing" entrepreneurs, the idea that in practice entrepreneurs are unable to assign a specific probability to future outcomes prevents such risk from being priced in correctly or at all in asset prices.

In other words, take the example of a mortgage lender. A mortgage lender might possess data on past default rates. Therefore, according to Knight (1921), such default rates can be accounted for in its mortgage rates. The lender does not know which mortgage borrower will default, but he knows that, based on past data, the likelihood of default among similar mortgage borrowers is, say, three percent. However, the "generators" of probability distributions and their parameters (mean, standard deviation, etc.) in economics, which are the very subjects or human beings involved, prevent us from assigning any specific probability to future outcomes, especially when the probability space is unbounded (although past data can give us the impression that it is "bounded" by its past, most extreme deviations[78]). No amount of non-catastrophic data can provide us with information on the catastrophic data, until such catastrophe hits (Taleb & Pilpel, 2004). Hence, since this is true, Knight's (1921) risk, which is then priced in capital, borders on the brink of irrelevancy. Since all this risk involves some type of human action (even when the impact or degree of human intervention that contributes to an economic outcome is minimal, for instance, in the case of natural disaster insurance), probabilities cannot be correctly estimated *a priori* and therefore not be priced in *correctly*. Hence, Knight's distinction between risk that is priced in and uncertainty that is not priced in but "borne" but the entrepreneur boils down to almost all future events being "uncertain" (not subject to *a priori* knowable probabilities) and therefore entrepreneurs either pricing in uncertainty. But then the price of capital has no obvious fixed point to arbitrage; but are merely influenced by the amount of uncertainty an entrepreneur is willing to bear and *not* price in. Moreover, Knight (1921) thought that this was a problem, principally, of "not having enough data." According to Taylor (1980): "As Knight has said the problem [of uncertainty] stems from the inability to accumulate sufficient empirical data relating to particular classes of subject and events" (p. 76). Hayek (1978) seemed to agree: "In the social sciences we have to deal with (…) phenomena which are not made up of sufficiently large numbers of similar events to enable us to ascertain the probabilities for their occurrence" (p. 5). Yet the idea of "having not enough data" is precisely the idea Taleb and Pilpel (2004) criticize; no amount of normal data can inform us about non-normal or extreme events.

In other words, in a financial theory of capital we more or less assume that future cash flows are given, and in no way uncertain. Yet, future cash flows (or income) are inherently uncertain and fraught with implicit probability estimates[79]. In other words, the net present value (NPV) of a firm, say a chemical plant, given its future expected cash flows (both positive and negative) could range from anywhere in the billions (if the firm is highly

---

[78] Consider a stock market investor in 1986 that, based on historical returns, takes into account a maximum daily drawdown of 6.7% (highest daily loss up till that point, which occurred in May 1962). This stock market investor would be blown out of the water by the 20.5% drop in 1987, which is more than three times the largest daily loss on the S&P 500 before its occurrence. The largest daily loss (in percentage) on the stock market is just the largest daily loss until the next largest daily loss.

[79] No businessman pins himself down at an exact net present value estimate: a range is considered and sensibilities are tested before decisions are made.

effective and successful) to literally zero (if its plant explodes and causes a chemical disaster). Any return on capital (including a 100% negative return) is, therefore, "uncertain."

However, Knight's "gedankenexperiment" is interesting. With perfect foresight, all risks (uncertainty) can be priced in the current prices of capital and assets. Indeed, under the assumption of competition between entrepreneurs, any positive margin that is left in a price when accounted for all risks should be arbitraged away. This was, implicitly, the position of economists during the entire first and second round of debates on capital theory: risk would only at a later point become a focal point in the intellectual debates. For now, it suffices to fathom Knight's initial work as the genesis of all future work on risk in the context of capital theory.

A future problem would arise with the development of modern finance that many theoreticians assume, contrary to this work, the entire future to be "knowable" *ex ante*, including in cases of Knightian uncertainty or "case probability." We tend toward the other edge, where even Knightian risk and class probability are actually about uncertainty.

There exists, of course, a difference between *ex ante* estimated probability and *ex post* observed probability. Human beings are fallible and entrepreneurs are no exception. They can either "get it right" or "get it wrong."

Risk often has been confounded with probability[80], that is, the odds of something happening. However, while probability refers to the odds of something happening, risk consists of two

[80] Ludwig von Mises (1949), for instance, compared class and case probability, but never made any explicit reference to risk as such. He does refer to uncertainty: "Life itself is exposed to many risks. At any moment it is endangered by disastrous accidents which cannot be controlled, or at least not sufficiently. Every man banks on good luck. He counts upon not being struck by lightning and not being bitten by a viper. There is an element of gambling in human life. Man can remove some of the chrematistic consequences of such disasters and accidents by taking out insurance policies. In doing so he banks upon the opposite chances. On the part of the insured the insurance is gambling. His premiums were spent in vain if the disaster does not occur. With regard to noncontrollable (…) events man is always in the position of a gambler." (p. 112). Moreover, Mises (1949) clearly understands second-order effects (and *n*-order effects), which turns the business of forecasting into a fool's game: "Every new datum brings about a reshuffling of the whole price structure." (p. 118).

elements: (1) the odds of something happening and (2) the potential pay-off or loss when it happens or simply exposure to a certain outcome (Taleb, 2012).

However, that leaves us with the question of the entrepreneur being the true "uncertainty bearer." After all, the resources invested are not provided by the entrepreneur, but by resource owners (savers or "capitalists"). Attributing this uncertainty-bearing to the entrepreneurial role seems therefore to be inaccurate from a theoretical point of view. Yet, as a consequence, we add "another layer" of uncertainty. If the entrepreneur is not the uncertainty bearer but the executor, then how can we

Hence, the broader question that arises is: is this how equilibrium should be defined? With risk priced in as exact event probability estimates?

3.9 Mises as Menger's Savior on Capital Theory

Mises's greatest feat was his amplification of Menger's (1888) original contribution to capital theory, especially in *Human Action* (1949). As Gunning (2004) notes: "The only prominent Austrian to follow Menger on the capital concept is Mises. Like Menger, Mises rejected the materialist definition of capital." (p. 4).

Mises (1949) was very clear as it came to capital and its relation to the acting entrepreneurs and economic calculation (modern accounting and capital budgeting):

> "From the notion of capital goods **one must clearly distinguish the concept of capital**. **The concept of capital is the fundamental concept of economic calculation, the foremost mental tool of the conduct of affairs in the market economy**. Its correlative is the concept of income (…). The whole complex of goods destined for acquisition is evaluated in money terms, and this sum – the capital – is the starting point of economic calculation (…). That amount which can be consumed within a definite period without lowering the capital is called income. If consumption exceeds the income available, the difference is called capital consumption. If the income available is greater than the amount consumed, the difference is called saving. Among the main tasks of economic calculation are those of establishing the magnitudes of income, saving, and capital consumption. (pp. 260-261) [emphasis mine]

Economic calculation, according to Mises (1949), involves:

> "The task which acting man wants to achieve by economic calculation is to establish the outcome of acting by contrasting input and output. Economic calculation is either an estimate of the expected outcome of future action or the establishment of the outcome of past action. But the latter does not serve merely historical and didactic aims. Its practical meaning is to show how much one is free to consume without impairing the future capacity to produce. It is with regard to this problem that the fundamental notions of economic calculation--capital and income, profit and loss, spending and saving, cost and yield--are developed. The practical employment of these notions and of all notions derived from them is inseparably linked with the operation of a market in which goods and services of all orders are exchanged against a universally used medium of exchange, viz., money." (p. 211)

It is therefore more or less equal to modern-day accounting (to establish the "outcome of past action") and capital budgeting (to appraise "the expected outcome of a future action") practices. As such, capital and economic calculation are inseparable. One could not exist without the other.

Moreover, according to Gunning's (2008) interpretation of Mises (1949), Mises defended a theory of interest very similar to Fisher's (1930). Mises is mostly known for his theory of time preference as the source or origin of interest. Yet, according to Gunning (2008), Mises:

> "(…) leaves open the possibility that the *market rate of interest* might be "led," so to speak, by entrepreneurial estimates of the profitability of current investment in new methods of production and in research. Allowing for this possibility is an extension or example of consumer sovereignty which I have argued is the key to understanding Mises's economics. **It allows for a seamless integration of an entrepreneur-driven theory of market interest with other entrepreneur-driven economic phenomena**. Instead of the rate of interest in a market economy being a rate that adjusts to some imaginary "pure rate of interest," it is led by entrepreneurial anticipations of the consequences of investment. **Such an integrated theory of interest includes not only consumer ends but also entrepreneurial knowledge of the means**." (p. 4) [emphasis mine]

Indeed, one of Mises's (1949) main contributions to Menger's (1888) initial work, was to integrate his (arbitrage) theory of entrepreneurship with Menger's non-material theory of capital (Gunning, 2004). Menger (1888) viewed entrepreneurial activity as a subset of labor services. Such specific entrepreneurial labor services consisted of appraising economic circumstances and economic calculation. Yet, as Huerta de Soto (2008) showed, the entrepreneurial function is inherently *creative*. Moreover, as Gunning (2004) shows, if the entrepreneurial function consists of using knowledge to determine whether an object or thing is part of capital, it is some type of circular logic to call entrepreneurship, then, a part of that same capital. If entrepreneurship involves determines whether a good is capital, then entrepreneurship values its own entrepreneurial contribution? Such a combination of the theory of capital with the theory of entrepreneurship is unsatisfactory.

Hence, Menger (1888) did not notice that specific entrepreneurial activity (or the *pure* entrepreneur) is covered by the theory of the 'return on capital' (what Mises later referred to as a *pure entrepreneurial profit*) which is an *ex post* profit (or loss) that shows how well an entrepreneur – exclusively in his capacity as entrepreneur – anticipated future price differentials. Mises (1949) did and added his theory of entrepreneurship, in which entrepreneurs can realize profits by arbitraging the price differential between inputs and outputs which ultimately depend on consumer spending – to his theory of capital, in which capital simply equals *financial net worth*, an entrepreneurial appraisal in terms of money of the asset or assets under his control. Capital is, thus, about appraisement, consumer preferences and capital (profit-and-loss) accounting, not about material capital goods.

A possible critique to Mises's work on capital theory, is his omission of *scale*. With scale, I refer to the idea that the theory of capital can be viewed at different scales: for example, a specific individual productive asset, a combination of assets (that is, a firm), an industry or a country. Mises (1949) limited his theory of capital to the second, that is, the firm. He implicitly assumed that only that scale matters, since in Mises's mind an entrepreneur is an entrepreneur at *the firm level*. He therefore directly controls a given set of productive assets, which leads to a *net worth* and allows for capital accounting on the firm level. In this sense, the firm's capital equals the net present value (NPV) of the future (net) cash flows it is able to generate. However, this same concept can be applied to different *scales*: such as the individual productive asset. A machine, as capital, simply equals the net present value of the future (net) cash flows, or it contribution to a firm's net cash flows. If not, there exists a profit arbitrage opportunity. Moreover, the machine, as it can be replaced, has a replacement value, which is a combination of, again, a smaller *scale* of 'capital'. The prices of the inputs to the machine should equal the price of the machine, otherwise a profit opportunity exists. Hence, the price of the inputs (necessary to replace a machine) must equal the net present value of the future cash flows, or their contributions to the machine's future yields, *they* are able to generate. The concept of "capital," thus, applies at different scales, not merely the *scale* at the firm level.

If we go to the other extreme, that is, *scales* of capital greater than capital at the firm level, we could also conclude that the concept of capital is applicable. The only difference is that, at times, claims on a "country's capital" cannot be bought and sold on the market. However, that such an aggregate notion of capital could be informative, should not be doubted. Mises's (1949) claim, in this sense, is that capital accounting only makes sense when applied to an individual or a group of individuals (the firm-level), but not to a nation or even the entire world. According to Mises (1949), aggregating capital accounts on national level is senseless. The contradictory idea inherent to this reasoning becomes clear when we consider that a nation and even the entire world population *are* groups of individuals. A large group of individuals that is, but it turns the separation of certain types of groups of individuals into an arbitrary value judgment by the allegedly omnipotent economist. When are we exactly allowed to estimate the capital of a specific group of individuals utilizing capital accounting? This seems a theoretical dead-end as there exist no sound criteria for distinguishing between groups.

Mises (1949), for instance, questions whether or not "the country's climate and the people's innate abilities and acquired skill would have to be included in the calculation of national wealth" (p. 217). Yet by recognizing the importance of monetary calculation, which paradoxically Ludwig von Mises recognizes as no other, he already provides the answer to his own question. Perhaps Mises merely remains unaware of his statement's full ramifications. That is, monetary calculation and thus capital comprise all goods that can or could be exchanged on the market (Gunning, 2004). There is no point to estimating something in terms of money when it cannot be exchanged against money on the market. A country's climate is thus not to be included in any monetary calculation. How enjoyable a climate might be, it cannot be exchanged on the market. It is not-transferable and uncontrollable. On the other hand, people's innate abilities and acquired skill could and, in fact, should be capitalized. Labor services are sold on the market against market prices and therefore we can consider people's abilities and skills as the underlying asset that renders such services. Since we can estimate or observe the market value of such labor services, it can be assessed in monetary terms and consequently be used for monetary calculation. It is perfectly legitimate to include an asset's estimated productivity over its lifetime, as in the services it renders, for accounting purposes. Labor skills do have market prices.

Following our line of reasoning, calculating a nation's total capital could provide important insights into the (economic) conditions of an industry, a nation, or any other arbitrary group of entrepreneurs[81]. The information that this would render could be useful in, with the limitations of monetary calculation in mind, comparing the (future) wealth of geographical areas or groups of people. It would, moreover, indicate the damaging effects of defective economic policy far earlier than can be observed[82]. Such applications fall, of course, outside the scope of economic *theory*. But denying that such insights could be valuable would be shortsighted. The concept of a total capital of a conglomerate, an industry, a city, a country, a continent, or any other specific group of people is not meaningless. The meaning of aggregating the capital of individuals or businesses as if they were part of a metaphorical fund of capital is in part to discover and scrutinize any differences in wealth and progress.

Hence, Mises (1949) limits his theory of capital to the firm-level (as he assumes that entrepreneurship only takes place at the firm-level), explicitly dismissing applying the same theory of capital to a smaller *scale* (individual assets, replacement value of individual assets) and to a larger *scale* (an industry's or nation's capital). From the point of view of the Misesian theory of the profit-arbitraging entrepreneur, such explicit exclusions are unfortunate.

---

[81] The present-day phenomenon of index trackers (index ETFs) are only a practical reflection of what we argue here. They are, in fact, a type of aggregate *capital* of many different firms, directly contradicting Mises's objections. Take, for instance, the MSCI Emerging Markets ETF or Vanguard's Healthcare ETF as good examples of the points we have raised here.

[82] In this sense, a stock market, as it directly discounts a government policy in stock prices, can provide an administration with direct and often important feedback on a given policy. If we would have to evaluate a public policy by waiting for its (sometimes very distant) future impact on income, we would be handicapped. On a more practical note, if we follow Mises's logic, a stock market index provides no information at all, since it is an aggregate *capital*.

In addition, Mises's way of distinguishing between entrepreneurs and "promoters" is untenable. An entrepreneur, according to Mises (1949) and very much in line with Clark (1899), "(…) discovers a discrepancy between the prices of the complementary factors of production and the future prices of the products as he anticipates them, and tries to take advantage of this discrepancy for his own profit" (p. 707). Therefore, "[a]n entrepreneur can make a profit only if he anticipates future conditions more correctly than other entrepreneurs" (ibid, p. 291). A "promoter," on the other side, is a pioneer, someone who upsets the price structure and generates 'disequilibrium'. What Mises (1949) in fact attempts here, by distinguishing his theory of the entrepreneur with his theory of the "promoter," is to reconcile his theory of entrepreneurial arbitrage (which leads to a, in practice, never attainable price equilibrium) with Schumpeter's theory of entrepreneurial "creative destruction" that upsets an economic structure rather than bringing it closer to equilibrium. Nevertheless, as we will see later[83], Schumpeter's theory of entrepreneurship is actually a specific application, a subset or category, of the broader (Kirznerian) theory of the profit-arbitraging entrepreneur.

In sum, Mises (1949) rescues the (subjectivist) Mengerian theory of capital in which no reference to material capital goods is necessary. Moreover, Mises (1949) furthers Menger's theory of capital by integrating it with his theory of the (profit-seeking) entrepreneur. Viewing the theory of capital in isolation of the theory of the entrepreneur is pointless: both are intimately related with each other.

Unfortunately, Mises (1949) himself sometimes gets sidetracked from his own non-material theory of capital. Mises (1949) talked about what distinguished a rich from a poor country was the number of capital goods. Or that a country's being richer since they had "more capital goods per worker" (p. 740). Hence, Mises (1949) himself sometimes confuses his own capital theory, straying from his financial theory of capital to a material theory of capital. Moreover, Mises (1949) ignores *scale* in his praiseworthy theory of capital and limits his non-material theory of capital to the firm-level. Yet, these are minor flaws to Mises' (1949) immense work to salvage Menger's capital theory and add to it with his theory of the entrepreneur.

[83] See p. 359.

3.10 A First Attempt to Incorporate the Term Structure into the Theory of Interest

We have seen earlier (e.g., Fisher, 1930; Menger, 1888; Mises, 1949) that a coherent definition and conceptualization of capital is the (net) *present value of all future income*. However, we are then faced with a problem because we have to choose an interest rate among many to discount future income to the present. Since there exists no single uniform interest rate in the economy, but rather a term structure representing interest rates at different maturities, a logical explanation of this difference between interest rates of different maturities is of great importance. Economists, until this first attempt to address the issue by Lutz (1940), have largely ignored the issue. Indeed, as Lutz (1940) observes:

> "It has long been customary in works on the theory of interest to talk about *the* interest rate, and to deal with the problem of the difference between rates on different maturities by adding a footnote to the effect that the author understands by *the* interest rate the whole "family" of interest rates. Although the incompleteness of this kind of treatment was generally recognized, it was not regarded as an essential defect of the theory, because it was assumed that the whole "family" of interest rates moved up and down together, and that furthermore there was a tendency towards equalization of the different rates." (p. 36)

Indeed, none of the other economists involved in the debates on capital and interest seem to have given any thought to the existence of the term structure. Hayek, Mises, Knight, Kaldor and Wicksell seem to have put no consideration into the idea of a "family" of interest rates, whereas Böhm-Bawerk, for example, flat-out rejected such an idea. Fisher (1930) at least acknowledges the existence of the term structure (or yield curve), but assumes the spread more or less as a given without giving it any due consideration. He, too, explains long-term interest rates as a mere extrapolation of future short-term rates (under the assumption of perfect foresight), which later became the basis of Lutz' first comprehensive formulation of the expectational theory of the term structure (Lutz, 1930).

The fundamental question from a theoretical point of view is: what exactly explains the spread between shorter-term interest rates and longer-term interest rates? And what are the ramifications of the origin of such a (yield curve) spread? We have seen that the difference between short-term and long-term rates *was* discussed decades earlier, mostly in the debates on shiftability at the time of the founding of the Federal Reserve[84], but never by economists who were engaged in a lively debate on the interest phenomenon.

Lutz's (1930) explanation of the term structure is very much in line with Keynes's theory of liquidity. Lutz (1930) sees "the long-term rate as a sort of average of the future short-term rates" (p. 38). Hence, long-term rates simply embody the expectations of all future short-term rates. Culbertson (1957) summarizes Lutz's position in slightly different wording: "(…) the expectational theory (…) [a]s developed by John R. Hicks and Friedrich A. Lutz (…) argues that the interest rate on a long-term debt tends to equal the average of short-term rates expected over the duration of the long-term debt" (p. 487). This theory was later coined the 'expectation theory of the term structure'.

Reconciling this theory with the empirical observation that yield curve spreads change, Lutz (1930) explains the fact that long-term rates sometimes move contrary to short-term rates by attributing this phenomenon to the expected short-term rates in the medium term. In other words, sometimes long-term interest rates fall and short-term interest rates rise. According to Lutz (1930), this can happen when present short-term rates increase, but expected future short-term rates decrease. Of course, since expectations cannot be observed, this theory will most certainly explain, at first glance, any movement or change in the term structure. It is more or less akin to explaining the league performance of FC Barcelona with the mood of Lionel Messi; if FCB loses, you can deduce that Messi must have felt bad. If FCB wins, it must be because Messi felt well. Nevertheless, Messi's mood is non-observable. Any outcome (that is, any shape or slope of the yield curve) can be rationalized with Lutz' theory *ex post*.

Explaining long-term rates as a series of expected *future* short-term rates, Lutz (1930) makes three key assumptions. As Luckett (1959) summarizes, Lutz (1930) assumptions are:
"(1) Everyone in the market knows what future short-term rates of interest will be, i.e., there is accurate forecasting; (2) there are no costs of investment for either borrowers or lenders; and (3) both borrowers and lenders possess complete shiftability as between obligations of different maturities." (p. 132). While the first two assumptions appear valid assumptions to

[84] For more on the shiftability debate of the early 20th century, go to p. 121.

explain some state of equilibrium, so that we can better understand how (financial) entrepreneurs arbitrage profit opportunities, Lutz's third assumption is questionable.

Is short-term credit a substitute of long-term credit? And is, as a result, the long-term interest rate an expression of short-term interest rates? Moreover, is complete and perfect arbitrage ("shiftability") along the term structure permissible? Lutz (1930) surely seems to imply so.

However, as Culbertson (1957) notes, Lutz implicitly assumes that arbitrage only occurs among lenders, not among borrowers. Put differently, why would under perfect foresight anyone want to *borrow* at long-term rates if long-term rates are merely perfect reflections of future short-term rates, with no other additional allowance or premium? No market for long-term rates would actually arise; the whole necessity to explain long-term rates thus vanishes. Culbertson (1957) effectively refuted Lutz's 'expectations theory of the term structure' on theoretical grounds, even though much later empirical studies would also increasingly contradict Lutz's theory of the term structure (e.g., Modigliani & Sutch, 1966).

Fisher (1930) was also very critical of Lutz's theory of the term structure. Fisher claimed that the term structure of interest rates was a spontaneous outcome of market forces, which cannot be arbitraged away (Culbertson, 1957). As Fisher (1930) wrote:

> "[It is necessary to posit] a theoretically separate rate of interest for each separate period of time, or to put the same thing in more practical terms, to recognize the divergence between the rate for short terms and long terms. This divergence is not merely due to an imperfect market and therefore subject to annihilation, as Böhm-Bawerk, for instance, seemed to think. They are definitely and normally distinct due to the endless variety in the conformations of income streams. No amount of mere price arbitrage could erase these differences." (p. 313)

Since the maturities of investments (as reflected in the term structure) ultimately depend on the underlying cash flows that are a result of consumer decisions as to *on what* and, more importantly, *when* to consume, they can never be arbitraged, since by merely transforming the financial maturities you would not be able to transform the underlying *economic* (consumption) maturities.

More or less at the same time, Keynes (1930) arrived at the same conclusion. Riefler's work (one of the key contributors in the debate on shiftability) was largely accepted by Keynes.

However, in his *General Theory* (1936), Keynes abandoned this theory (Culbertson, 1957). In fact, Keynes turned the theory upside down, basing his new theory on short-term expectations with regard to long-term interest rates (Keynes, 1936). Keynes had a change of heart and began explaining long-term rates as the product of mere psychological considerations (Culbertson, 1957).

Hicks (1939) was among the few that recognized the earlier treatments (e.g., Wicksell, 1898; Keynes, 1936) of interest as an overly naïve simplification. Hicks (1939) wrote: "Economists, in their discussions of interest problems, often talk about the determination of *the* rate of interest. It would seem that they must have some such reduction as this in mind; yet the rate of interest which they discuss is more usually the long rate. (…) *The* rate of interest in Mr. Keynes's *General Theory* is [for instance] the long rate." (p. 148).

Yet Hicks (1939) himself deviated to Lutz's (1930) theory of the term structure. Hicks (1939) wrote that: "[T]the differences between rates of interest (…) arise from differences in the duration of loans." (p. 144). "These also tum out to be partly a matter of risk; but they are also influenced by other considerations." (ibid). Yet, Hicks (1939) in his own analysis, reduces the term structure ("various durations") into a single unique one-week rate, implicitly adopting the 'expectations theory of the term structure'. In Hicks' (1939) own words: "[I]t is possible to build up the whole system of interest rates, using the short rate as unit." (p. 150). Consequently, "the long rate is an average of current and forward short rates." (ibid., p. 15).

To sum up, the entire literature on the term structure of interest rates – despite an extensive treatment of the theory of interest – consisted in Lutz's (1939) 'expectations theory of the term structure.' Critical follow-ups to Lutz's theory of the term structure came rather late. Luckett (1959), for instance, would address Lutz's seminal paper decades later, much closer to (although not explicitly considered in) the third round of controversies on capital (the famous "Cambridge controversies").

Unfortunately, despite recognizing the issue at first, Hayek (1941) decided that the difference between the short end of the yield curve (the more liquid side) and the long end (the illiquid side) was of no importance: "The mobility of capital, then (like the closely connected concept of liquidity), is a magnitude which can be adequately represented only in two dimensions, one giving the range of dates at which the alternative returns from a given resource are obtainable and the other the magnitudes of these alternative returns." (p. 328). In effect, Hayek refers to the magnitude of the loss involved when “liquidating”. Yet Hayek (1941) then turns his back to his own definition as: “(…) probably very few useful generalisations can be made about the problem as a whole, except for the negative statement that any sharp division into two distinct categories of capital goods, such as circulating capital and fixed capital, is likely to do more harm than good.” (ibid). Of course, under perfect foresight there is no need for liquidity (nor money) for households, yet in reality at least to some extent households hold their savings in liquid, short-term monetary assets such as a demand deposit. If, however, these savings are invested in illiquid, not-mobile capital, problems arise.

3.11 Macaulay’s Contributions to the Term Structure, Maturities and “Duration”

In 1938, Canadian economist Frederick Macaulay published his seminal work *The Movements of Interest Rates*. If only economists would have paid heed to Macaulay, who had an actuarial background, no Cambridge controversy on capital would have existed (we will show why at a later point).

The motive of Macaulay’s close interest in interest rates was very practical: Macaulay (1938) looked for a way to accurately calculate the impact of a change in yields (interest rates) on the price of a bond. A quick example should clarify Macaulay’s principle argument. Take, for instance a 10-year bond with a face value of $100 and a coupon yield of 8% selling at $114 compared to a 10-year bond with a face value of $100 and a coupon yield of 6% selling at $100. Now, both bonds have a similar maturity (10 years), but that does not imply that both bonds will have an equal price increase/decrease in response to a change in market interest rates. Therefore, we need some type of weighted-maturity to know (a) how bond prices will adjust and (b) how long it takes for the initial cost of the bond (or the initial outlay of an investment project) is repaid. If it takes longer before the principal is repaid, *ceteris paribus*, the risks are greater.

Macaulay (1938) was one of the first to propose a solution to the interest rate and maturity question: “For many economic purposes the 'yield' of a bond must be considered as an *average* of various rates of interest used during successive future periods.” (p. 29). In effect, Macaulay (1938) proposed to “weight” the maturity to get to a *duration* different from mere maturity. As Hallerbach (1999) summarizes: “Since the weighted average time-to-maturity takes into account the timing of all cash flows and not only of the principal, duration is more meaningful in this respect than (…) maturity.” (p. 2). Hicks (1939) and Samuelson (1945) would later elaborate on Macaulay's (1938) initial work by showing how bond prices and interest rates interact (put differently, their work involved examining the elasticity of bond prices to changes in yield)[85]. Thus, Macaulay (1938) used the term "duration" to indicate the time element of a loan or investment, which essentially boils down to an adjusted term-to-maturity. In discussing the difference between "maturity" and "duration," Macaulay (1938) remarked that:

> "It is clear that 'number of years to maturity' is a most inadequate measure of 'duration'. We must remember that the 'maturity' of a loan is the date of the last and final payment only. It tells us nothing about the sizes of any other payments or the dates on which they are to be made. It is clearly only one of the factors determining 'duration'. Sometimes, as in the case of a low coupon, short term bond, it may be overwhelmingly the most important factor. At other times, as in the case of a long term, diminishing annuity, its importance may be so small as to be almost negligible. Because of its nature, length of time to maturity is not an accurate or even a good measure of 'duration'. 'Duration' is a reality of which 'maturity' is only one factor." (pp. 44-45)

The theory of (Macaulay) duration has some key implications for the theory of capital:

(1) The lower the coupon yield, the higher the convexity. In other words, if interest rates are low, any increase in interest rates will have a stronger impact on the bond price then if interest rates were higher;
(2) The duration is always shorter than the maturity of an investment; yet how much shorter depends on the timing of the cash flows;

---

[85] At this point it is important to note that Samuelson (1945) still "believed" in the Böhm-Bawerkian notion of interest rates affecting the profitability of investments with a certain degree of "roundaboutness" and time-to-maturity (in short, their *duration*). Samuelson would later discard the theory of *duration* because of the observed phenomenon of reswitching, which was one of the most important points debated in the third round of capital controversies (and Samuelson was one of the key participants). Nevertheless, we will see that the phenomenon of reswitching is largely based on a misconception, see p. 242. One could say that there exists a big gap between the pre-WWII and post-WWII Samuelson.

(3) The concept can be applied to both individual (productive and financial) assets as portfolios or combinations of assets (for example, a firm).

In our above example, with two bonds with identical yields (for the first you have to pay more but it has a higher coupon yield) and identical maturities, the duration of the first bond in our example would equal *7.44 years,* whereas the duration of the second bond would equal *7.80 years*[86], the difference being that the former with a coupon yield of 8% returns the capital invested sooner than the latter.

Clearly, this is very much in line with a financial interpretation of 'capital-intensiveness' and the 'period of production' to which we referred many times before. In effect, what Macaulay (1938) did for bonds, equally applies to the theory of capital. That is, if we consider both the maturity (i.e., 'period of production', time) *and* how fast an investment is repaid (i.e., 'roundaboutness' or 'capital-intensiveness'), we will get to *Macaulay duration*. **Macaulay duration is the *sine qua non* of capital theory**. Capital-intensiveness ('roundaboutness') and the (average) period of production make no sense when considered from a technical or material point of view, but do when they are considered from a subjective, financial point of view. This was also, essentially, the view that Samuelson (1945), for instance, adopted, until he abandoned the theory in the third round of controversies on capital.

Moreover, Macaulay's (1938) insights are also extremely valuable in light of the earlier debates on shiftability and banks' maturity mismatching. The longer the maturity of financial asset, the greater the fluctuations in its price are. Therefore, the greater the price fluctuations over a short period by buying and selling a financial asset with a long maturity (Macaulay, 1938), compared to simply buying a short-term financial asset which is held until maturity. Macaulay (1938) himself concluded, with regard to the practice of maturity mismatching by banks: "Unlike short time loans, long time loans are not 'self-liquidating'. Prior to its distant maturity, nobody has to buy or retire a particular long-term bond at a particular time or go into bankruptcy. **This is why it is so peculiarly inappropriate for banks to place any large percentage of their demand funds in long term bonds**." (p. 43).

[86] Although duration is often expressed in years, Hallerbach (1999) argues that it should be interpreted correctly as a (net) present value. For our purposes, the difference is close to irrelevant.

Curiously, the *duration gap* is widely used by financial intermediaries such as banks and pension funds to measure duration mismatches between assets and liabilities. However, it is mostly treated as an indicator of net worth sensitivity (and thus future solvency risks). A duration gap of 0 means, simply, that the cash inflows and cash outflows are matched to such an extent that net worth cannot be impacted by changes in interest rates. A positive duration gap (with longer durations of assets than liabilities) would cause assets to be more “sensitive” to interest rate changes, which would reduce the bank’s net worth (solvency). However, liquidity, as opposed to solvency, is often not considered in this context. Managing the timing of cash inflows and cash outflows as an economic principle of (aggregate) liquidity instead of a narrow banking principle of (individual) solvency is entirely disregarded[87].

To sum up, Macaulay’s (1938) made a key contribution to economic science by providing a financial (subjectivist) foundation under the notions of the ‘period of production’ and ‘capital-intensiveness’. Moreover, Macaulay (1938) clearly identified the problem of banks’ engaging in maturity mismatching (investing in long-term loans or securities). Moreover, banks’ modern-day practice to match the durations of assets and liabilities using derivatives protects banks against solvency issues, not liquidity issues. And, last but not least, at low rates of interest, there is high negative convexity to increasing interest rates, as any rise in interest rates disproportionately affects the present value of investments and financial securities[88].

As such, the concept of duration was first developed by Macaulay (1938). Thereafter, it was occasionally used in some applications by economists (Hicks, 1939; Samuelson, 1945), and actuaries (Redington, 1952).

Curiously, as Osborne (2014) would later recognize: “[B]y its very construction, the formula captures pattern in the cash flows: if the bulk of the cash flows are returned early in the sequence (they are front-loaded) then [Macaulay Duration] is low, and if the bulk of the cash

[87] This closely relates to many banks’ attempt to “hedge” themselves to higher interest rates by using long-dated interest rate swaps. However, long-dated interest rate swaps protect a bank against solvency issues, not liquidity issues. We will analyze the difference in greater detail in Section V: Capital in Disequilibrium, p. 408.

[88] In this light, one should be very fearful for rising interest rates after years of historically record-low interest rates.

flows are returned late in the sequence (they are back-loaded) then [Macaulay Duration] is high." Osborne (2014) would develop an approach that includes using the unorthodox rates of any polynomial function, such as a net present value calculus; by including the unorthodox rates in a cash flow schedule, the effect is very similar to applying Macaulay *duration* to the factor discounts. This modern-day vindication of Macaulay (far beyond, possibly, his own wildest dreams) is of utmost importance in this context, but needs further development before it can be fully integrated and used in both theory and practice.

3.12 The Dao of Corporate Finance

As Spitznagel (2011) mentions, Williams (1938) explains that a stock's "value" depends on its "discounted value of its stream of future dividends," which was later popularized into the discounted value of its stream of net cash flows (which are either invested back into the business if returns are high or returned to its respective shareholders in the form of dividends, which was what Williams considered when he valued stocks). Hence, Williams (1938)'s work is a primer in the net present value approach, which is defended by this work. It is essentially forward-looking (as it involves the future yields and income that an investment is able to generate). In basic terms, Williams (1938) proposed that the value of a combination of productive assets (that is, the value of a firm) is equal to (Spitznagel, 2011):

$$Value = \frac{Dividend}{y - g}$$

Where:

Dividend = Expected dividend

y = Discount rate

g = Constant growth rate of dividends

In a certain sense, it is surprising that Fisher (1907) first published on capital as the discounted value of a stream of future, before Williams (1938), albeit not as a theory of investment value, but as a starting point to a coherent subjectivist theory of capital.

Now, in modern-day applications, dividends have simply been replaced by (net) cash flows, but the underlying valuation principle remains the same. In addition, in equilibrium, we would expect the price of firm to equal its value (Williams, 1938). Moreover, more broadly put, we would expect the price of capital to equal its value[89]. As Spitznagel (2011) proves, this is equal to saying that we expect the implied return on invested capital to equal the weighted-average cost of capital: “We are thus separating the wheat from the chaff, whenever $\frac{Value}{Invested\ capital}$ is high it simply and precisely means that implied ROIC exceeds WACC, and this is really all we need to know.” (p. 3).

This “law” of investment value not only applies to capital at the firm-level, as its initial application proves. We could easily rephrase the “dao of corporate finance” to the “dao of capital.[90]” If we take an individual productive asset, for instance, as a narrow type of “capital,” we will see that as we define the value of that capital as its present value of its future contributions to yields or present value of its yields, we have defined the implicit, capitalized ROIC of that asset. Now, the WACC of that very same capital, are simply the prices of the assets necessary to create the productive asset, i.e., its replacement value. Yet, its very replacement value (its WACC), is equal to the net present value of its inputs, that is, the present value of the contribution of each of the assets to the productivity of the productive asset it renders. This, of course, applies then to *all* assets in an economy and *all* combinations of capital. In equilibrium, no price differentials exist (net of interest) between the various assets: every asset’s price reflects the net present value of its future (monetary) contributions. Hence, the concept of capital can be applied to *any* scale. Whether there will be arbitrage, however, and bilateral price equality in equilibrium, depends on whether the ownership of a given combination of assets (that is, capital) can be exchanged on the market.

The work of Williams (1938) is noteworthy, since it foreshadows a later contribution by Tobin and Brainard (1976) to connect capital to its replacement value. Yet, replacement value simply equals capital, but at a different *scale*. And there will be arbitrage in-between different

[89] As we have discussed earlier, capital can be capital at any scale: the assets to create a specific productive asset, a single productive asset, a firm (a combination of productive assets), an industry, a country, etcetera. See p. 96.
[90] This is, coincidentally, also the title of Spitznagel’s book (Spitznagel, 2013).

scales[91]. An example is the equity *q* ratio, where the market price of the equity of a set of firms (in this case, the firms that make up the S&P 500) is contrasted with the replacement value of the equity of the same set of firms (that is, the underlying value of the productive assets needed to replace the equity).

That is, if we look at a S&P 500 index ETF, then the price (assuming equal units) of the S&P 500 index ETF should equal the price of the equity of the underlying firms which should equal the price of the underlying productive assets held by these firms, which should equal the price of the productive assets needed to replace these productive assets. Since all these prices, in equilibrium, equal the present value of their future contributions to yields or yields, we have circled back to Williams's (1938) valuation principle:

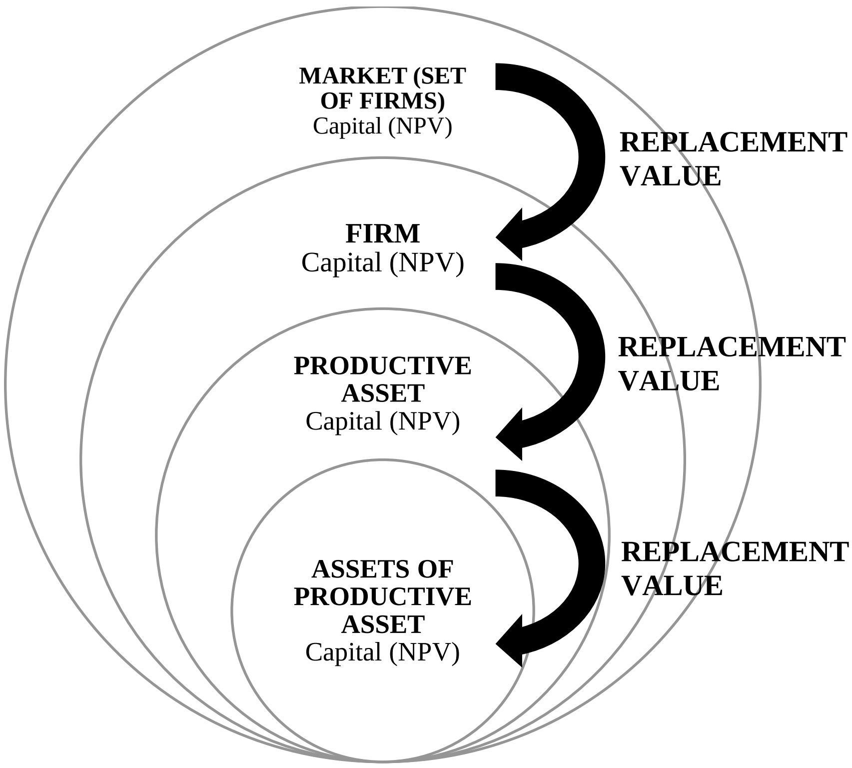


Figure 26: The different scales *of capital: the dao of capital is the NPV approach.*

[91] This applies all the way to individual commodities: in equilibrium, their prices both reflect their contributions to the yields of the investments (productive assets) in which they are used, while the prices of the assets used to extract or mine a commodity should be, in sum, equal to the NPV of the commodity's contributions. It is always important to note that, ultimately, all yields depend on (present and future) consumer spending.

Chapter 4: The Third Round (1953 – 1968): The Cambridge Controversies

4.1 The Neo-Ricardians versus the Neoclassicals: The Closing Fight on Capital

The last round of controversies on capital, the famous "Cambridge controversies," came off the back of Robinson's publication in 1953 and Solow's and Swan's articles on growth in 1956. Solow belonged, of course, to Cambridge, U.S., while his main adversaries were based in Cambridge, U.K.: hence, the name Cambridge controversies. Solow was mainly interested in questions of economic growth and development, since the Second World War came to an end but left the (developed) world in shambles. What determines the long-term growth of an economy?

This third round of capital controversies, in hindsight, marked a peak in constructivist thought. Constructivism, according to Hayek (1973), is the notion that all institutions in a society are "deliberately constructed by somebody" (p. 10). Constructivists "habitually argue on the assumption of omniscience" (p. 13). The narrow focus of economics of growth models and forecasting, led to a series of newly proposed economic growth models: the Harrod-Domar model, the Solow-Swan model and the Mankiw-Romer-Weil model, among others.

Joan Robinson, who was working at the Economics department of the University of Cambridge (U.K.), made her first contribution to capital theory in 1953. Her article *The Production Function and the Theory of Capital* (1953) could be rightly called the first punch in the Cambridge fight over capital. Her article was promptly followed by an (indirect) response by Solow (1956). Robinson and Solow ended up publishing books, for different reasons, with a lot of overlap in about the same time frame. In 1956, Robinson ended up publishing her most important work on capital, *The Accumulation of Capital*. This book was an attempt to extend the Keynesian theory into the long-run[92].

Solow, on the other side, published an article on growth theory, titled *A contribution to the theory of economic growth* in 1956, followed up by publications on technology and investment in 1957 and 1960 and the birth of the famous Solow-Swan model in 1956 (the

[92] As we have seen earlier, Keynes's *General Theory* was largely geared to short-run disequilibrium, p. 182.

model is partly named after the Australian economist Trevor Swan, who never got too much involved in the Cambridge debates). Most of Solow's work would culminate in his book with Paul Samuelson, *Linear Programming and Economic Analysis* (1958), revealing Samuelson's growing interest in capital theory. This eventually led Robinson to publish her response to Solow's work in 1962, *Essays in the Theory of Economic Growth,* and she would later elaborate what would later become known as the "Cambridge growth theory" together with Nicholas Kaldor.

Alongside Joan Robinson and Nicky Kaldor, Piero Sraffa played an important role in the Cambridge capital controversies. Sraffa already participated in the second round of capital controversies (the so-called Sraffa-Hayek controversy). Sraffa (1960) tried to salvage Ricardo's theory of value (grounded in the labor theory of value), in stark contrast to the neoclassical theory of value, and is often credited with beginning the Neo-Ricardian school of thought.

By this time, Friedrich Hayek had already left the London School of Economics for the University of Chicago. As Johnson (1982) notes: "To understand [this] controversy, one has to understand something about the politics of academic debate in England; more particularly, to appreciate the sharp division between the London School of Economics and Cambridge over Hayek versus Keynes in the 1930s (…). But by the 1950s (…) there was no one left in London either capable of or interested in debating pure "capital theory" with Cambridge. Had it not been for Cambridge (U.S.A.)—I mean MIT not Harvard—responding eagerly to Joan Robinson's challenge to "orthodox" production function theory in order to display its mathematical-economics muscle." (p. 121).

Hayek's absence was rather damaging to the reputation of the Austrian School of economic thought, since much of the criticism fell on the shoulders of earlier 'Austrian' thinkers. It should be no coincidence that the Austrian School was deemed refuted and irrelevant, especially after Samuelson (1966) argued that the reswitching phenomenon debunked the 'Austrian' theory of capital and interest.

Robinson's primary goal was, according to Johnson (1982), to show that capitalism could not possibly work. By destroying the rational foundation of theories behind profit maximizing, income distribution and price determination, Robinson could show that prices are at mercy of the whims and fancies of entrepreneurs and capitalists. Nicholas Kaldor largely joined

Robinson in her criticism of the neoclassical production function, albeit for other reasons. The name Kaldor should ring bells, since he was one of the participants in the second round of capital controversies (e.g., Kaldor, 1937; Kaldor, 1939), next to the earlier mentioned Piero Sraffa.

Yet Kaldor was in it for other motives than Sraffa. As Johnson (1982) argues: “Nicholas Kaldor once commented that if you really believe in capitalism, it is worthwhile doing all the work required to explain how it functions. If you do not believe in capitalism, it is not worthwhile exploring how it is supposed to work in order to show why it doesn’t. But to do all that work in order to show that it cannot work is a waste of time” (p. 122). In contrast to the two earlier debates, which were highly theoretical and apolitical, the Cambridge controversies were clouded by ideological convictions.

Kirzner (1996) reached the same conclusion. He writes: “The ideological element in the main-stream approach here being attacked is thus the ‘consumer sovereignty’ claim held to be implicit in [the mainstream neoclassical] approach. The Cambridge (UK) critics, denying the supply-and-demand determination of income distribution under capitalism, were denying that capitalist income shares benignly express (and cater to the satisfaction of) the preferences of the consuming public.” (p. 5).

In summary, the Cambridge controversies on capital of the 1950s and 1960s evolves around three basic premises:

(1) One cannot add up the monetary values of capital goods since their monetary value depends on a discount rate, that is, the rate of profit (the rate of return on capital). This boils down to a circular argument: the money value of capital depends, according to the neoclassicals, on the rate of return on capital, yet the rate of return on capital depends on the money value of capital. Therefore, the former cannot determine the latter and the latter cannot determine the former. We will discuss this in the fifth paragraph of this chapter.
(2) If a financial variable, that is, the discount rate (according to the neoclassical interpretation of capital the rate of return on capital) determines the “dollar-valued” capital stock, any physical interpretation of the money capital stock becomes self-contradictory. The capital stock could fall or rise whenever a financial variable (i.e., the rate of return on capital) changes. Therefore, the dollar value of the capital stock is not

interpretable as the physical capital stock. We will discuss, largely, in the second paragraph of this chapter.

(3) Any fall in the rate of interest (that is, according to neoclassical interpretation, the rate of return on capital) will lead to a longer "period of production" and more roundabout production methods. However, at one point, if the rate of interest falls further, some projects "switch back" to a shorter "period of production" and a less roundabout This is known as the reswitching paradox which allegedly disproves both the 'Austrian' theory of capital and the neoclassical theory of capital of the participants involved (e.g., Solow, 1963). We will discuss this in the third paragraph of this chapter.

As Desai (1991) briefly summarizes: "The Cambridge-Cambridge controversy was fought around the issues of capital heterogeneity, capital labor substitutability and the relationship between the interest rate (profit rate) and capital." (p. 56).

The entire neoclassical theory of economic growth[93], as it is grounded in a *physical* or *material* theory of capital, is therefore fallacious. Indeed, Cambridge, U.K. concluded that the claim that richer countries became rich by using "more" capital was meaningless[94], especially since many high-income countries have relatively little *physical* capital but actually have a large service economy. This observation was later dubbed the "Leontief

[93] We will refer to any user of the production function as "neoclassical," although strictly taken, Solow is generally viewed as a "neo-Keynesian." Neo-Keynesianism is, however, a mere synthesis and continuation of prior neoclassical thought with some additions, such as the IS-LM (investment saving – liquidity preference money supply) model and the Philips curve, which for the scope of this work is largely irrelevant. Neo-Keynesianism is different from, for example, monetarism. Nevertheless, the disagreements between neo-Keynesians and monetarists largely revolve around monetary policy and inflation. Both lines of thought would fall under our use of "neoclassical" as both would either use or agree with the fundamental theories behind the (Cobb-Douglas) production function and the economic growth models that are based on it.

[94] Even Mises (1949) was partly swayed by this physical or material fallacy. In *Human Action*, Mises writes: "(…) every increase in the supply of capital goods available results in a lengthening of the period of production (…)." (p. 495)

paradox" (Leontief, 1953), a paradox that vanishes as soon as we interpret capital *not* as something physical and include "human capital." [95]

It is vital to note that the insights of Menger (1888), Fisher (1930) and Mises (1949), at this point, have been completely forgotten. The view that capital is a financial concept of *net worth* was upheld by not a single participant in this last round of capital controversies. Moreover, the time preference theory of interest had also completely vanished, long forgotten by the both Cambridge schools who debated capital.

In addition, it is important to note that Solow's work[96] was largely a continuation (or critique) to the earlier Harrod-Domar growth model (Harrod, 1939)[97]. That model assumed: (a) capital is a 'stock' of physical capital goods that yields an output; (b) the marginal product of capital is constant (no diminishing returns on capital, in other words, a fixed capital-output ratio); (c) capital is a prerequisite to have output; (d) part of output is saved (savings rate multiplied by output) which equals investment and adds to the 'capital stock' as long as it is in excess of depreciation; and (e) any change, net of savings, in the 'capital stock' amounts to a fixed rate of depreciation. Hence, to grow an economy at a faster rate, you can either increase savings, increase the product of capital, or decrease the depreciation rate. The Harrod-Domar growth model, more than anything else, sparked the interest of various neoclassical economists who later proposed their own growth models.

4.2 Solow's, Swan's and Samuelson's Neoclassical Sins and Virtues

In a sense, Solow, Swan and Samuelson abandoned the work of Frank Knight (1934) and Fisher (1930), joining the ranks of Böhm-Bawerk (1888). In line with Böhm-Bawerk, these neoclassicals defended capital as a heterogeneous mix of different capital goods, supplemented with labor and technology (which are largely exogenous factors since they are

---

[95] Later economists "solved" the Leontief paradox by distinguishing between skilled labor and non-skilled labor, which proves our point regarding the error to exclude human capital from capital (e.g., Böhm-Bawerk, 1888; Kaldor, 1938)

[96] Solow's background was mostly in statistics (econometrics) and (Leontief) input-output models, which would mark profoundly the mathematical intensity of the third round of debates on capital.

[97] Both the Harrod-Domar and Solow-Swan models were named after their creators not as collaborators, but as independent researchers that arrived simultaneously at (more or less) similar conclusions.

determined outside their models; population growth and new inventions are, in that sense, "external shocks"). Ideally, therefore, production functions should be used to reflect the wide variety of different capital goods. In fact, Solow (1963) himself criticized the production function for this oversimplification of reality. However, as econometrics does not allow for such variety, a capital aggregate (*K*) is allowed for all practical purposes, as Solow (1963) defends. How else should theory be applied to practical matters such as economic growth forecasts? As Birner (2002) argues, "Solow the econometrician was not inhibited by the results of Solow the theoretician (p. 93).

4.2.1 A Brief Overview of the Solow-Swan Growth Model and Its Flaws

We will now briefly summarize the growth models as presented by Swan (1956) and Solow (1956) in their respective papers. We will recognize that both models are mere applications of the earlier Cobb-Douglas production function. Curiously, as we should remember, Cobb and Douglas (1928) never intended their production function to reflect any "economic" relation, but merely the physical productivity of, in their example, the manufacturing industry. Nonetheless, Swan (1956) and Solow (1956) applied this fallacious model to whole economic structures in their later empirical econometric work. Swan (1956) assumes in his version that:

(1) The 'capital stock' is a stock of heterogenous capital goods measured in dollar terms;
(2) Savings equals a given ratio to output (output equals income), that is, sY;
(3) Savings represents the annual addition to the 'capital stock' (with the annual growth rate of capital being $s\frac{Y}{K}$);
(4) Perfect competition assure that the wage and profit rate equal the (physical) contributions of capital and labor to output (and are thus expressed as a share of output);
(5) All savings are profitably invested and there is full employment;
(6) The rate of growth of the labor force, technological process, and the savings rate (s) are exogenously determined;
(7) The capital-output ratio moves toward equilibrium, which at any given time is determined by technological progress and the rate of labor growth.

In essence, the Solow-Swan model (Solow, 1956; Swan, 1956) is almost the same as the Harrod-Domar model, *except* for the fact that it assumes diminishing returns to capital input (that is, *not* a fixed capital-output ratio). Hence, if accumulation of capital exceeds the growth in the labor force, diminishing returns of capital to output are the result.

The implications are manifold: first, increased savings does not increase the level of output per head; second, capital accumulation does not lead to equilibrium (the "stationary state"); third, technological progress is *prima facie* responsible for growth in output; fourth, money is neutral; fifth, consumer preferences (of output) do not change, *ever*. While Swan (1956) notes that capital might be prerequisite for technological progress (and so the real effect on increasing capital might not show), he does not consider this point in any greater detail.

Subsequently, Swan (1956) considers land as something separate from capital, and assumes that the land is a "fixed factor of production." While land does not serve as an "input" in Swan's model, it does reduce the returns to scale of capital and labor (that is, where capital and labor earlier added up to 1, now capital, land and labor add up to 1, yet land does not add to output). Hence, including factors such as land that cannot be increased in (physical) quantity, simply reduces the returns to scale of the other "augmentable" inputs, such as labor (population growth) and capital (savings).

In addition, Swan (1956) lays the foundation for differentiating the natural rate of growth and the observed rate of growth, as he takes Harrod (1939) initial work and applies it to his own production function, in which he thus distinguishes between a "warranted" and "natural" rate of growth (of output). That is, a decreasing output-capital ratio (which the Harrod-Domar assumes fixed) would lead to a lower rate of profit, which would bring wage rates back to their "appropriate" levels.

Solow (1956) devised roughly the same growth model as Swan (1956). Solow (1956) assumes:

(1) One commodity-output;
(2) Savings is a share of output (sY);
(3) The 'capital stock' (K) is an accumulation of the share of savings of past output;
(4) Net investment equals the rate of increase in the capital stock ($K = sY$);
(5) Output is net of depreciation of the capital stock;
(6) There is no, what Solow (1956) calls, "nonaugmentable" resource such as land;
(7) There are either natural constant returns to scale (of inputs, in the case of assumption six); in absence of the previous assumption, that is, if land is not scarce, then decreasing returns to scale in both capital and labor will be assumed;
(8) The labor force (L) is an exogenous factor that increases at a constant rate n;

(9) Diminishing marginal productivity of capital (K).

Solow (1956) observes that, as a result: “If the initial capital stock is below the equilibrium ratio, capital and output will grow at a faster pace than the labor force until the equilibrium ratio is approached. If the initial ratio is above the equilibrium value, capital and output will grow more slowly than the labor force. The growth of output is always intermediate between those of labor and capital.” (p. 71). All in all, the Solow-Swan growth model is a mere application of the earlier discussed Cobb-Douglas production function:

$$Y = F(K, L)$$

Where, simply:
Y = Physical output
K = Capital stock
L = Labor (hours)

Whereas Harrod (1939) assumed a fixed proportion between labor and capital inputs, the Cobb-Douglas production function assumes marginal diminishing returns.

We can now derive the key tenets of Swan’s and Solow’s capital theory:

(1) Inputs and outputs are physical and material; the ‘capital stock’ is a complex of heterogenous production goods that, nevertheless, is measured in terms of money, labor are physical labor units measured in terms of labor hours, and output is some combination or a single commodity *as if any output* is valuable to consumers;
(2) The rate of interest is equal to the rate of profit (as share of physical output);
(3) The wage rate is equal to its contribution to physical output;
(4) There is no money or, if money is assumed, it is neutral;
(5) Savings are physical, not financial, and the savings rate is exogenous and thus does not depend on, for instance, the rate of interest.

Essentially, Solow (1956) and Swan (1956) are wrong on economic growth. Economic growth is not an increase in a physical capital stock that leads to greater labor productivity and higher physical output per capita. But this is not the economic problem (e.g., Kirzner, 1960) and completely ignores the idea of “coordination” as a cause of economic growth; profits, after all, are a sign of discoordination. The economic problem is an increasingly efficient use of (manifest and latent) resources in function of the individual aims of a society's members. For instance, the coordinating role of peer-to-peer and sharing apps such as Uber and Airbnb do not add to the *physical* capital stock (cars and houses), but they *coordinate* the use of such physical objects more efficiently given certain consumer preferences and thus add economic value. In other words, Solow (1956) and Swan (1956) defy the very goal of economic science. Paradoxically, Solow's inputs, if measured in prices, is equal to Solow's outputs. In a state of equilibrium, Solow's "capital" input would simply equal the present value of future output. This leads to a circular reasoning, recognized and criticized by Cambridge, U.K.

What the production function in fact is, is *a* basis of a *narrow* theory of production (or technical optimization). What Solow and Swan did was take this narrow theory and apply it to an aggregate economy and turn it into a *general* theory of economic growth and progress. However, this is inadmissible. What is true for one firm, is not true for an entire economy. A rate of increase in a firm's profit indicates that the firm is doing a good job in managing scarce resources. A rate of decrease in an economy's profit level, in stark contrast, would indicate a "healthy" economy, as arbitrage opportunities are being arbitraged away and we get closer to an unassailable equilibrium, where no further exchange can improve the situation of a single individual (in a non-Pareto sense).

Circling back to the earlier defined Solow-Swan model, Solow (1970) cites the following growth facts:

1. Real output per worker (per hour) grows at a constant rate over long periods of time;
2. The capital stock grows at a constant rate and exceeds growth in the supply of labor (that is, exceeds in crude terms population growth);
3. Real output and the capital stock grow at an equal rate; the capital/output ratio remains stable;

Yet, there are severe problems with the Solow-Swan model (Solow, 1956; Swan, 1956), also defended later by Samuelson. We will briefly summarize the main troubles with the modern-day application of the neoclassical production function:

(1) The Solow-Swan model does not show or elucidate the role of financial intermediation. As such, it does not show the value added of financial intermediation in the accumulation of capital and investment of scarce savings. Indeed, savings are "physical" (not financial)

and the rate of interest equals the rate of profit which is also, thus, “physical” rather than financial;

(2) The Solow-Swan model does not show or elucidate how or why distortions in capital markets could end up *destroying* capital. As such, the Solow-Swan growth model is incapable of differentiating sustainable economic growth from unsustainable economic growth, with huge practical consequences (even legitimatizing and vindicating bad policy). Solow (1956), for instance, assumes a single-commodity output that is consumed and demanded *ad infinitum*. No *value* problems arise in the Solow-Swan model; *economic value* takes a backseat to physical production. Consumers have no preferences and if they have preferences, any real-world problem of satisfying consumer preferences is completely assumed away;

(3) Moreover, the Solow-Swan model defines “savings” as a share of output that is then added to the “capital stock.” This ignores the intricacies involved in the composition of savings, especially with regard to time. Capitalists (savers) actually make decisions with regard to the maturities at which they wish to save: the Solow-Swan model, however, completely ignores the term structure and its relation to savings. There is a fundamental and important difference between “savings” at a three-month maturity and “savings” at a ten-year maturity, which is completely ignored by the neoclassical production function and theory of growth. It seems that the Solow-Swan model implies that people save at perpetuity without reference to why they would do so;

(4) Capital is, loosely, operationalized as the purchase cost (in money terms) of all capital goods, excluding labor[98] and technological innovation (capital thus comprises the prior accumulated “capital stock” plus a “flow” of savings/investment). However, this “dollar amount” (or amount in money terms), at least in case of the “capital stock,” is a historical amount which is then depreciated at a fixed rate of depreciation. In addition, the model assumes that any new investment has a positive return. As a consequence, the entrepreneurial function is completely ignored (the entrepreneur is primarily responsible

[98] The Mankiw-Romer-Weil version of the growth model includes the stock of human capital, which depreciates at the same rate as “physical” capital.

for his or her investment), as well as the institutional setting in which the entrepreneur acts[99];

(5) The Solow-Swan model unjustly contributes to the negative public opinion of the "rich getting richer." A prime example is Piketty (2014), in which capital is essentially backward-looking, has no relation to consumer valuations and is permanent since no losses occur (or are allowed for in the model);

(6) Labor is defined as labor hours spent or effective labor hours spent, which has no bearing on economic value creation. What counts is the value of labor, which can and should be capitalized resulting in "human capital." However, the notion of human capital is either disregarded or badly integrated in practice by explaining away and attributing econometric residuals to some intangible "human capital"-measure that is operationalized by (public) educational spending and/or college degrees (again, historical cost);

(7) There exists an optimal "capital-labor" ratio, which makes no sense, since any labor is capital and capital is not an input but a tool of entrepreneurial economic calculation. At some instances, the optimal capital-labor ratio does not depend on capital K, but on specific human skills that can be used on the same, previously existent capital goods (for instance, advances in human knowledge and abilities with a same given set of equally capable computers or hardware);

(8) Given the manner in which the law of diminishing returns is applied to capital, K, means that the model logically implies that sustained economic growth is impossible. Even worse, the assumption of diminishing returns on capital, and a fixed rate of depreciation, implies that output eventually ceases to grow. In other words, new investment, according to the explanation of the model, will not be enough to offset depreciation charges, leading to an impairment of capital and hence lower output;

---

[99] It is rather disgraceful that some authors (e.g., Lucas, 1990) seem to wrestle with the idea that no "convergence" of output from rich countries to poor countries exists. Of course, if you implicitly assume in your growth model that *any* investment is successful, you have just eliminated the whole challenge that economics has set out to resolve. Instead of going from an equilibrium model to the real world and adding complexity to the model by introducing new variables that can "account" for the difference between predicted and observed phenomena, they should first revise the assumptions that went into the model to begin with. Taleb (2012) would call this, "*via negativa*", which "[c]onsists in focusing on decision making by substraction, via the identification of errors. (…)" (p. 65).

(9) The Solow-Swan model is, in practice, an excellent example of fitting the model to the data instead of the other way around (Taleb, 2012). The model is assumed correct, while we know it is theoretically full of leaks and holes, and hence the "residual" is explained by means of "soft factors," according to the likings of the investigator. Moreover, the model only projects some assumed or projected long-term growth; since the model is fitted to the data, it is completely incapable to forecast economic downturns and underestimates growth rates in times of (artificial) economic booms;

(10) Technology and innovation are a completely exogenous variable in the Solow-Swan growth model, but it is actually captured in our inclusion of human capital into the theory of capital and requires capital to emerge (that is, Bill Gates would not have created an operating system if he did not have the access to a computer, that is, capital). Greater capital accumulation allows for and funds more research and development. Solow's model (1956) implicitly assumes that any investment in R&D would be a waste of money. Swan (1956) recognized this possibility, but did not give it any second thoughts;

(11) Solow's growth model assumes closed borders and absolutely no mobility of capital; that is, a nation can use and attract the capital of another country's citizens to finance production. There is no cross-border movement of capital and institutional ("soft") variables are excluded from the analysis, while technology as some mysterious and impromptu residual is overemphasized.

(12) As we have seen earlier (see page 183 for example), human labor is always considered a subtractive element and never considered an additive element. In augmented Solow-Swan growth models, an increase in the population lowers the capital invested per worker, adding to misleading static models of unsustainable population growth. The model indeed predicts lower levels of income (output) as a result of a higher rate of population growth, since every human being added is a net cost, not a net gain in terms of capital. However, even the Malthusian law of population would imply any economist applying the Solow-Swan growth model to first prove that we are past the optimum size of population. As Mises (1949) notes: "As long as the employment of additional hands results in a more than proportionate increase in the returns, harmony of interests is substituted for conflict. (...) An increase in population figures does not curtail, but rather augments, the average shares of the individuals." (p. 663);

(13) In Mankiw-Romer-Weil's growth model, which is a more recent variant of the Solow-Swan growth model, human capital is separated from capital. Human capital is then seen as an input that generates a return on investment in the form of physical output. However,

separating human capital from the rest makes no sense, since in many cases investments in human capital are a consequence of an increase in capital, not a cause. It only makes sense to invest in human capital if the complexity of capital increases. Moreover, it confuses spending for investment. Investment necessarily only adds to capital if its return is positive. You cannot simply take all spending on education in a given country and depreciate it at the same rate as a building to some historical cost. Some of that investment might even yield negative returns and are therefore a net loss to the 'capital stock' (that is, capital) of a country. Yet the idea of negative returns is alien to the neoclassical growth models. The "engineering" approach to economic growth, which is induced by the simplicity of the Solow-Swan and Mankiw-Romer-Weil's model, seems to justify (public) investment in education at any cost since any amount of it yields a greater output[100], which in the end might never materialize, resulting in enormous societal losses.

To sum up, the trouble with the mentioned neoclassical growth model, is the fact that they are used for "social engineering" (e.g., Solow, 1956) and top-down decision-making. They give an illusion of knowledge and turn economics into a science of mechanical optimization (Huerta de Soto, 2010).

The "Golden Rule" savings rate is an excellent example of economists becoming social engineers, reducing mankind to mere pawns on a chessboard. In reaction to the Solow-Swan growth model, Phelps (1961) tried to show mathematically that an "optimum" exists between savings and present consumption. The "optimum" is the rate of consumption that maintains

---

[100] In Guatemala, for starters, it should be clear that an increase in education and investment in human capital makes no sense if that human capital cannot be sold (and used) profitably within the country. It makes no economic sense to train a man in quantum mechanics if the labor markets only offers jobs in the coffee industry, unless the worker moves abroad where he can apply his knowledge and where it can command a market price. The incoherence of all this should be clear to anyone. The local government assumes the costs of the investment, while the worker and another country assume the direct benefits from the investment. However, as the Mankiw-Romer-Weil model implies, any spending on education is be "capitalized" and depreciated in the growth model without any bearing on the value and return of the investment. **Savings invested only add to the capital stock if the return on investment is positive**, which requires entrepreneurs to assume costs for decision-making in a continuous process of "tinkering" (trial and error, assuming costs).

capital per worker constant (that is, the rate of consumption that allows sufficient *s* – or savings – to account for depreciation of capital and growth of population), but does not increase nor decrease capital. From there on, if the rate of increase of capital, or so was thought, equaled the rate of increase of output, and both would grow at exponential rates, a "golden age" could be created (Phelps, 1961). A government would only need to estimate the optimum value for *s* (savings) to obtain the highest possible steady state value for *c* (consumption). While Phelps (1961) would probably succeed in putting a smile on his reader's face as he explains his model with a story that takes place in the fictional Kingdom of Solovia (named after no one other than Solow himself), his theoretical arguments are less convincing. Phelps (1961) commits the same previously summarized errors as Solow. A brief selection:

a) He assumes that the return on capital is a function *of* capital, the mere holding of physical capital gives automatically rise to a return on that capital, which is independent of the market, the way the capital is actually invested and/or the existence and efficacy of entrepreneurial arbitrage. Capital provides a return just as an apple tree provides apples;
b) He assumes that no changes in consumer preferences could possibly occur and therefore capital could never be impaired, or otherwise his "optimal" value for *s* (savings) would no longer be able to maintain the capital stock constant;
c) He assumes time preferences of society are of no importance and can simply be imposed from above by the authorities.

Hence, given the above assumptions, the Solow-Swan growth model (and all its variant) assume an "optimal" savings rate. Then, if a society saves less or more, its behavior is deemed suboptimal and thus condemned.

Phelps' (1961) article, although one would sincerely hope he meant it as a joke, reveals the dangers of the Solow-Swan growth model (as well as later variations on the Solow-Swan growth model) as a social engineer's wet dream: "With Oiko's inspiring words still ringing in their ears, the Solovian people pressed the King for a program to attain the golden-rule path. So the King proclaimed golden-rule growth a national purpose and instituted special levies. Once the golden-rule path was reached, investment was continuously equated to profits and Solovians enjoyed (…) maximum social welfare ever after." (p. 643). It leads to the illusion that a government should ask whether or not an economy is saving "too much" or "too little" and use public policy to impose an "optimal" rate[101]. These authors seem oblivious to the fact that the ratio consumption/savings out of present income is a result of the respective time preferences of consumers; markets for intertemporal exchange will move, in a certain sense, on their own toward an optimal distribution between present consumption and investment. In fact, if they assume their model's flaws, it could very well be that they urge a government to reduce the savings rate under a false pretext with severe macroeconomic consequences, such as lower levels of material wellbeing.

4.2.2 Solow's "Computer Paradox" (Or "Productivity Paradox")

Solow's computer paradox refers to the fact that Solow's residual was close to zero, despite advances in information technology in the 1970s and 1980s. The Solow-Swan growth model attributes a residual, if there is a residual, to technology and innovation. Yet, even if it was clear for everybody that enormous technological advances were made, there was no residual at all. Numerous superficial attempts were made to explain the productivity paradox, by referring to:

Yet after our critique of the Solow-Swan growth model, not much is left to remark regarding Solow's computer paradox.

There is nothing paradoxical about this. We know that Solow-Swan and their adaptations assume a fixed depreciation rate of 10%. However, technological innovation also diminishes the future income that previously used (and now largely obsolete) technology is able to generate. Hence, the present value of a capital stock's future yields is impaired. Paradoxically, then, the rate of depreciation can be much higher in eras of high economic progress. Therefore, the Solow-Swan model overestimates the value of the capital stock (as input) when depreciation exceeds 10%, and underestimates the value of the capital stock (as input) when depreciation is lower than 10%. In other words, overestimating the value of the

[101] This seemingly trivial concept has been used as justification for Japan's extremely high public debt – relative to both other countries as historical levels. Japan's private savings rate was higher than the supposed "optimum", which supposedly justifies the government to make up for the difference. In various parts of the world, consumption taxes as well as capital gains taxes are imposed as "incentives" to save or (in a majority of cases) consume. See, for instance, Chamley (1986) and Frankel (1998).

capital stock with an equal rate of output, would diminish the Solow residual. Sudden shifts in consumer preference would also lead, *ceteris paribus*, to higher rates of depreciation. But since the whole economic problem of coordination and *value* creation is disregarded, such considerations do not show up in the Solow-Swan mode.

The problem of the neoclassical treatment of capital can be divided into three sections:

(1) The treatment of labor (measured in – effective – labor hours) separate from capital

"Labor" can be hardly considered "measurable" by adding up human beings or labor hours: as is well understood by anyone, not all labor is made equal: why, then, is such an assumption suddenly accepted in the Solow-Swan model? Moreover, "labor productivity" is measured by simply dividing the output by labor hours (derived from working age population), which makes no sense, even in the material sense of the Solow-Swan model. Moreover, the attempts include human capital by including (public) spending on education is futile: human capital is simply the net present value of future wages and its value will tend to approach its contribution to total present and future income. Hence, it is not an "input."

(2) The alleged circularity argument against the neoclassical theory of capital

An apparent circularity in the neoclassical reasoning as to the "value of capital" has been an important point of critique by Sraffa (1960), Robinson (1953) and Kaldor (1957). Hénin (1986) explains it as the notion that the "value of capital must be known to determine the rate of profit, yet this rate of profit is assumed given in measuring the value of capital" (p. 198). Hence, this circular logic disproves the validity of Solow's and Swan's attempt to conceptualize capital as the monetary value of production goods (that is, in modern finance terms, the present value of the future income a combination of capital goods is able to generate). Measuring in capital in money terms is a theoretical dead-end. Indeed, Robinson (1953) begins her attack by arguing this apparent circularity in the neoclassical production function:

> "Should capital be valued according to its future earning power or its past costs? When we know the future expected rate of output associated with a certain capital good, and expected future prices and costs, then, if we are given a rate of interest, we can value the capital good as a discounted stream of future profit which it will

earn. But to do so, we have to begin by taking the rate of interest as given, whereas the main purpose of the production function is to show how wages and the rate of interest (regarded as the wages of capital) are determined by technical conditions and the factor ratio. Are we then to value capital goods by their cost of production?" (p. 81)

Robinson (1953) answers the last question with a resounding "yes," effectively retrogressing to the outdated Marxian labor-theory of value. She proposes in one of her first contributions to the famous Cambridge controversies that capital should not be explained by its discounted present value (and inherently forward-looking perspective), but rather by labor-units. Robinson (1953) is effectively tricked into regressing to the long-refuted labor theory of value because of this alleged circularity in the neoclassical notion of capital.

Now, concerning the alleged circularity, it is admissible to take a project's inputs (an individual firm's "capital") and calculate its individual return on capital. It is also admissible to take such capital as a sum of various individual "capitals" and average the rate of return on that capital. What is not admissible, however, is to treat an economy as if it were a single firm. It is true, after all, that a single firm has an output (its individual return on capital) that, in this rather questionable case, at the very same time represents an economy-wide opportunity cost (an economy-wide return on capital), which must be used to value the capital (that is, to discount its future cash flows to the present).

However, what is admissible for a firm (a combination of capital goods) within an economy (that is, to use a rate of interest that depends on an economy-wide opportunity cost to discount the future yield on capital goods to arrive at its present value, which is necessary to calculate a *return on capital*: to calculate a return on capital you need to know the return as well as the value of the total capital). What is inadmissible, on the contrary, is to assume that an economy consists of a single firm with one single rate of profit that is then necessary to value the firm's (and thus economy's) capital by discounting future profits at that very same rate of profit. More simply put, what is exogenous to a single firm, is endogenous to the system. Consequently, assuming that an economy *consists* of a single firm, then leads to fallacious conclusions.

Moreover, we should introduce the idea of *ex ante* and *ex post*, which is a concept proposed by Shackle (1989). To determine the *ex ante* value of capital, we have to estimate a rate of return that serves as an opportunity cost. We can then proceed, assuming we have estimated an expected profit, by estimating our return on capital. In contrast, to determine the *ex post* value of capital, we *could* use the current rate of return as discount rate (as immediate opportunity cost) or we can use a future expected rate of return as opportunity cost. This forward-looking perspective is often ignored in the course of the Cambridge controversies, with both sides to blame[102]. In other words, by introducing the concepts of *ex ante* and *ex post* the sudden paradox and circularity disappears. Even if it were admissible to assume capital, in Solow's sense, as a single capital good with its corresponding physical output (instead of a multitude of different capital goods with varying rates of return), reintroducing the concept of time and the notion of *ex ante* and *ex post* is a clear rebuttal to Sraffa's (1960) and Robinson's (1953) critique of the circularity inherent in the neoclassical production function.

Of course, when we refer to physical outputs and static equilibrium with no *ex ante* and *ex post* or the notion of the entrepreneur, this circular argument becomes problematic. But since this is not the case, Cambridge, U.K. is mostly attacking a strawman.

(3) The complete omission of time and duration

By both the neoclassical side *and* the neo-Ricardian side, the role of time was completely misunderstood. Previous contributions (e.g., Mises, 1949; Fisher, 1930) were ignored and in vain. Moreover, the "period of production" was completely, especially due to the alleged *reswitching* problem which we will discuss below.

4.2.3 A Brief Note on Fischer Black's Reinterpretation of the Neoclassical Production Function

Black (1990) was trying to formulate his ideas on capital markets to his colleague by using "their language." Their language effectively meant through the Cobb-Douglas production

[102] The Cambridge, U.K. side (e.g., Pasinetti, 1966) could be accused of an excessive obsession with the mathematical side of the debate, while Cambridge, U.S. (e.g., Solow, 1963) could be accused of an excessive fixation on the equilibrium angle of the debate, neglecting the role of time in capital theory and economics. A greater emphasis on the subjective nature of economics could have prevented this type of mathematical and equilibrium tunnel-vision.

function. But what Black (1990) in fact, amounted to nothing short of a devastating critique of his neoclassical colleagues. As Mehrling (2000) explains:

> "For Black, the standard aggregative neoclassical production function is inadequate because it obscures sectoral and temporal detail by attributing current output to current inputs of capital and labor, but he tries anyway to express his views in that framework in order to reach his intended audience. Most important, he accommodates the central idea of mismatch to the production function framework by introducing the idea that the "utilization" of physical capital and the "effort" of human capital can vary over time.
> (…)
> [T]he meaning it expresses remains very far from familiar to the trained economist. For one, the labor input has been replaced by human capital so there is no fixed factor. For another, both physical and human capital are measured at market values, and so are supposed to include technological change. **This means that the A coefficient is not the usual technology shift factor (the familiar "Solow residual") but only a multiplier, indeed a kind of inverse price earnings ratio, that converts the stock of effective composite capital into a flow of composite output**. In effect, and as he recognizes, Black's production function is a reduced form, not a production function at all in the usual sense of a technical relation between inputs and outputs. What Black is after comes clearer when he groups terms and summarizes as $Y_t = AE_t K_t$, where Y is output, E is composite utilization, and K is composite capital. Here the effective capital stock is just a constant multiple of output, and vice versa. It's just an aggregate version of Black's conception of ideal accounting practice (1993c) wherein accountants at the level of the firm seek to report a measure of earnings that can be multiplied by a constant price-earnings ratio to get the value of the firm." (p. 13) [emphasis mine]

Hence, Black (1990) effectively redefined the production function by showing that the capital stock (in value terms) is a *result* of output, rather than a *cause* of output. This would, of course, be sufficient to completely invalidate the production function as the basis of economic growth forecasting, developmental economics and other later (modern) applications of neoclassical growth theory, which are one way or another derived from Solow's initial work.

Inherent to Black's disagreement with his non-finance colleagues, is the difference between an economic versus a finance perspective on the nature of the economy. This is essentially what distinguished Fisher (1930) from his adversaries in the second historical round of debates. "The classical economists habitually thought of the present as determined by the past. In Adam Smith, capital is an accumulation from the careful saving of past generations, and much of modern economics still retains this old idea of the essential scarcity of capital, and of the consequent virtue attached to parsimony. The financial point of view, by contrast, sees the present as determined by the future, or rather by our ideas about the future. Capital is less a thing than an idea about future income flows discounted back to the present, and the quantity of capital can therefore change without prior saving." (Mehrling, 2000, p. 14) Indeed, the backward-looking view of capital can be traced back to classical thinkers such as David Ricardo and observed in modern neo-Ricardian and (at times) neoclassical and Austrian economic thought. In contrast, the forward-looking view of capital can be observed in a handful of Austrian thinkers (e.g., Menger, 1888; Mises, 1949) and neoclassical thinkers (e.g., Fisher, 1930; Black, 1990).

4.3 The Never-Ending Debate on Capital Reversing and Reswitching

According to Stiglitz (1974), even though capital can be aggregated across society, reswitching does not occur as such in the aggregate, but rather in an underlying project or marginal capital investment and has therefore no important ramifications for an economy. However, the reswitching critique is completely valid as soon as we try to construct an aggregate production function and imagine the aggregate to be a single stream of future (net) cash flows discounted to the present by one single rate of interest. Rather, this aggregate implies millions and millions of different streams of future cash flows which can even be discounted to the present at various rates of interest. The economic significance of the reswitching debate is therefore negligible. As long as we do not fall in the neoclassical trap of attributing an independent life to a capital aggregate[103], there is no reason why the reswitching critique is in any way different from making a mountain out of a molehill. Indeed, this was Levhari's (1965) argument: "[Reswitching] may indeed be observed in the production of a single good (…) [yet] it is impossible with the whole basis of production." (p. 99). Yet Pasinetti (1966) would later refute Levhari's work (Levhari was a student of Samuelson, while Pasinetti was a student of Sraffa) by proving that reswitching *was* a possibility, even at the aggregate level.

The phenomenon of reswitching (or capital-reversing) was first noticed by Robinson (1953) and Sraffa (1960). However, Levhari (1965) and Pasinetti (1966) were among the most vocal participants with regard to this critique on Solow (1956) and company (Harcourt, 1972).

Let us go through an example of reswitching. We, as Fisher (1907) did, will assume two income streams:

1. Income stream #1: 5 dollars in 10 years and 100 dollars in 100 years
2. Income stream #2: 15 dollars in 25 years

As Osborne (2014) shows, in a classic discounted cash flow (DCF) valuation this would equal:

---

[103] As Schumpeter (1908) urges, every aggregate or macroeconomic phenomenon must be able to be traced back to the acting individual. This "methodological individualism" is the hallmark of modern-day economics. In fact, the Lucas critique (Lucas, Econometric Policy Evaluation, 1976) evolves around the very fact that aggregates cannot be used in econometric forecasting models, since the parameters of such models are not structural, but rather a result of individual conditions and adaptations to the "rules of the game." Any macro-economic forecasting should be grounded, therefore, in parameters that govern *individual* behavior. We cannot predict the impact of a tax reform on economic growth with historical macroeconomic aggregate data, since individuals might not share the same "microfoundations" as in the past: for instance, perhaps individuals expect the tax reform to be overthrown by the next administration or perhaps they are bound by other variables such as rising housing costs. Models on individual behavior should therefore underlie models on aggregate behavior. Unfortunately, in modern (mostly neoclassical) economic forecasting models (also called dynamic stochastic general equilibrium or DSGE models) that prevailed in the wake of the Lucas critique, it is still common to observe an aggregate production function.

1. $PV_1 = \frac{5}{(1+r)^{10}} + \frac{100}{(1+r)^{100}}$
2. $PV_2 = \frac{15}{(1+r)^{25}}$

Where:
PV = Present value
r = rate of interest

Whether we prefer PV_1 over PV_2, depends on the rate of interest. Yet unfortunately, the traditional discounted cash flow method yields the reswitching phenomenon. That is, PV_1 is higher at a low interest rate (lower than 4%, lower at a medium interest rate (4% to 7%), but higher than PV_1 again at a high interest rate (higher than 7%).

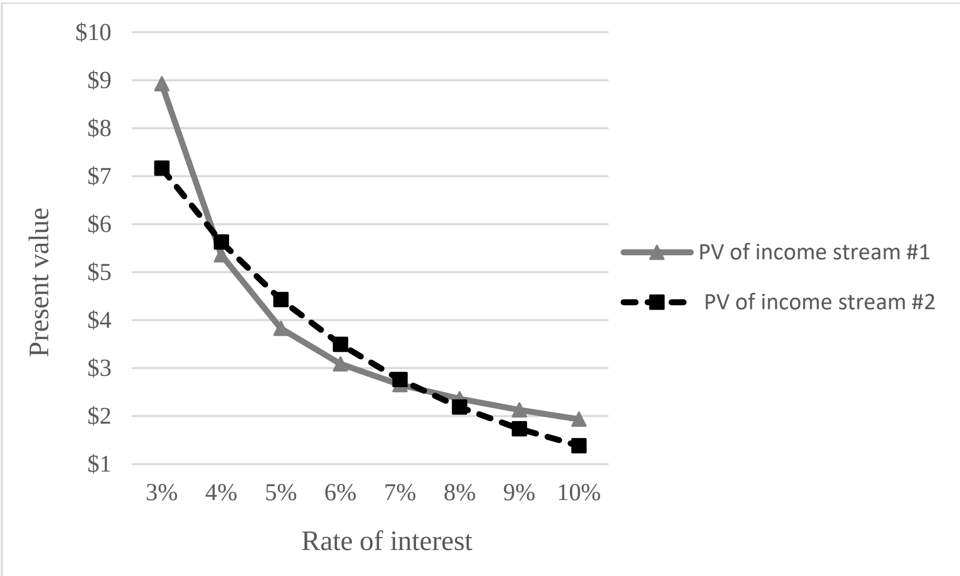


Figure 27: Fisher's (1907) example of "reswitching"; apparently, there is something wrong. Even though PV_1 is supposed to be more "roundabout", to have a longer "duration" (t=100) and to have a longer (average) "period of production," it becomes more profitable again at higher rates of interest than PV_2. This inconsistency allegedly refutes Austrian and neoclassical capital theory.

Alternatively, Sraffa (1960) used an example in which two production techniques are compared (with varying labor inputs at different points in time), but output is assumed equal. At what interest rate is production technique A more profitable than B and vice versa? Which production technique should be picked? The foundation of both neoclassical and Austrian economics would assert, according to Sraffa (1960), that at some higher rate of interest one technique is more cost-efficient, whereas at some lower rate of interest the other technique is more economical. Curiously, Samuelson (1966), who is or was considered to be on the neoclassical side of the debate, admitted that the issue of reswitching runs contrary to neoclassical *and* Austrian thought:

> "The phenomenon of switching back at a very low interest rate to a set of techniques that had seemed viable only at a very high interest rate involves more than esoteric technicalities. **It shows that the simple tale told by Jevons, Böhm-Bawerk, Wicksell, and other neoclassical writers** — alleging that, as the interest rate falls in consequence of abstention from present consumption in favor of future, technology must become in some sense more "roundabout," more "mechanized," and "more productive" — **cannot be universally valid**." (Samuelson, A Summing Up, 1966, p. 568) [emphasis mine]

As Merlo (2016) notes, "[reswitching] would imply that the investment's profitability might be enhanced by increasing the cost of capital" (p. 43), which would run counter to the idea of profit maximization which is inherent to neoclassical theory but also to Austrian theory (albeit not in an equilibrium sense but in a tendency toward equilibrium, e.g., Mises, 1949). According to Kirzner (1996), "(…) it was the reswitching and capital reversal paradoxes which cast the deepest shadow upon the mainstream neoclassical theory of distribution" (p. 7). Nevertheless, according to Kirzner (1996), "(…) these paradoxes present no problem at all for the explanation of interest contained in the Misesian theory of capital and interest" (ibid.). While for the moment we might doubt the latter, we will later see that the reswitching paradox indeed was no problem for the original Misesian theory of capital and interest.

In Samuelson's 1966-article, he takes aim at, on one hand, Böhm-Bawerk and Hayek and, on the other hand, his neoclassical predecessors (e.g., Fisher, 1930). As Samuelson (1966) comments, the mathematical case of reswitching shows that by reducing present consumption and increasing capital formation, interest rates can actually rise (and, hence, the time preference theory of interest is refuted):

> "[I]n the conventional model, successive sacrifices of consumption and accumulations of capital goods lead to lower and lower interest rates. This conventional neoclassical version of diminishing returns is spelled out at length in my *Economics*. Unfortunately, until reswitching had alerted me to the complexity of the process, I had not realized that the conventional account represents only one of two possible

outcomes. (…) [The] story can be reversed: **after sacrificing present consumption and accumulating capital goods, the new steady-state equilibrium can represent a rise in interest rate**!" (ibid., p. 579) [emphasis mine]

Reswitching would, thus, break with this foundational assumption in economics that businesses and entrepreneurs are able to maximize profits by picking the least expensive (or most economical) production method *and* with the time preference theory of interest. Take the following example of producing a single good *X* with either technique A or B (with different labor inputs in different periods), which was presented by Samuelson (1966):

Period	**Labor input**	
	Technique A	*Technique B*
-3	0	2
-2	7	0
-1	0	6

Figure 28: Two production techniques, A and B, with different labor inputs in different periods. Which production technique is more economical at what rate of interest?

There are slight differences with Fisher's 1907-example: in this case, we are not *discounting* (future cash flows), but *compounding* (past labor inputs). The general principle, however, remains the same. At various rates of interest, the present value (or capital cost) *switches back* from one technique to another.

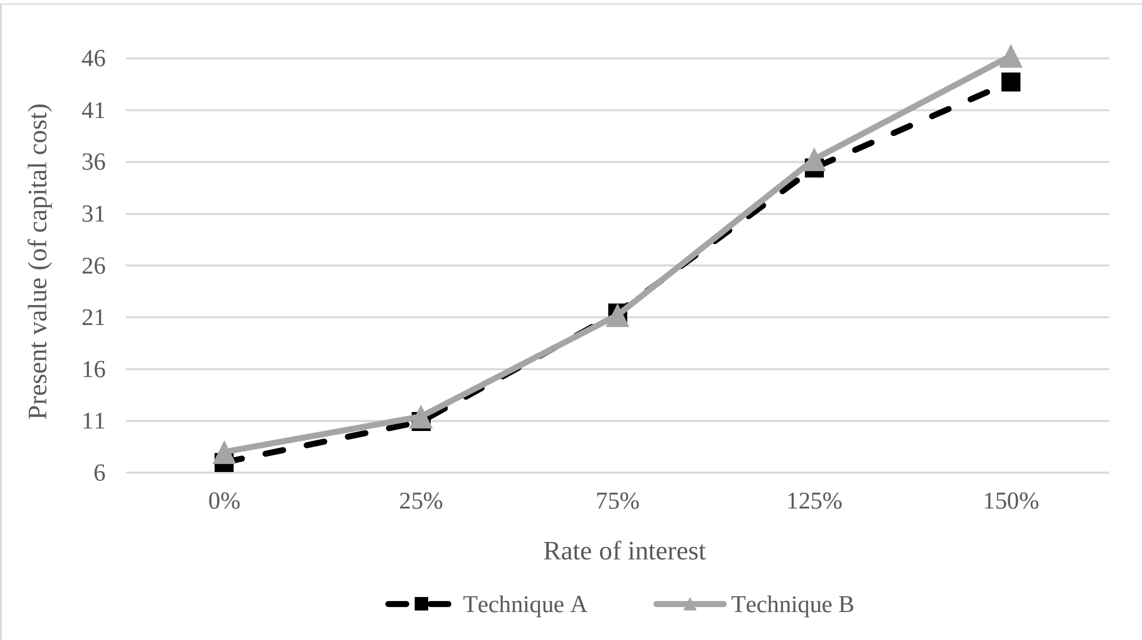


Figure 29: Samuelson's (1966) example of "reswitching"; this time around, we are looking at the "cost" of a capital good in terms of its labor inputs. Technique A and B yield the same single unit of a given capital good. When we compound past labor inputs at different rates of interest, one technique results in a lower cost than the other. However, there is "reswitching": at low interest rates, the cost of technique A is lower, at intermediate interest rates, the cost of technique B is lower, but the cost of technique A becomes lower again at high interest rates, hence, reswitching.

Note, first of all, that one of the greatest troubles with both the Cambridge, U.S. *and* the Cambridge, U.K. side concerns the very fact that labor is assumed to a homogenous input which is equal across different methods of production. This assumption seems to be a stretch. Knowledge that can or could be applied (or must be applied) in production process A, is in many cases different to the knowledge that is applied in production process B. You can get a massage from either a Shiatsu or a Swedish massage therapist which might alleviate back pain, but a Shiatsu therapist is usually unable to replace a Swedish massage therapist and vice versa.

Second, note the backward-looking perspective in Sraffa's and Robinson's work (and, regrettably, Samuelson's interpretation of Hayek's and Böhm-Bawerk's work on capital and interest). Output is a result of historical labor inputs. A capital good is, too, a result of accumulated labor inputs. Yet what gives value to a given capital good are not the historical outlays for its creation, but rather the future income that this capital good is able to generate.

Now, if we take the above labor inputs (as dollar amounts) and derive their future value (which, then, would indicate which production technique is cheaper at a given rate of interest), we can produce the following table:

Rate of interest	**Technique A**	**Technique B**
150%	*$43.75*	$46.25
75%	$21.44	*$21.22*
25%	*$10.94*	$11.41
0%	*$7.00*	$8.00

*Figure 30: Both techniques A and B for the production of a given capital good with their past labor inputs taken (compounded) to the present (*t *= 0) as in Samuelson's (1966) example. The more economical technique at a given interest rate is highlighted in italics. Reswitching between techniques*

occurs: at 0% and 25%, technique A is more efficient, at 75% technique B, and at 150% technique A becomes more efficient again (reswitching).

The conclusion of such theoretical exercises is that the assumption that a market economy is unable to allocate resources in line with the preferences of consumers. Moreover, it allegedly refutes the idea of a rational law of supply and demand. As Hausmann (1981) aptly summarizes the Cambridge, U.K. critique: “If capital is an input into production whose value measures its quantity and whose marginal product decreases with its quantity, and if the rate of interest is proportional to the marginal product of capital, then the rate of interest and the value of capital must be inversely related. How can one regard capital as a factor of production if (…) firms find it profitable to use less capital when its price (the rate of interest) declines? When an input becomes cheaper one should expect firms to find it profitable to use relatively more of it. Something is drastically awry” (p. 76).

Nevertheless, such hypothetical mathematical examples disregard the fact that not *one* single interest rate exists, but rather a wide variety across different maturities. In other words, in such hypothetical examples the existence of the *term structure of interest rates* (or *yield curve*) is not taken into account.

Now, let us — by way of contrast — assume the same cash flows as Fisher’s initial example. However, instead of discounting at a single rate of interest, we discount the cash flows according to a sloped yield *curve*, or *term structure of interest rates*. That is, instead of discounting at a single rate of interest, say 3%, we discount the first ten years at 3%, the next fifteen years at 3.5% and the longest maturity (100 years) at 4%. We assume a conservative positively sloping yield curve spread here. For instance, the average term spread between ten-year rates and thirty-year rates does not generally amount to fifty basis points, but is closer to approximately a hundred basis points[104]. In contrast to Fisher’s (1907) and Sraffa’s (1960) findings, this yields a consistent yield curve with no reswitching (only a *single* switch):

[104] As of writing, the 30-10 Year US Treasury yield spread is 30 basis points. However, less than a year ago (2017) the spread amounted to 65 basis points. Yet even that spread is historically low. In 2011, for example, the spread stood at 150 basis points.

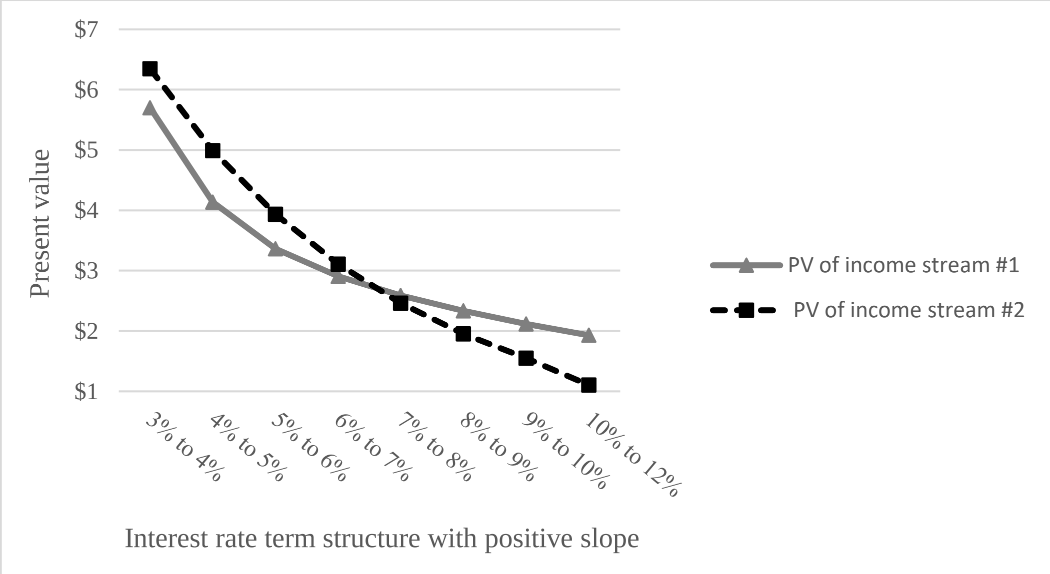


Figure 31: The same example as Fisher (1907), but this time with a yield curve spread applied to the cash flows (that is, the cash flows are discounted at different rates according to a positively sloped yield curve or term structure). After taking into account the term structure, no reswitching occurs.

Hence, even though reswitching cannot be completely ruled out by integrating the term structure of interest rates into both Fisher's as Sraffa's reswitching examples, is surely diminishes and avoids its occurrence to a very large extent. What was a phenomenon that would occur at extremes (extreme as in the size of the cash flows, the timing and duration of the cash flows and the number of periods), just became a phenomenon that is even less likely to occur or that would only occur at even *extremer* extremes. Given a more or less regular or usual yield curve (combined with avoiding sign changes by separating negative and positive cash flows), reswitching tends to disappear in both theory and practice.

Similarly, Tiwari (1994) makes a case for using multiple single-period rates as discount factor rather than one single rate (or some average of multiple single-period rates), albeit for other reasons. According to Tiwari (1994), the rate of interest changes over time and this uncertainty should, one way or another, be taken into account net cash flows from more distant periods. Hence, even though Tiwari (1994) takes another angle, it is clear that using a single rate of interest to discount *any* future cash flow at any period in time tends to lead to erroneous capital budgeting decisions.

Now let us do the same with Sraffa's example. Let us assume the same two production techniques, A and B, with the same labor inputs spread across three periods. Since this example requires more "extreme" rates of interest to observe reswitching (0, 75 and 150 percent), I have taken the present spread between 1-year, 2-year and 3-year US Treasuries. Then, I applied the relative spread to the 75 and 150 percent "interest rates" of the initial Sraffa-example. I assume that, in case of for instance the 75 percent-example, the 1-year rate equals 75% and adjust the 2-year and 3-year rates accordingly with the current percental *observed* spreads I mentioned earlier (this would yield a 1-year rate of interest of 75%, a 2-year rate of interest of 87% and a 3-year rate of interest of 93%). I repeated the same procedure for the 150-percent case. The result is as follows:

Term structure			**Technique A**	**Technique B**
1 yr	**2 yr**	**3 yr**		
150%	174%	186%	*$52.55*	$61.79
75%	87%	93%	*$24.48*	$24.88
0%	0%	0%	*$7.00*	$8.00

Figure 32: The more economical technique at a given interest rate is in italics. After taking into account a positively sloping yield curve, no reswitching occurs. In fact, no switching *occurs: technique A remains the most optimal (economical) choice under all considered interest rates.*

In the above example, by including the term structure of interest rates (instead of one single interest rate), reswitching disappears: *technique A becomes the more economical (lower cost) technique to produce the capital good at all interest rates*. Visually, this can be seen with the following chart:

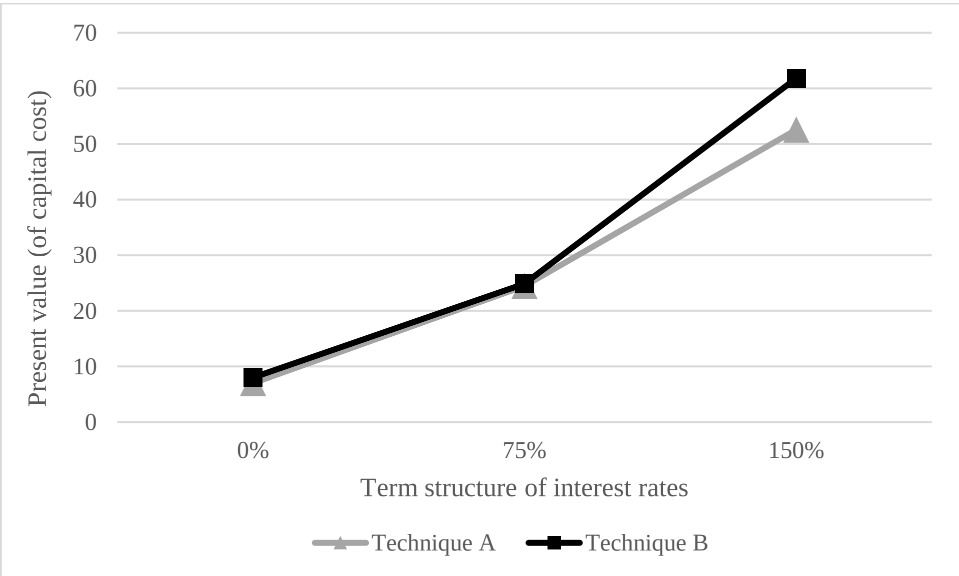


Figure 33: Instead of using one single rate of interest, we took the term spread on 1-year, 2-year and 3-year US Treasuries and applied these to the interest rates used in Samuelson's (1966) example. By using the term structure, reswitching disappears: technique A is less costly at all (term structures of) interest rates.

Yet even in showing the flaw of discounting cash flows (or labor inputs) of different periods at one single rate of interest, we should make an explicit acknowledgment at this point. In contrast to what many economists appear to suggest (e.g., Sraffa, 1960; Samuelson, 1966), the "Austrian" theory is not based on the notion that time preference suggests that long-term interest rates should *always* exceed medium- and short-term interest rates. The Austrian theory only suggests that, with regard to their non-present (postponed) consumption, individuals have preferences: for instance, X amount (for a given consumption) in period t = 1 and Y amount in period t = 10. A (segmented or preferred habitat[105]) term structure is an expression of this rather Austrian idea of time preference. Hence, *reswitching* might occur, albeit not in the sense that Sraffa and others would have imagined. Since savers could decide to massively save at five years, the rate of interest on 5-year debts could be lower than the equivalent on 10-year debts[106]. Even though there is *a degree* of arbitrage, it will never be perfect, since investments with different maturities are not perfect substitutes. On the margin, savers will prefer other maturities (shorter and longer) and borrowers will likewise prefer other maturities (shorter and longer). Yet this imperfect arbitrage does never result in a perfectly arbitraged yield curve. One can attempt to build a vegan hamburger by replacing beef with soy and mushrooms, but not all beef hamburger eaters will substitute their preference for a, to them, inferior vegan hamburger: there exists imperfect arbitrage between vegan hamburgers and real hamburgers, just as there exists imperfect arbitrage between different maturities on capital markets. Although the idea of a 30-year bond trading at a discount to a 10-year bond might sound outrageous, this is what actually can be observed in

---

[105] Below we will elaborate further on the importance of the term structure, see p. 261.

[106] This also proves, as our discussion below will point out (p. 261), that the *expectations* theory of the term structure is false. The *expectations* theory of the term structure posits that long-term (real) rates are a product of *expected* future short-term (real) rates. Hence, according to this theory, a 30-year bond trading at a discount against a 10-year bond would be completely irrational. The liquidity premium theory of term structure suffers from the same defect: it cannot provide a logical explanation for inversions.

real markets, both in the past and present. Longer maturities have often traded at a discount to shorter maturities (e.g., Homer & Sylla, 1996).

We have just seen, in a simple example, how by applying a term structure of interest rates the phenomenon of *reswitching* largely disappears. However, there is a similar, important, contribution by Osborne (2014) which goes in the same direction. As we have seen in Fisher's example (Fisher, 1907), we are comparing two income streams: the first gives us $15 in 25 years, the second gives us $5 in 10 years and $100 in 100 years. Let us assume that we can buy either stream for $5 dollars and that the rate of interest is 4%. The present value of the first stream is $5.36 (which would give us a *net* present value of $0.36), whereas the present value of the second stream is $5.63 (or a *net* present value of $0.63). Apparently, you would thus pick stream 2 over stream 1 since you are considered to maximize your return on investment. The only problem with this overly simplified (mostly unmentioned) assumption in Fisher's streams, is that the second stream, even though it NPV appears higher, takes hundred years before it materializes, whereas the first stream would take *just* 25 years to materialize.

One way in which practitioners take this factor into account in their capital budgeting decisions (which are exactly the decisions we are putting under scrutiny in the debate on reswitching), is by using other measures of return on capital investments. It is unlikely that entrepreneurs depend their capital investment decisions on a mere NPV calculus. In fact, entrepreneurs often use the payback period next to a discounted cash-flow valuation (NPV). The payback period (and its variations) method indicates the risk and liquidity of a project (Brigham & Ehrhardt, 2008). Considering the NPV method in isolation, generally ignores the *illiquidity* (or *degree of liquidity*) of a project. Given Fisher's extreme example, where income stream 2 only fully materializes after 100 years, it seems unlikely "reswitching" would actually occur in practice as practitioners take into account other measures of profitability.

We should therefore take into account that NPV, as other measures of return on capital investments, are *tools* that practitioners use and academics can observe and analyze, but are certainly not *defining* for the underlying economic cause-and-effect relationships. Even without NPV, entrepreneurs would have cues to deduce whether short-term or long-term investments are less or more profitable at any given state of the economy. In this sense, the critique that the NPV under certain assumptions and without due consideration of the mathematical impact of polynomials can result in contradictory results, is akin to criticizing double-entry bookkeeping (which allows us to determine retroactively whether a profit or loss was made) because of changes in the purchasing power of the unit of measure (Mises, 1949). Economics should be concerned with real choice and real decision-making, not with hypothetical choice and mathematical anomalies given a multitude of theoretical "if's" and "but's"[107].

Macaulay *duration*, which we discussed earlier, is perhaps a better way to account for the "illiquidity" of any given future income stream and the role of time in the NPV-approach to capital budgeting.

Osborne (2004; 2012) discovered that by including the unorthodox rates to the orthodox rate (which, equals, paradoxically the *weighted* time-to-maturity or correct Macaulay *duration*) the phenomenon of reswitching also disappears. Since the reswitching is caused by the existence of polynomials, the unorthodox rates are normally ignored in solving a net present value formula. As Osborne & Davidson (2016) argue: "The calculation and analysis of interest rates in discrete time have to date failed to use the full properties of the underlying polynomials on which they are based." (p. 1). With regard to the specific Sraffa-example (of labor inputs), Osborne & Davidson (2016) write:

> "For some, reswitching undermines the foundations of neoclassical economics because it belies the idea of a monotonic relationship between relative capital values and factor price. The reswitching equation is an *n*th degree polynomial having *n* roots, implying the existence of *n* interest rates. Conventional analysis uses one interest rate but ignores the others. We argue that the others should not be ignored because all rates are determined simultaneously, and when one rate shifts, all rates shift. We demonstrate that the Samuelson reswitching model possesses a 'dual' expression containing every interest rate, the rates being compressed into a composite, interest-rate variable, thereby establishing a role for interest rates previously thought lacking

[107] Aswath Damodaran, a highly praised NYU finance professor, when asked about reswitching commented: "Reswitching is a math problem, not a finance problem. Let me put it this way: in the list of sins that bedevil investing in companies, this does not make the top hundred."

in use and meaning. **The relationship between this composite interest rate and capital value does not exhibit reswitching.**" (p. 1)

Curiously, the approach advocated by Osborne (2004), is equal to an adjusted Macaulay *duration*. That is, the product of the *unorthodox* interest rates represents the adjusted Macaulay duration of the original cash flows (or the original labor inputs as is the case in the above Sraffa-example). Here we can see how earlier theories previously deemed unrelated to capital theory suddenly becomes immensely valuable (Macaulay, 1928). Let us apply Osborne's (2004) method to see what happens to the labor-input example:

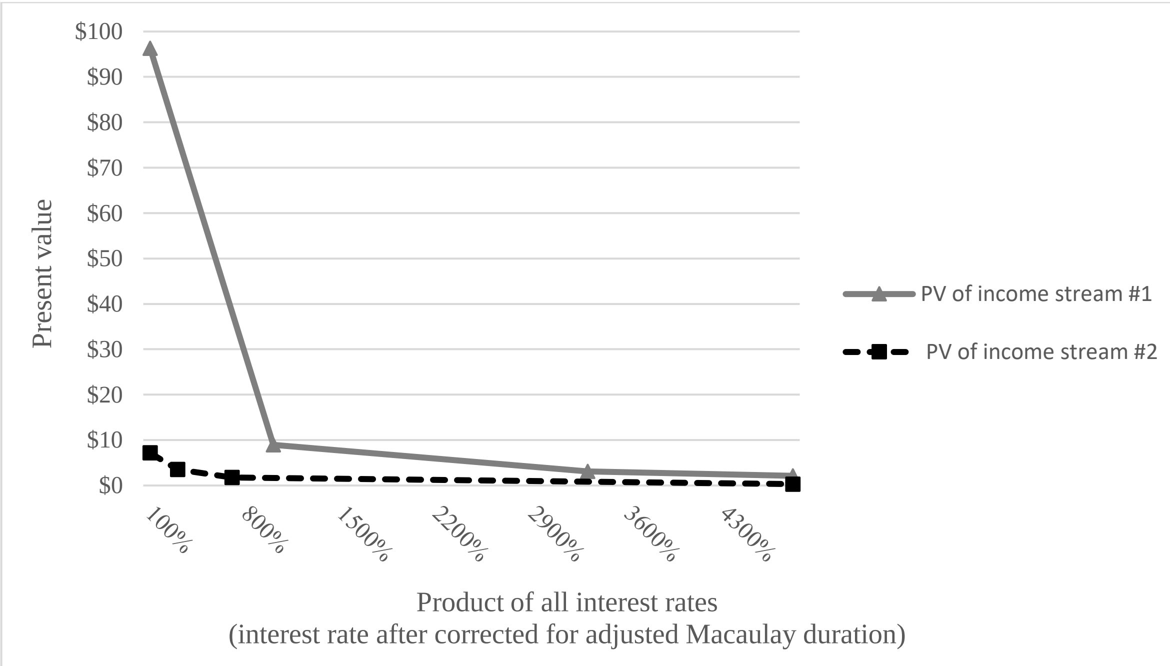


Figure 34: The reswitching phenomenon disappears in Fisher's (1930) example, as well, when we apply the multiple-interest rate approach advocated by Osborne & Davidson (2016).

As we can appreciate above, the problem of reswitching disappears as soon as we include the *unorthodox* rates. Let us also apply the same method to the Fisher (1930) example, which is about cash flows rather than labor-inputs and involves discounting the future rather than calculating the present value of past labor inputs (which is much closer to our notion of capital):

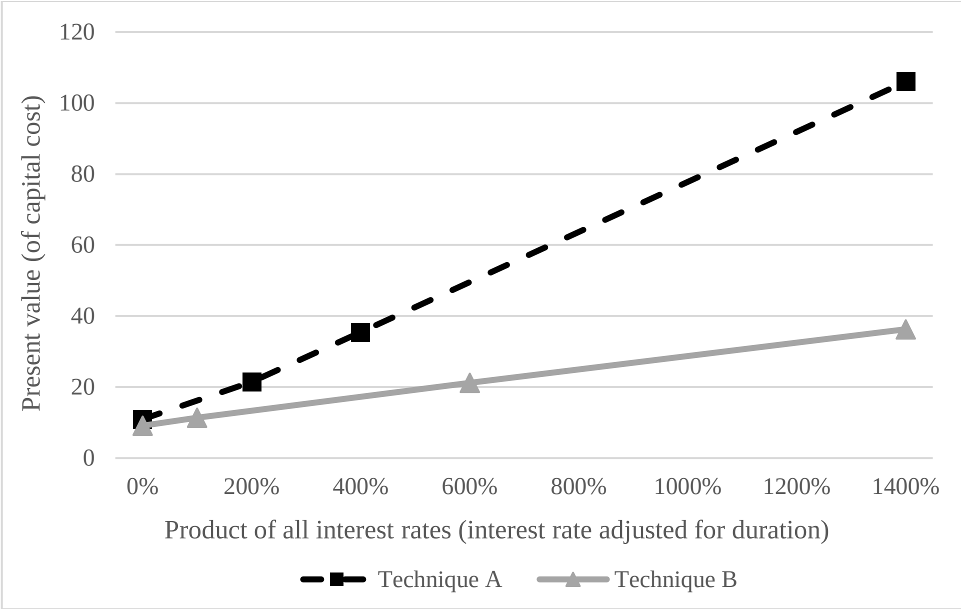


Figure 35: The novel approach by Osborne (2004) also resolves the Sraffa-paradox of past labor inputs. By including the unorthodox rates that are normally ignored (yet include valuable information according to Osborne), we get a time-weighted rate of interest that solves the reswitching phenomenon.

Moreover, in practice, entrepreneurs separate the present value calculations of the outlays and inflows of a certain projects or good, which at least avoids sign changes which reduces the frequency of (potential) reswitching. This reduces the odds of anomalies in NPV models considerably.

Moreover, Brigham and Ehrhardt (2004) inconsistent rankings are most of the times the result of investment projects with “large cash outflows either sometime during or at the end of its life” (p. 1102). This is not often the case in real life capital allocation decisions, as evidenced by the fact that most investment require an initial (large) outlay rather than a future large, sudden outlay.

Yet, even if all the quibbles about reswitching *would* have merit, it is of no importance that entrepreneurs pick a more “profitable” project over a less “profitable” projects from their individual perspectives[108]. Rather, an individual should be able to compare his or her

[108] We do not pretend here that, especially large businesses, decide on capital expenditures based on a set of different options compared along different profitability measures, such as traditional NPV or IRR methods.

proposed project and whether it generates a yield compared to its opportunity cost. If so, he or she will embark upon the project and the project should be deemed profitable. Afterwards, or at intervals, the individual is able to retrospectively determine whether his investment was successful. If it is profitable, they have bid up some prices (so-called "inputs") and lowered other prices (so-called "outputs"). Successful arbitrage of prices will show a profit. And a profit, any profit for that matter, will bring us closer to a hypothetical state of equilibrium where no further arbitrage is profitable, regardless of whether the entrepreneur picked a *more* profitable investment over a less profitable one.

As Mises (1949) explains, "the custom of computing interest pro anno is merely commercial usage and a convenient rule of reckoning. (…) It does not affect the result whether one calculates with an unchanging rate of interest and changing prices of the principal (…) (p. 533). In other words, interest could easily be expressed in a quantity of money instead of a percentage rate. That is, we could easily state the logic behind interest to be "I prefer $100 dollars today over $105 in a year, but $106 in a year over $100 dollars today." This would allow us to take the nominal cash flows of an asset (or combination of assets) and subtract the time value of money as a quantity rather than discounting nominal flows at a percentage rate. This would avoid any reswitching or reversals. Macaulay (1938), in this context, remarks that: (…) sometimes seems easier to think in terms of *price* than in terms of *yield*. Why should we not substitute price for yield in all our discussions? Probably the simplest way to answer this question is to point out that 'yield' may often be a better way to measure price than prices themselves. It measures a corrected rather than a raw price. **It may be considered as the reciprocal of an adjusted price**—a price that has been corrected for varying coupon rates and maturities." (p. 41) [emphasis mine]

4.4 The Inadmissible Exclusion of the Time Preference Explanation of Interest in the Cambridge Debates

Kirzner (1996) notes that the neoclassical explanation of interest as a fruit of the physical productivity of capital "was the *only* view advanced as an alternative to the neo-Ricardian perspective on interest as a surplus" (p. 135). As Kirzner (1996) adds, Robinson's and Sraffa's neo-Ricardian did not appeal to market exchange relations, but to "the relations between workers and capitalists including possibly their relative bargaining power" (Hausman, 1981, p. 167) Interest is thus a surplus that arises from capitalists exploiting workers.

As Stiglitz (1974) highlights, a part of the controversy revolved around the phenomenon of interest and its relation to savings. Cambridge, U.K., explained the rate of interest as a simple function of economic growth: savings are, generally, retained earnings, says Cambridge, U.K., therefore savings are a share of profits and thus depend on economic growth or, in other words, profit growth (Stiglitz, 1974). The rate of interest, then, equals the marginal productivity of capital. Nevertheless, the controversy revolves around the fact that Cambridge, U.K. sees the causality starting at the savings rate, which determines the marginal productivity of capital. Cambridge, U.S., however, sees the marginal productivity of capital as the causal factor, which then determines the rate of interest. Stiglitz (1974) limits his critique to blaming Cambridge, U.K. for being unable to present a theory of productivity and a theory of imputed prices of both labor and capital goods. According to Stiglitz (1974), Cambridge, U.K. is unable to explain the prices of individual factors of production: they only establish (since the rate of interest is equal to the marginal productivity of capital) that the *average* rate of profit should be equal to the average rate of interest, without explaining how individual wage rates and profit rates are determined. Nevertheless, one wrong does not right another wrong. Cambridge, U.K., was completely right to assume savings as a causal factor of the rate of interest, as the theory of time preferences proves (e.g., Fisher, 1930; Mises, 1949). Whether or not the distinction between household savings and retained earnings matter, is less clear. Whenever a firm retains earnings ("saves"), it does not return money back to its current shareholders. Since shareholders are *ultimately* households (ultimately a financial intermediary might hold the stock, but its liabilities are ultimately held by households as part of their portfolio). Hence, the distinction between retained earnings and household savings is one of degree, not of kind. What Cambridge, U.S. fails to explain, is why profit rates are not arbitraged away by profit-eager entrepreneurs until they cease to exist. Hence, despite the fact that Cambridge, U.K. lacked a theory that explains the price of individual capital goods, they were right in criticizing Cambridge, U.S. for their theory of savings and interest, since it clearly lacks a causal link between savings and the rate of profit.

None of the two Cambridge's even considered the rate of interest to be determined by, on one hand, time preferences (that is, abstinence) and, on the other hand, "investment opportunities" (Fisher, 1930). They considered the rate of interest as an equal of the rate of profit, a fallacious concept that we have already extensively discussed over the first two rounds of controversies on capital.

4.5 The Measurement of *K*

Despite several valid theoretical objections against measuring or estimating the capital stock (K) as one of the inputs to growth models based on the Cobb-Douglas production function, many authors (e.g., Solow, 1956; Swan, 1956) went ahead and used neoclassical growth models for pragmatic reasons.

In fact, K is operationalized as the sum of past investments or the "physical capital stock," depreciated at a fixed annual rate[109]. Moreover, K is subject to the law of diminishing returns. This implies that at a certain point the rate of depreciation of the capital stock exceeds output and therefore, per definition, investment. This means that the model logically implies a point at which we inevitably end up with *negative* net investment. As a consequence, the measurement of K implies that sustained economic growth is impossible and virtually assumes that we are heading for a complete disaster, as output and capital accumulation will grind to a halt. Assuming a fixed rate of investment to consumption (as shares of output) even implies a complete disinvestment of the entire capital stock, which is nothing short of absurd.

Robinson (1953) was very critical of attempts to "quantify" capital, explicitly rejecting the new present value approach and, moreover, the credibility of capital markets:

> "[C]apital may be conceived of as consisting either in the cost or in the value of the plant. If cost is the measure, should money cost actually incurred be reckoned? It is only of historical interest, for the purchasing power of money has since changed. Is the money cost to be deflated? Then by what index? Or is capital to be measured at current replacement cost? The situation may be such that no one in his senses would build a plant like this one if he were to build now. Replacement cost may be purely academic. But even if the plant is, in fact, due to be replaced by a replica of itself at some future date, we still have to ask what proportion of the value of a brand-new plant is represented by this elderly plant? And the answer to that question involves future earnings, not cost alone.

[109] In many applications of the Solow-Swan growth model, for instance, a fixed 10% per annum depreciation rate is assumed.

If the capital is to be measured by value, how decide what the present value of the plant is? The price at which it could be sold as an integral whole has not much significance, as the market for such transactions is narrow. To take its price on the Stock Exchange (if it is quoted) is to go before a tribunal whose credentials are dubious. If the capital-measurer makes his own judgment, he takes what he regards as likely to be the future earnings of the plant and discounts them at what he regards as the right rate of interest for the purpose, thus triumphantly showing that the most probable rate of profit on the capital invested in the plant is equal to the most appropriate rate of interest.

All these puzzles arise because there is a gap in time between investing money capital and receiving money profits, and in that gap events may occur which alter the value of money." (p. 84).

Hence, Robinson's (1953) critique revolves around three elements: (a) using past prices of capital goods (cost of purchase) assumes we are able to account for inflation, but adjusting for inflation is arbitrary; (b) using current replacement cost would involve replacing an apple with a peer, since we would never replace one single capital good with the exact same capital good; (c) the present value approach is fraught since the market might be illiquid so other market participants will not inform on the present value; (d) financial markets, as such, are completely unreliable indicators of present value; (e) if the person in question produces his own valuation, and given the fact that the rate of interest is *equal* to the rate of profit, he should always get a net present value of exactly zero, which is uninformative.

Robinson (1953) makes several valid points (her first point is, for example, completely valid), especially in relation to for instance Solow (1956), yet she also makes some contradictory claims: with regard to (b), what matters in economics is not the *physical* replacement, but the replacement in *value* terms. Concerning (c), this is irrelevant, since entrepreneurs appraise their own assets (in this case, a plant or a prospective soon-to-be-built plant) against market prices, hence, making it more or less irrelevant if the underlying markets are liquid or not liquid. With regard to (d), this simply seems a completely arbitrary and ideologically-driven value judgment (as we will see later with our equity q, there is a reversion to the mean and, thus, at least equity markets are sufficiently 'efficient' in the long run). Regarding (e), *even* if you accept that the rate of interest is equal to the rate of profit, there is no true contradiction.

Since the entrepreneur might value the plant according to his own (future) cash flow estimates, as long as he uses as the rate of discount the *market* rate of interest rather than his own (opportunity) rate of profit, then there is no contradiction. Capital is, in a first instance, a theory of entrepreneurial appraisal, not some objective top-down measure of wealth or a material input to some physical production process.

On the other hand, we have the Cambridge, U.K. economists who deny the existence of “human” capital. As we have seen earlier, Knight (1935) argued that human capital is just as much capital as nonhuman capital. Hence, wages and labor cannot be separated from profits and capital. A person *can* be considered capital (and can either enhance or destroy that capital) and his income is a return on capital. However, since only labor services are sold on the market, and not the asset responsible for such services (that is, the human being – in absence of slavery), prices cannot be bid up to reflect the present value of a person’s expected future yields. But the same principles still apply. Moreover, just as other factors of production, labor is not homogenous. Therefore, while some labor income (that is, wages) might rise, others might fall. It is a fallacy to assume labor as some sort of homogenous and fungible input to production. The theory of capital reflects, in a certain sense, a theory of *value* production, not a theory of *material* production.

Interestingly, despite what many economists seem to grasp, the relationship between interest rates and the value of capital is nonlinear. When low interest rates decrease even further, the present value that equals capital goes up further than when a high interest rates decreases by the same number of basis points: hence, potential distorting interest rate signals (i.e., financial intermediaries succeeding to arbitrage down interest rates) lead to a disproportional impairment of economic capital. One of Taleb’s (2012) main critiques is that economists fail to understand nonlinear relationships. This might be one of them. Potential losses and risk of capital consumption increase when low interest rates become even lower.

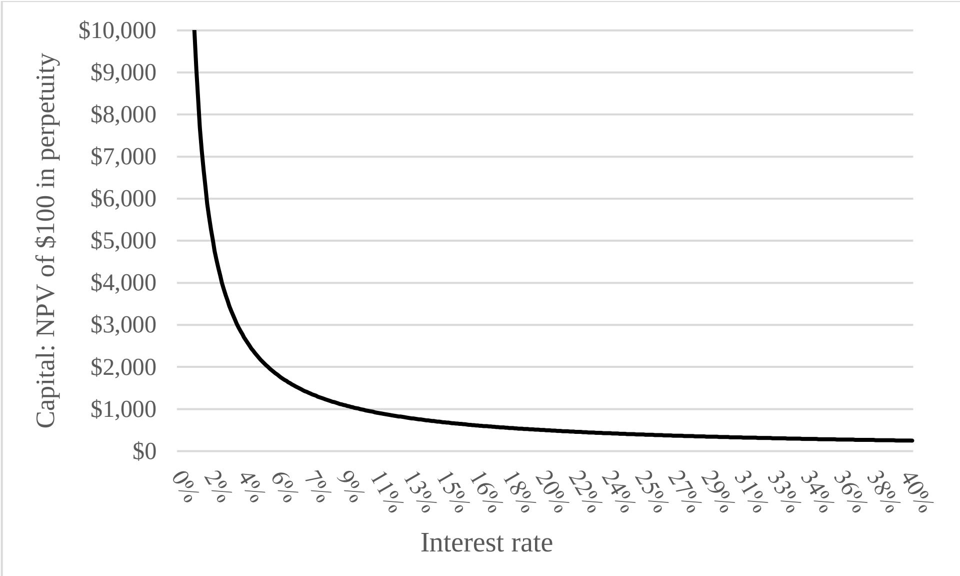


Figure 36: The nonlinear relationship between capital and interest rates

This implies, *ceteris paribus*, that miniscule increases in interest rates from a low base generate disproportionate large losses in capital. Conversely, increases in interest rates from a high base generate relatively mild losses in capital. The relationship between capital and interest rates (time preferences) is nonlinear. This principle does not only apply to capital in a broad sense, but also to stock market prices. Moreover, the same principle applies to financial collateral. Fragility increases when financial valuations are pushed to extremes. This is a concept completely foreign to the Cambridge Controversies of the 1950s and 1960s.

4.6 The Return of the Term Structure

We have seen earlier how early-day explanations of the term structure exclusively relied on an expectations theory (e.g., Lutz, 1930; Böhm-Bawerk, 1888), viewing long-term rates as a mere product of *expected* future short-term rates. This debate became wider and more important, particularly in the 1960s and 1970s. Modern-day defenders of the expectation theory of the term structure are, for instance, Campbell and Shiller (1987), who defend the expectation theory despite empirical evidence to the contrary. They reason any discrepancies between theory and practice away as “behavorial bias.”

Culbertson (1957) was one of the first to disagree with Lutz (1930). Culbertson (1957) explains the yield curve spread through four elements: (1) liquidity differences among different maturities, (2) the attractiveness of debts (investments) of different maturities given the expected future changes in their prices (very much related to the earlier discussed

Macaulay *duration* or price sensitivity to changes in yields), (3) short-term effects of changes in the supply of debt at a given maturity accompanied by a rigid demand for debt at the same, given maturity, and (4) differences in lending costs related to maturity (rolling over debt tends to elevate such costs).

Culbertson's (1957) theory of the term structure was later termed the "market segmentation hypothesis." The market segmentation hypothesis, according to Modigliani and Sutch (1966), suggests "that both lenders and borrowers have definite preferences for instruments of a specific maturity, and for various reasons, partly due to institutional factors and regulations constraining financial intermediaries, will tend to stick to securities of the corresponding maturity, without paying attention to rates of return on other maturities" (p. 183).

Modigliani and Sutch (1966) take Culbertson's "market segmentation hypothesis" and attempt to integrate it with a less naïve expectation theory, which results in their "habitat theory" (later coined the *preferred*-habitat model). This model allows for *some* arbitrage, albeit far from perfect: "Under this model the rate for a given maturity, *n*, could differ from the rate implied by the Pure Expectation Hypothesis by positive or negative "risk premiums," reflecting the extent to which the supply of funds with habitat *n* differs from the aggregate demand for *n* period loans forthcoming at that rate. If the *n* period demand exceeded the funds with *n* period habitat, there would tend to arise a premium in the *n* period maturity, and conversely. Such premiums or discounts would tend to bring about shifts in funds between different maturity markets, both through the "speculation" of investors tempted out of their natural habitat by the lure of higher expected returns and through "arbitrage" by intermediaries induced to "take a position" by borrowing in the maturity range where the expected return is low, and lending where the expected return is high." (p. 184).

Every maturity class is a "habitat", and if it is profitable enough investors will leave their "maturity habitat" to find their luck in different habitats, but only if the reward is exceptionally high. This assures that a single habitat (maturity class) does not move extremely out of line with other maturity classes. Or, according to Vayanos and Vila (2009), "(…) there are investor clienteles with preferences for specific maturities, and the interest rate for a given maturity is influenced by demand and supply shocks local to that maturity" (p. 1). In this Modigliani-Sutch model, the maturities are therefore "segmented." Instead of one intertemporal market, there is a different one for every maturity. Taken to its extreme, every maturity class would be completely isolated from other maturity classes and its corresponding interest rate would *exclusively* depend on the supply and demand at that specific maturity. This would be true in a model with no arbitrage. However, there is arbitrage. First of all, if interest rates in a single maturity class rise relative to others, as Modigliani and Sutch (1966) signaled, there is an incentive not limited to professional arbitrageurs, but rather for savers to pick different maturities and adjust their plans accordingly, as long as the difference is large enough to account for optionality (which has a premium due to uncertainty over future interest rates) and transaction costs.

Still, the "arbitrage" by intermediaries that Modigliani and Sutch (1966) refer to is far from perfect. As we will see in section IV, maturity mismatching by financial intermediaries can lead to ugly consequences. The only way for intermediaries, hence, to arbitrage, would be to adjust their corresponding assets as well to match their new liability maturities. Hence, any "arbitrage" ultimately must come from "capital users," e.g., businesses that require funds for investment. This arbitrage is, therefore, minimal. As Greenwood and Vayanos (2008) describe it: "[W]hile clienteles can substitute to maturities away from their "preferred habitat," **such substitution is imperfect**." (p. 1) [emphasis mine].

Modigliani and Sutch (1966), in effect, identify two "laws" that underlie the term structure:

(1) An individual does not need his income for, let's say, 10 years. If he would invest his savings in a 10-year bond, he would know the expected outcome at the end of ten years. If, he would invest in shorter maturities (let's say 10 times 1 year), his outcome would, on the other hand, be uncertain; his outcome would then depend on the future course of short-term rates at maturity of each 1-year bond. Moreover, he incurs transaction costs for rolling over each 1-year bond. Thus, given risk aversion, he chooses the 10-year bond unless he expects short-term interest rates to be high enough over the 10-year period to compensate for the transaction costs and risk. **In other words, the long-term rate is capped by the option to pick short-term rates; if long-term rates get too high, there is arbitrage, because investors such as this individual will switch to short-term credit**.

(2) On the other hand, this same individual does not invest in maturities longer than 10 years (let's say, in a 30-year bond), since he would be exposed to the uncertainty regarding the price he might obtain for his bond when he attempts to sell it after ten years. **This would cap, or provide a ceiling to, the long-term rate, given the option to pick longer-term rates**.

The implications of this model across the financial chain (from savers to final borrowers) have been mostly overlooked. This model implies that consumers make plans over time regarding their savings (with higher or lesser degrees of "optionality") and adjust their portfolios accordingly. Such "time preferences" are then transmitted across the financial system to the "capital users." The most important distortion in transmitting such time preferences arise out of balance sheet mismatches, principally due to maturity transformation. As such, the primary borrowers and primary lenders which have a specific preference for a given maturity (or set of similar near-identical maturities) and arbitragers have only limited means and incentives to arbitrage the yield curve. In Modigliani's and Sutch's (1966) words: "(…) the spread could also be influenced by the supply of long- and short-term securities by primary borrowers (i.e., by borrowers other than arbitrageurs) relative to the corresponding demand of primary lenders, to an extent reflecting prevailing risk aversion, transaction costs, and facilities for effective arbitrage operations." (p.184).

As Vayanos and Vila (2009) observe, Culbertson's and Modigliani's views have not attracted much attention: "Even though the preferred-habitat view is relevant in practice and has been proposed more than half a century ago, it has not entered into the academic mainstream; it has typically been confined to a short discussion in MBA-level textbooks." (p. 1). Moreover, it has not gained prominence due to its apparent "no-arbitrage" assumption (Vayanos & Vila, 2009).

Yet Culbertson's view, although initially criticized on empirical grounds, has also been largely supported by empirical studies. Greenwood and Vayanos (2008), for instance, show that "a one standard deviation increase in the relative supply of long-term bonds is associated with a 39 bps increase in the term spread and a 2.31 percent increase in long-term bonds' expected excess returns." (p. 2). Moreover, they show that over the business cycle arbitrageurs become resource constraint and more risk averse, leading to a greater degree of "segmentation" of maturity classes.

Indeed, what many miss, is that the whole existence of the term structure implies Taleb's concept of "optionality" (Taleb, 2015). Shorter maturities imply higher optionality, that is, an "option-like situation by which an agent has a convex payoff, that is, has more to gain than to lose from a random variable, (…) that is, can benefit from volatility and dispersions of outcomes." Such an option does not come without a price. Indeed, the price mechanism demonstrates such volatility and dispersions of outcomes, which makes the optionality of shorter maturities valuable. As soon as profit opportunities arise, this optionality can be put to use. We could use the term “liquid”, but this concept has been permeated with the fallacious idea of “exogeneous liquidity.” But the reliance on exogenous liquidity requires, especially in times of crisis, other economic agents to have endogenous liquidity. If the financial sector is structurally illiquid, the degree of optionality can turn out to be a mere illusion in, as long as the music plays apparently liquid markets. Hence, liquidity preference, or degrees of optionality, are a partial explanation of a non-arbitrable yield curve.

Smith & Xu (2017) also make a convincing case to value *any* investment (in their case, stocks and bonds) according to a *term structure of interest rates*. According to Smith & Xu (2017), there are two important problems with using a single discount rate (an “obvious” and a “subtle” error):

(1) Using a single rate of interest to value investment with different maturities, “(…) will overvalue long-term [investments] relative to short-term [investments] when the term structure is upward sloping and undervalue them when the term structure is downward sloping” (p. 62);
(2) Using a single rate of interest to value investments with different coupon rates / annual expected returns, will “(…) undervalue high-coupon [investments] relative to low-coupon [investments] when the term structure is upward sloping and overvalue them when the term structure is downward sloping.” (ibid).

Some authors (e.g., Damodaran, 2008), however, argue that the difference is negligible and can therefore safely be ignored. Yet, when Smith & Xu (2017) used historical term structure data to measure the impact of using the term structure instead of a single rate of interest in six discount models. Their conclusion is that, indeed, valuations errors are high when discounting at a single rate of interest: “(…) twists and turns in the term structure cause valuations based on a single interest rate to fluctuate substantially around valuations based on the complete term structure.” (p. 66)[110]. Curiously, Smith & Xu also found “(…) a relative overvaluation of growth stocks when the term structure is upward sloping and an undervaluation when the term structure is downward sloping.” (ibid). Smith & Xu’s findings

[110] If anything, with short-term interest rates at near zero over the past few years, the valuation errors have never been as large as today.

become especially important if our conclusion is valid: that is, the ‘expectations theory’ of the term structure is false.

Chapter 5: More Recent Contributions to the Theory of Capital, Interest and Money (1980 and Beyond)

5.1 Fat Tails and the Pricing of Risk in Capital

We have at an earlier stage commented on the work of Knight (1921) and, to a lesser extent, Mises (1949) on risk and uncertainty[111]. In the 1960s, however, a large part of what is now wrongly called "modern finance", a large literature on risk developed: Markowitz (1952), Sharpe (1964), Lintner (1965) published their work on the capital-asset pricing model (CAPM) and risk in the 1960s and 1970s. Fama (1971) was one of the first to criticize Sharpe and others. As Fama (1971) grounded his work on Mandelbrot's insight (Mandelbrot, 1963) that market returns do not follow a normal distribution and that therefore standard deviations do not even exist (such as in a Paretian power-law distribution), putting a bomb under more than a decade of research on risk and the pricing in of risk. If the future is characterized by incalculable risk, the construction of an equilibrium state with "risk priced in" is not an appropriate tool to study the dynamic processes that would, theoretically, lead to such equilibrium.

Both "Austrian" and modern neoclassical capital theory assume, in effect, that any profit will be arbitraged away by profit-seeking entrepreneurs. Nevertheless, this assumes that entrepreneurs are perfectly capable of pricing in (or capping)[112] risk, so that in fact only risk-adjusted profits are arbitraged away. Otherwise, what entrepreneurs perceive as profit

---

[111] See p. 201.

[112] Contracts can be used to cap potential losses out of hidden or difficult to estimate (nonlinear) risk probabilities. Insurers are way ahead of bankers in this sense. Insurers exclude certain fat tail events from coverage, which allows them to "price in" risk even though they do not know the exact probability (or probability distribution) of an event occurring. Contract theory addresses the issue of pricing in risk even though the underlying probability distribution remains unknown.

arbitrage, is in fact a nonlinear loss function[113]. Let us illustrate our point by means of an example.

Let us assume that a bank (or investment fund) is offered a stake in a newly-formed mortgage-backed security (MBS). A $100 stake in a high-quality interest-only mortgage loans with 30-year maturities at a fixed rate of interest of 5%. Now, if the market rate of interest is 4%, we arrive at a NPV of this claim on a pool of mortgage loans of **$117.29** ($100 face value, 30 years, $5 in annual interest payments, discounted at 4%) as long as we assume zero risk. Our current opportunity cost could be a 30-year government loan at 4% (a NPV of **$100** since our interest earned is equal to our rate of discount), again assuming zero risk[114].

Nevertheless, the cumulative default rate (with a loss severity of 1, that is, 100%) is estimated to be 6%. Of course, losses do not occur linearly. Hence, assuming that the probability distribution of expected losses on defaults and foreclosures equals a normal distribution or Gaussian bell curve, such differences are expected to even out over time in a cumulative rate of loss (that is, "risk"). At a cumulative default rate of 6%, the NPV of the claim on MBS is not **$117.29** but rather **$110.25** ($117.29 x (1 - 0.06)). It should be clear from this example that an apparent profit opportunity exists if the market price of the security would be **$100**, *if* our risk estimates are correct. We would net a profit of $110.25 - $100 = $10.25. Hence, the bank in our example would buy the security, increase demand, and bid up its price. It would thus arbitrage away any profit margin until no longer any potential profit exists.

Nevertheless, our *if* is a strong assumption. In the real world, since risk estimates of future events are uncertain, we will rather have different actors with asymmetric *ex ante* risk

[113] This is, of course, precisely what happened to Long-Term Capital Management (LTCM). What LTCM viewed as "profit opportunities," adjusted for risk, that it could arbitrage away while netting a profit, was in fact a misappraisal of risk. By not accounting for the nonlinear nature of risk, LTCM thought the profits the fund was earning were risk-adjusted when in fact they were not. However, risk does (a) not appear linearly over time but is often times clustered and (b) can be underestimated when risks are priced as if risk, that is, the probability of losses (or complete ruin), is modeled according to a Gaussian probability distribution (see Taleb, 2012).

[114] It is common in practice to refer to this rate as a "risk-free rate." However, outside of academia, no such risk-free rate exists.

appraisals[115]. If the *ex post* cumulative default rate turns out to exceed 6% and reaches, for instance, 15%, then our risk-adjusted NPV would equal **$99.70**, lower than our purchase price. We would incur a $0.30 loss. What we thought was profit arbitrage, was in fact a misappraisal of risk: the real-life distinction between real and illusionary profits can be deceptive.

Now, the main problem is that losses are nonlinear and risk (probability distribution) does not follow a normal Gaussian bell curve distribution. As Taleb & Tapiero (2010) write:

> "Under a nonlinear loss function, increased exposure to rare and latent events may have the effect of raising costs of aggregation while giving the impression of benefits — since costs will be borne during rare, but large-impact events." (p. 3504).

Taleb's (2015) point boils down to a systemic mispricing of risk by financial entrepreneurs. Moreover, such mispricing risk only becomes apparent *over time*, when the risks materialize. However, *financial* entrepreneurs that misprice risk and arbitrage illusionary profits will gain more resources and create even greater mispricing of risk. They earn higher returns and attract resources based on these, temporarily, higher returns which are, risk-adjusted the right way, actually lower (often times substantially lower) than the opportunity cost.

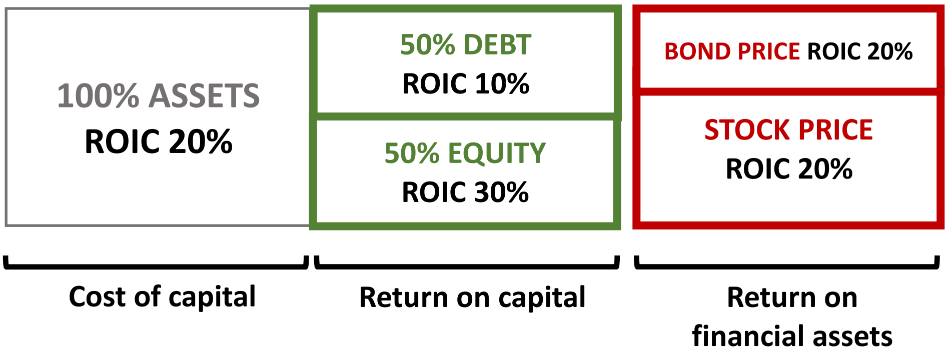
Bankers could and historically did use contracts to limit or cap the downside of loans and investments, for instance by limiting the loan-to-value, by forcing homeowners to take out mortgage or unemployment insurance or by transferring credit risk through the use of derivatives. Hence, what *can* be arbitraged and priced in is in fact *not* risk, but payoffs.

Turning to another common misconception, the Modigliani-Miller theorem is often misconstrued as "leverage does not matter."

---

[115] Probably the worst possible consequence of taxpayer-funded bailouts is not so much the simplistic assumption of "moral hazard," but rather the fact that entrepreneurs that systematically *underestimate* risk will outbid actors that *correctly* estimate future risk without being eliminated by the market process. Bailouts, hence, dramatically increase systemic risk, much more than the simple notion of "moral hazard" presupposes. Moral hazard is deliberate excessive risk-taking; what we emphasize here is nondeliberate excessive risk-taking by keeping excess risk-takers alive at the expense of entrepreneurs that correctly *anticipate* and price in risk. This is commonly not understood by decision makers.

This asymmetry of risk bearing is not often appreciated. If savers would predominantly prefer to hold debt instead of equity, the cost of debt would be low and the cost of equity high. Businesses would opt for debt financing rather than equity financing. However, by substituting equity for debt[116], businesses become more fragile, since the cost of debt is a fixed and compulsory cash outflow whereas the cost of equity is not. Moreover, debt assumes repayment of principal at some point in the future, unlike equity.



In this case, the current return of 20% on underlying assets is divided in the sense that debt holders are paid 10% in interest, whereas the remainder accrues to shareholders, which represents a 30% return[117]. However, on financial markets, *financial* entrepreneurs can borrow at 10% and invest in equity that yields 30%. Hence, financial asset prices begin to reflect the discrepancy, as stock prices on secondary markets are bid up and bond prices are bid down. Let us assume that, for simplicity sake, the equilibrium returns on both debt and equity are thus an equal 20% (which they are not, due to the seniority of the claim on the underlying productive asset), arbitraged back to the return of the underlying assets. In this strictly limited sense, of a perfect equilibrium, does the Modigliani-Miller theory hold. Yet, the real world is not characterized by perfect equilibrium, so the most that we can conclude is that to the degree the differences become larger and more extreme, the greater is the

[116] This is in fact what has been happening since the crisis of 2008 due to historically ultra-low rates of interest.

[117] Let us appreciate for a moment that in this visual representation of the interaction between the three “structures” the ROIC of one is the WACC of another and *vice versa*.

entrepreneurial incentive to bring the returns on different instruments back in line with each other.

Therefore, a “Fisher’s pendulum of returns” exists in the composition of funding[118]. There is constant arbitrage on various levels of both financial and nonfinancial entrepreneurs to equalize returns across assets, industries and financial instruments.

5.2 Money, the Law of Reflux and Asset Quality

We have earlier briefly referred to Fullarton (1845), who was one of the first to formulate what some now call Fullarton’s law of reflux. There have been some more recent attempts, however, to revendicate the law of reflux. The law of reflux, simply put, a theory that explains that issuing liabilities (which are used as means of payment) inevitably leads to a run on liquidity. An important consequence of the theory of reflux is that intermediaries – banks – cannot lend more than deposit holders are willing to save (not spend or withdraw) without suffering a drain on their reserves and thus losing assets.

5.3.1 A Brief Overview of the Law of Reflux

The conventional formulation of Fullarton’s law of reflux should be clear: whenever the supply of an issuer’s liabilities (in Fullarton’s case, notes) exceed the demand for them, a reflux occurs, that is, the liabilities are returned (or return) to the original issuer. As a result, the issuer loses assets (in Fullarton’s classical formulation, gold reserves), while the supply of the issuer’s financial liabilities contracts again. In Glasner’s (1992) words:

> “In Fullarton’s formulation, the law of reflux asserts that any excess note issue of the banking system must necessarily either be returned to the banking system in exchange for assets held by the banking system, for example, securities and IOUs or gold, or be converted into equivalent liabilities such as deposits.” (p. 877)

Some economists (e.g., Huerta de Soto, 1998) have argued that banks “create demand” for the deposits they create, in the sense that banks could simply lower lending standards and lower the interest rates it offers on loans. While it is true that banks can lower interest rates to induce borrowers, it cannot lower interest rates on the other side of their balance sheet: on deposits and debt. The rate of interest banks must pay to depositors and bond investors is,

[118] We will discuss the Fisher’s pendulum of returns theory further below, see p. 295.

under competitive banking, dependent on their willingness to hold bank liabilities in their portfolio. And whenever the interest rate differential between assets and liabilities is negative, banks would have to require higher interest rates on their assets.

According to this meaning, banks have no special power compared to other financial intermediaries. Tobin (1963), for instance, highlighted the difference between a bank and another nonbank financial intermediary: “[T]he fact that a bank can make a loan by "writing up" its deposit liabilities, while a savings and loan association, for example, cannot satisfy a mortgage borrower by crediting him with a share account. The association must transfer means of payment to the borrower; its total liabilities do not rise along with its assets. True enough, **but neither do the bank’s for more than a fleeting moment**.” (p. 7) [emphasis mine]. Tobin, here, describes precisely the law of reflux. As soon as a loan is made (and thus a deposit is added to the bank’s deposit liabilities), it will get spent, which will induce a “reflux” and will thus eliminate once gain the very same deposit that was created earlier. Hence, Tobin’s reference to “for more than a fleeting moment.”

Now, the more pressing question is: in which cases does the law of reflux apply and in which cases not?

Convertibility, to some authors (e.g., Glasner, 1992) is for instance a requirement for the law of reflux to function properly. Banks’ demand deposits (a large part of a bank’s liabilities) are convertible into reserves (a bank’s assets). The source of reflux of modern commercial banks can come from, at least, two sides:

1. Whenever a holder of a liability “spends” the liability and the liability ends up in the hands of a competing bank. The competing bank then requests redemption of the liability into reserves, either directly or (modernly) indirectly through a clearinghouse. In the latter case, the source of reflux is thus an adverse clearing (Selgin, 1888).
2. Whenever a holder of a liability *directly* converts the liability into cash. For now, deposit holders can simply withdraw their balances and convert them into physical currency, which are a liability of the central bank rather than the commercial bank itself. Hence, the bank loses assets, there occurs an “asset drain.”

The law of reflux thus implies that no single commercial bank can lend more than other depositors are willing to hold, as part of their portfolio’s, in that bank’s liabilities (commonly, deposits). Selgin (1988) justifies his theory with the law of reflux (“adverse clearings”), however, also relies on convertibility for the theory to work. Hence, central banks, are exempted from the law of reflux since in modern times central bank liabilities are no longer (directly) redeemable into the underlying assets of the central bank. Yet, according to Selgin (1988), reflux only happens in function of some exogenous determined amount of bank reserves (either monetary gold supply or, modernly, central bank liabilities). There is no other factor that limits the expansion of bank deposits. Glasner (1992), among others, show that the portfolio demand for money is what really triggers the law of reflux. Hence, the source of reflux is determined endogenously, by the amount of bank deposits that a capitalist is willing to *hold* (hold as opposed to borrow and spend).

Hence, according to these authors (Glasner, 1992), Fullarton's theory of reflux only works *as long as* there is convertibility of bank liabilities into some type of asset. Glasner (1992), for instance, writes:

> "For Smith certainly did not believe that by simply lending on the security of real bills an entire banking system could avoid an inflationary overissue. **It was convertibility that prevented inflation**. However, Smith took it for granted that, without special privileges conferred by the government, ordinary banks could operate only by committing themselves to convertibility." (p. 875) [emphasis mine]

A coherent version of the law of reflux means that intermediaries can hold an amount of assets (investments) equal to the demand for its very own liabilities. Yet, according to Glasner (1992), whenever there is *no convertibility*, and there exists *no external drain on the assets of a bank* since liabilities cannot be directly converted into assets, the amount of liabilities cannot be reduced directly by the holders of the liabilities. In the example of Glasner, redemptions or withdrawals would extinguish financial sector liabilities. In today's system, redemptions are impossible and therefore we, as holders of currency or commercial bank deposits, cannot extinguish financial sector liabilities. This is the core of the argument of Glasner in defense of convertibility.

Glasner, however, ignores the alternative that the liabilities are held and can be bought or sold to other third-parties. That is, even without convertibility, bank liabilities would be discounted in secondary markets. This drop in the market value of these liabilities would force a bank to defend its liabilities by selling assets or lose out to competition. Now, if the quality of its assets was weak to begin with, this bank basically finds itself in a death trap.

The bank will not be able to withdraw the deposits issued in excess of the demand for them (again, to *hold* them as part of a portfolio), given the market value of the assets that back the issued deposits.

Nowadays, however, legal tender laws prevent discounting of commercial or central bank liabilities[119] (they are to be accepted at face value). But that does not prevent the law of reflux from occurring. Legal tender laws only apply to domestic transactions. Foreigners can exchange a currency (say, the dollar) for their respective domestic currency. As a consequence, legal tender laws *externalize* the law of reflux to an entire country or currency-area. An expansion of liabilities is not apparent in domestic markets, but becomes apparent on international markets. The demand for liquidity moves from demand for domestic bank liabilities (both of the central bank and commercial banks) to foreign bank liabilities or substitutes such as MMMFs in domestic financial markets.

While the post-Keynesian line of thought cannot be easily summarized and is actually quite fractured, there have one common trait: money creation is endogenously determined by the system and not by either physical limits (e.g., gold) or fiat (e.g., a central bank). Some of these proponents of theories of endogenous money (e.g., McLeay et al., 2014) emphasize the "demand side" of credit as limit to the supply of credit, while other authors (e.g., Lavoie, 2000) emphasize the "supply side" of credit (the amount households are willing to save in monetary financial assets versus non-monetary financial assets or, indeed, the portfolio demand theory).

As McLeay et al. (2014) for instance suggest: "The behaviour of the non-bank private sector influences the ultimate impact that credit creation by the banking sector has on the stock of money because more (or less) money may be created than they wish to hold relative to other [financial] assets (such as property or shares)." (p. 20). What McLeay et al. (2014) miss, however, is that the law of reflux does not require repayment of debt. McLeay et al. (2014) write: "[T]he money may quickly be destroyed if the households or companies receiving the

[119] In the case of shares of money market funds (MMFs) discounting still occurs, since legal tender laws do not apply to MMF liabilities (or equity, to be entirely correct), but only to bank liabilities. For instance, in 2008, Cipriani et al. (2009) show that twenty-nine MMFs "broke the buck" and reached a net asset value (NAV) below $0.995 and that before the bankruptcy of Lehman Brothers, already five MMFs broke the buck.

money after the loan is spent wish to use it to repay their own outstanding bank loans. This is sometimes referred to as the ‘reflux theory’.” (p. 20). This view is false: what the reflux theory says is that as soon as a new deposit has been created in the process of credit creation (that is, a loan and deposit are made on the asset and liability side of the bank), that deposit is used in purchases, so that the bank suffers a reflux as soon as a competing bank refluxes the deposit for conversion.

Thus, a bank always requires *holders* of its liabilities (that are willing to hold, not spend such liabilities) to be able to lend liabilities to others that are, in turn, used to purchase assets. Put differently, no repayment of debt by the receiver of the deposit is necessary, since any competing bank would simply demand reserves of the issuing bank, either directly or, as happens in modern times, through a clearinghouse. Hence, banks who expand aggressively their lending without attracting depositors are bound to lose assets (which must be converted into reserves) to other banks, incurring costs in the process (Selgin, 1988). If such banks think the loss is temporary, they can put up current assets (e.g., US Treasuries) as collateral and borrow the reserves necessary to bridge the time necessary (which also involves a cost). However, borrowing assets comes at a cost, which ultimately reduces the profitability of the borrower. If the loss is not temporary, the eventual loss will be even larger to the expanding bank.

Thus, in absence of legal tender laws *and* in absence of convertibility, the funding/lending capacity of any financial intermediary would depend on:

D^{liab}: The demand for its liabilities
D^{liab} x $\hat{S}^{liab}$ = P liab: The market value of its liabilities
P_{assets}: The market value of its assets
i^{liab}: The interest rate the bank pays (or must pay) for its liabilities
i^{a}: The interest rate the bank earns on its liabilities
$i^{a} - i^{liab}$: The interest rate spread of the bank

S^{liab} is constant because there is no convertibility; if there was convertibility, S^{liab} would not be constant and would basically conform to the conditions outlined by Glasner (1992) which the original version of the law of reflux explains.

Beyond convertibility, whenever the interest rate spread of the bank is negative, its net worth will be impaired (due to accounting losses) and as a result the bank is forced to restrict its holdings to strictly its more profitable assets and/or lower the rates it pays on liabilities (which would, all other things equal, reduce the demand for its liabilities and thus lead to a decline in the market value of its liabilities). Here, we observe that the asset quality of banks is of utmost importance.

Hence, to summarize briefly our above scheme, a financial intermediary *not* faced with convertibility, is constrained by:

- A positive interest rate spread (a negative spread would lead to losses and would impair capital, resulting in an insolvent bank).
- The market value of its liabilities (it faces an exponential need to expand the amount of liabilities as businesses and households want to *borrow purchasing power* not nominal monetary sums). If its liabilities are discounted in the market, borrowers will need to borrow more units of the intermediary's liabilities to be able to use their bank credit for its original purpose, which would further dilute the market value of its liabilities.
- The market value of its assets (the quality of its assets; quality referring to the market value of its assets relative to book value) allows the intermediary to sell assets and withdraw liabilities to defend P^{liab} in the market. On the contrary, if the assets are of low quality, it might be the case that even by selling all of its assets, the bank cannot possibly withdraw all its liabilities at face value. In other words, insolvency would be the outcome.

As such Glasner (1992) points out, "(…) despite the incentive of banks to increase their issue of notes indefinitely, it is the public, not the banks, that determines the composition of bank liabilities." (p. 880). The law of reflux is set into motion whenever the supply (the issuance) of a bank's liabilities exceeds the (portfolio) demand for them. Whether convertibility is the defining factor (that is, required) for the law of reflux to work is questioned; in case of non-convertibility, the law of reflux still works, but effects the market value of the liability. However, ever since legal tender laws, people are obliged to accept these liabilities at face value. Nevertheless, it should be noted that convertibility still applies to modern commercial banks, albeit not to modern central banks.

5.3.2 The Relation Between the Law of Reflux and Asset Quality

As Glasner (1992) observes, the law of reflux is basically a theory about how the liability side of a bank's balance sheet reacts to the portfolio demand of these same liabilities by savers, whereas the theory of real bills or (better described) asset quality is a theory about how the composition of the asset side of a bank's balance sheet should behave. In Glasner's (1992) own words: "The law of reflux shows that powerful market forces induce a bank to accommodate a desire by members of the public to reduce their holdings of its money. How best to do so is the problem for which [the theory of asset quality], applied to an individual bank, prescribes a solution." (p. 888).

Yet, Glasner (1992) commits a mistake when he says that the real bills doctrine only applies to individual bank and not the banking system as a whole. As Sproul (2010) argues: "[Glasner and others] wrongly claim that the real bills doctrine is not valid for an entire banking system. If the real bills doctrine is understood, as it should be, to require only that money be issued in exchange for assets of adequate value, then in a banking system where all banks issue money in exchange for assets of adequate value, the quantity of money will automatically move in step with the assets backing that money, and inflation will be avoided both by individual banks and by the banking system as a whole." (p. 13).

In short, the portfolio demand for bank money is key ("consumer sovereignty"). If banks issue more bank money than savers are willing to *hold* (as part of their portfolio), **the law of reflux will return the bank money back to the bank for redemption**. If, however, banks are unable to meet redemptions (or only while sustaining severe losses) because they have invested in 30-year illiquid mortgages. **Hence, asset quality determines to which extent a bank is able to liquidate assets to meet the incoming redemptions**. Moreover, as we will see later, if banks invested in long-term credit, they will lower the long rate of interest, triggering the devastating consequences described by the 'Austrian' business cycle theory. In that case, secondary markets will be of no help since *exogenous* liquidity (that is, the liquidity of other banks) will dry up as no bank as fundamentally liquid. As Sproul (2010) writes:

> "What maintains the value of the shilling? Is it reflux or is it backing? Clearly it is backing that is of primary importance. Without assets backing the shillings, the government is not capable of buying them back, and no reflux is possible. But reflux also matters, since it is only through some kind of reflux (i.e., some kind of

> convertibility) that the assets backing a currency can ever be paid to currency holders." (p. 7)

In other words, the law of reflux is a theory about the quantity of money; the bank asset quality theory is a theory about the value of money.

With regard to the latter, Sproul (2010) notes: "What maintains the value of the shilling? Is it reflux or is it backing? Clearly it is backing that is of primary importance. Without assets backing the shillings, the government is not capable of buying them back, and no reflux is possible. But reflux also matters, since it is only through some kind of reflux (i.e., some kind of convertibility) that the assets backing a currency can ever be paid to currency holders." (p. 7)

Yet, what happens when bank money is inconvertible (as is the case of central bank liabilities)?

According to Sproul (2010), direct conversion (convertibility) is only one possible *channel* of reflux[120]: "Convertibility, like reflux, can be physical or financial, instant or delayed, certain or uncertain, restricted to some customers or open to all, at the customer's option or at the issuer's option, etc." (p. 7)

One example of such a reflux, in absence of convertibility, is whenever a central bank issued deposits to banks (which banks hold as reserves) while it simultaneously lends to banks. In this example, whenever the central bank increases the monetary base (reserves), recipient bank could use those reserves to pay down their debts: in this case, both the monetary base (reserves) and the central bank's asset side would shrink again. This is precisely how reflux would work in absence of convertibility. However, which channels of reflux are functional at any given point in time, depends on the exact circumstances of the bank. Moreover, the reflux might only impact the issuing bank *after a delay*, for instance when a central bank holds fixed maturity debt (say US Treasuries) that is repaid (by reducing the US T-account of the central bank's liability side). Hence, we can appreciate that even in absence of convertibility, the law of reflux continues to function to a degree, albeit with a delay.

---

[120] The possible *channels* of reflux is an important subject that deserves a more treatment that is beyond the scope of this work.

As a result, according to Sproul (2010), "A paper money that cannot be redeemed for a physical amount of metal could give the false impression that it is unbacked, when in truth just one channel of reflux has been closed. (…) [E]conomists, seeing that most modern paper money is not physically *convertible,* wrongly concluded that it is *unbacked*." (p. 8)

A correct bank asset quality ("backing") theory of money, thus, does not state that any overissuance of monetary liabilities is prevented by the law of reflux, but rather that there exists arbitrage against the overissuance of monetary liabilities in the sense that different channels of reflux arbitrage the value of these monetary liabilities back to their backing. This is of great importance, since suddenly the quantity of money does not matter, but rather the quality[121].

5.3.3 Critiques Against the Law of Reflux

Huerta de Soto (1998) would later criticize the law of reflux: "The theory of "monetary equilibrium" does not recognize that the *supply of fiduciary media generates, to a large extent, its own demand*." (p. 29). He argues that supply creates demand (especially as banks can simply lower the rate of interest on loans and/or lower lending standards to increase the demand for loans). Yet, seen from a portfolio view of the demand for money, we observe that this principle does not apply in the way Huerta de Soto (1998) envisages it.

Someone "demands a loan" to convert the deposit as soon as it is able to in some other (productive) asset. The borrower's deposit, thus, gets refluxed as soon as he spends it. Therefore, the bank can only make a loan whenever *someone else* is willing to hold an equivalent deposit saved up as part of his portfolio. We could say that critics of the law of reflux confuse the demand for bank money with the demand for credit (loans): although both types of demand have a corollary in a bank deposit as means of transfer, one triggers the law of reflux (the loan) whereas the other does not (the "saved", unspent deposit). One should never equate *holding* a cash balance with *spending* a cash balance. While banks could create demand for loans, they cannot create demand for deposits. That does not mean that banks do

[121] This insight is notable. Other views on modern central banks would imply that central banks could burn, throw away or sell off all of its assets, would make no difference for the value of a currency. In Section VI, we will present our own objections to this idea.

not create deposits when they lend money: they obviously do. Yet, what it *does* mean is that a deposit “created” in order to be lent, will be immediately spent and cause a reflux.

Moreover, while a bank could lower its rate of interest to attract borrowers, it has to protect at the same time the purchasing power of the liabilities it issues. Given the fact that borrowers demand *real resources* (present goods), and not bank deposits, they will borrow whatever they deem necessary given their investment needs. Hence, any sustained decline in the market value of an issuer’s liabilities will *raise* the real cost of borrowing (the cost of borrowing, thus, consists of both the expected decline in market value *and* the rate of interest charged by the bank).

It was this idea that led Tobin (1963) to write:

> “Neither individually nor collectively do commercial banks possess a “widow’s cruse.” (...) [Commercial] banks are limited in scale by the same kinds of economic processes that determine the aggregate size of other intermediaries.” (p. 6).

And (ibid.):

> “One often cited difference between commercial banks and other intermediaries must be quickly dismissed as superficial and irrelevant. This the fact that a bank can make a loan by "writing up" its deposit liabilities, while a savings and loan association, for example, cannot satisfy a mortgage borrower by crediting him with a share account. The association must transfer means of payment to the borrower; its total liabilities do not rise along with its assets. **True enough, but neither do the bank’s for more than a fleeting moment**.” (p. 7) [emphasis mine]

Therefore, the law of reflux is based on the portfolio demand for *bank money*, not the entrepreneurial demand for *bank credit*. The former does not cause reflux (as the balance is not spent), while the latter does (as the balance is spent almost immediately).

5.3.4 A Critique of the Naïve Real Bills Doctrine (RBD)

The real-bills doctrine (RBD) is a monetary doctrine that dates back centuries. Among its chief proponents was no one less than Adam Smith (1776)[122]. To sum up, a bill of exchange

[122] For our brief reflection on Smith’s real bills doctrine, see pp. 55-57.

is basically a you-owe-me (UO-ME) instead of an I-owe-you (IOU). Now imagine that company B buys goods from company A, offering a bill of exchange in return. The bill of exchange consists of a promise that company B will pay company A an amount X on future date Y (on maturity of the bill of exchange). Thus, when company A accepts a bill of exchange, it is holding a liability of company B that stipulates that company B has to pay company A in the future on a specified date.

Now, when company A sells the bill of exchange to a bank for a discount, it will receive a bank liability in return for a liability of company B. According to the defenders of the RBD, the bank is "underlying" the credit. Note that we can only reasonably speak of "underlying" a credit in the case in which the liability of non-payment is indeed transferred to the bank. In many cases, however, the bank possesses recourse against the note's endorser (usually the payee as assumed in this example).

The process outlined above is nothing out of the ordinary. What makes the assertions brought forward by the RBD different from other doctrines and, therefore, extraordinary, are mainly the following two things: (1) it is not necessary that the "discounting" of bills is supported by prior savings in the form of bank liabilities (that is, deposits); and (2) the bills, as long as they are backed by (or used in exchange for) "present" goods, can never be inflationary. This could be called the "naïve" real bills doctrine. As Glasner (1992) remarks: "[This] strong and fallacious form of the real-bills doctrine (…) denies that banknotes issued in exchange for real bills can ever be overissued. (…) [The] anti-Bullionists, applying it to the entire banking system, asserted that acquiring real bills only would allow a monopolistic bank of issue or an entire banking system (…) to avoid issuing too many notes and causing inflation." (p. 878).

The RBD, however, overlooks that genuine savings are needed to support the discounting of bills of exchange. Whenever the bank liabilities (that is, deposits or notes) that are exchanged against the bill of exchange are used to transact[123] (that is, to purchase other goods), they will *ceteris paribus* result in an adverse clearing. Whenever the discounting of real bills is not backed by prior savings in the form of bank liabilities, that is, by depositors who opt to hold more bank deposits or bank notes, it will negatively affect the clearing balance of the bank in question with respect to its rival banks. Hence, it is irrelevant whether or not the bills of

[123] We can generally assume that this is the case, since otherwise there would be no reason to sell the bill of exchange at a discount to the bank.

exchange are represented or “backed” by “real” goods. Rather, the supply of bank liabilities to lend depends on the demand to hold bank liabilities, whether the supply of bank liabilities is utilized to exchange against bills of exchange or against any other promise to pay (you-owe-me).

It is impossible for a banking system to accommodate savings in different forms than savings in bank liabilities. For instance, if people would prefer to save money proper instead of bank liabilities, the bank's function would be reduced to a mere triviality. In this case, banks could only reap profits as warehouses, where individuals would deposit their money holdings against a 100% reserve ratio. If people would prefer to hoard apples instead of bank liabilities redeemable in money, banks could still not expand loans on the basis of apples. Other banks in the clearing system would opt to redeem the bank liabilities exchanged against the apples in money (e.g., gold), even if this means that apples, a present good, are being saved and not consumed.

In short, the ability of a bank to expand bank liabilities is not without limit. Hence, there is a strict limit to how many bills of exchange a bank may accept. In addition, whenever the benefits of issuing bank liabilities to owners of bills of exchange is outweighed by the benefits of issuing bank liabilities for other purposes, that is, expanding other types of loans, banks would opt for the latter.

Friedman and Schwartz (1963) therefore rightly conclude with regard to this naïve version of the RBD: “(…) the real-bills doctrine would set no effective limit to the quantity of money.” (p. 169).” Yet, while Friedman’s and Schwartz’s (1963) statement is strictly true, it does not account for the law of reflux. The law of reflux *does* limit the quantity of money (to match the portfolio demand for money). And whenever there is reflux, the asset composition of the issuer suffering the reflux *matters*.

The real bills doctrine, then, well-defined, would no longer be a real bills doctrine: it would be a theory of the underlying asset composition, *not* limited to one single instrument: the assets’ quality, yields, *duration* and other legal characteristics and considerations are what matter: therefore, a different term, such as the theory of (bank) asset quality (that captures all of the above), would be recommendable.

The most important takeaway with regard to this present discussion of the real bills doctrine is, in a certain sense, that financial intermediaries ought to limit maturity mismatching. One way to avoid maturity mismatching is by limit lending, in proportion to the amount of deposits and short-term debts issued, to investments in short-term fixed income[124].

5.2.5 The Modern Austrian "Free Bankers": A Critique

We could certainly look with a critical eye to some of the most outspoken defenders of "free banking theory" (e.g., Selgin, 1988; White, 1999). Their general theory of free banking can be summarized in the following way:

(a) Commercial banks are limited by the law of reflux (or, as Selgin [1988] calls it, the principle of adverse clearings); whenever a single bank issues too many notes or deposits (lends too freely), its notes or deposits end right back up in the issuer's hands who will, in turn, lose bank reserves.

(b) If commercial banks expand credit in unison, the volatility of adverse clearings will go up and banks will, as a general rule, prefer to error on the safe side and keep larger reserves and/or restrict lending. Hence, the commercial banking system as a whole, cannot expand by itself if base money (reserves) does not increase, even when banks "expand at the same time." The mechanism of adverse clearings is, apparently, the *only* limit to overissuance.

(c) However, the central bank is responsible for creating bank reserves. If a central bank issues bank reserves, commercial banks are no longer restricted by the principle of adverse clearings. There are enough bank reserves to allow for a general expansion of bank credit, unbacked by prior savings. There exists a money multiplier effect; the central bank creates reserves and commercial banks increase lending in response.

[124] This is precisely what money market funds do (MMF). Regulations, however, have mostly favored banks, with exception of, of course, *Regulation Q*. MMFs rarely get into trouble. In 1978, First Multifund for Daily Income (FMDI) went bankrupt and investors took a 6% loss. Yet the average maturity of FMDI's assets was longer than two years. In 1994, the Community Bankers Fund "broke the buck," leading to a 4% loss for shareholders; yet, no "redemption run" occurred. In 2008, Reserve Primary Fund "broke the buck" due to exposure to Lehman, but eventually paid back 99 cents on the dollar (1% loss). The MMF industry is much more resilient than the banking industry as they tend to avoid maturity mismatches.

(d) The *only* thing that prevents commercial banks from expanding are reserve losses and adverse clearings[125]. A higher demand for a (commercial) bank's liabilities (e.g., demand deposits) leads, all other things equal, to lower (potential) reserve losses, according to Selgin (1988).

(e) A commercial bank's asset composition plays no important role. That is, a commercial bank with a 5% reserves-to-assets ratio and a loan book consisting of mere 30-year mortgages is less leveraged than a commercial bank with a 2.5% reserve-to-assets ratio and a loan book consisting of monthly maturing CDs.

(f) There exists no tendency toward centralization of bank reserves in absence of government (White, 1999).

While omitting the lack of empirical data backing up some of the claims of the free bankers (such as the fact that an increase in base money does not necessarily precede a credit expansion), there are some more pressing theoretical issues.

With regard to point (d), for instance, we see that Selgin (1988) and White (1999) rule out any role for the net interest margin (NIM) of banks to put an effective limit to commercial bank expansion. That is, no mention is ever made of the cost of capital of a bank and the return on capital (assets) and how the interaction between the cost of a bank's assets and liabilities affect a bank's ability to expand credit. Moreover, the attractiveness of holding commercial bank liabilities depends on their respective yields and compete with rival portfolio alternatives (such as money market mutual funds or US Treasuries). This implies that if the aggregate demand for financial assets does not increase, banks are effectively restricted by the fact that interest on assets decreases while interest on liabilities cannot be lowered.

If we look at point (e), free bankers overemphasize the role of bank reserves and underemphasize the role of a bank's asset-liability structure. For instance, in a historical account of free banking, White (1999), attributes the alleged success of the 18th century Scottish (largely) free banking system to: competitive note issue, the principle of adverse

[125] Selgin (1988) does consider the option of deposit holders to redeem their deposits for base money / currency (in Selgin's words, exchanging inside bank money for outside money), but rules it out as an anomaly that can only be caused by some type of external shock. In this sense, Huerta de Soto (1998) is right to criticize Selgin (1988) for assuming that such adjustments are always exogenous.

clearings (law of reflux) and proper reserve management, shareholder liability in case of losses and/or defaults, well capitalized but not overly dominant banks, and branch banking. There is absolutely no mention at all of the Scottish bank's balance sheets. White (1999) only acknowledges that from the second half of the 18th century to the first half of the 19th century the average ratio of "specie" (gold reserves) to demand deposits dropped from roughly ten to twenty percent to a substantial lower number of 0.5 to 3.2 percent, without mentioning the general composition of Scottish banks' balance sheets. Take a look, for instance, at this simplified balance sheet of Aberdeen's Banking Company in 1770 (Munn, 1982):

Figure 37: Aberdeen Banking Company's simplified balance sheet as of March 1, 1770

Assets		**Liabilities + equity**	
Cash (gold and silver)	33,705	Notes issued	39,718
Correspondents (debts from other banks)	3,925	Deposits	9,351
Advances (real bills)	63,592	Correspondents (debts to other banks)	23,178
Property	525	Capital (equity)	29,500

And a simplified balance sheet of the Bank of Scotland in 1871 (Munn, 1982):

Figure 38: Simplified balance sheet of the Bank of Scotland as of February 28, 1871

Assets		**Liabilities + equity**	
Cash (gold, silver and notes of other banks)	587,337	Notes issued	651,902
Bills discounted (advances)	8,270,602	Deposits	8,260,688
Government securities (and cash/short-term loans in London)	2,433,726	Correspondents (debts to other banks)	1,412,096
Railway debenture stocks, Indian government debt	215,357	Capital (equity)	1,371,117

In both instances, we can observe that the balance sheets show little to no maturity mismatches. In fact, in the first balance sheet, we can see that the sum of liabilities with short maturities are more than sufficiently backed by cash and bills discounted, at a coverage ratio of close to 1.40. In the second balance sheet, the sum of short-dated liabilities is *again* backed by ample amounts of liquid assets with short maturities (in this case, I compared cash and bills discounted to notes and deposits issued). The coverage ratio is close to 0.99 in the second example, but the relative share of bills discounted relative to cash has increased, indicating a more efficient banking operation. At any rate, both balance sheets show little maturity mismatching. The fact that deposits are matched by an equal amount of short-maturity bills of exchange would be of no importance to the above cited free bankers; they are exclusively fixated on cash reserves.

More specifically, while citing the great success of temporary suspension of payments and the general confidence in the Scottish banking system, they seem oblivious to the fact that if it were not for the liquid (short-maturity) assets of both banks (apart from cash reserves), a suspension of payment would lead to far greater losses and longer effective periods of suspension, which would easily crush any confidence in the system on the part of the general public. Liquidating self-liquidating short-term bills of exchange[126] allow for the almost immediate payment in case of large, sudden redemptions without jeopardizing the solvency of the issuing bank.

With regard to point (e), White (1999) cites Vera Smith (1936). However, Smith herself does not offer much prove other than a “Bahegot said so” and a general statement that banks must closely watch their reserve positions. Smith (1936) seems oblivious to the existence of centralized clearinghouses, where as a general rule collateral (that is, reserves, or in earlier times, physical gold reserves) is pledged and centrally held on the books of the clearinghouse. Even modern privately owned and controlled clearinghouses, such as LCH, work with centralized members’ reserves to clear financial transactions. Assuming that the natural course of financial markets is toward decentralized clearing balances make no sense; clearing would be enormously inconvenient and insecure. Moreover, centralized storage (as was the case with gold reserves) offers economies of scale on insurance and warehousing, besides eliminating the need for any costly and precarious physical gold transports between bank vaults; adverse clearings are simply accounted for by reassigning bar ownership from the

[126] I do not pretend that real bills of exchange are the *sine qua non* of banking as some defenders of the real bills doctrine do; rather, I am pointing out that in 17th and 18th century Scotland, the bills of exchange that were discounted were generally extremely liquid and quickly maturing IOUs from a variety of businesses.

reserve-losing bank to the reserve-gaining bank. The idea that in a free competitive market there exists no tendency toward centralization of bank reserves is therefore completely unwarranted.

5.2.5 The Fundamental Flaws of the Money/Deposit Multiplier Model

While the dominant paradigm *within* economics is still that commercial bank expand credit on top of the bank reserves which are created at the discretion of the central bank. As Jordan (2017) explains, this "*reserves first*" model is "the standard textbook framework for the fiat money creation process": "[O]pen market purchases by the central bank increase central bank money (the monetary base) and lending by commercial banks increases private deposits (via a money multiplier)" (p. 370). This is also called the central bank's "transmission mechanism," as it can create or destroy bank reserves which then in turn are used by commercial banks to increase or decrease lending (and thus broader measures of the money supply, such as M1 and M2). Many influential authors followed or still follow this "deposit multiplier" model (e.g., Friedman & Schwartz, 1987; Hayek, 1929; Selgin, 1988), in which the idea is that banks expand credit as a function of bank reserves. Historically, such reserves were equal to physical (monetary) gold. Nowadays, reserves are created by central banks. Hence, there exists a causal "multiplier effect" between what central banks on one hand create in reserves and what commercial banks on the other hand lend to prospective borrowers. Create more reserves and thereby increase lending, or so goes a majority of economic banking models (e.g., Selgin, 1988).

There is, of course, no denying that a ratio exists between central bank deposits on the one hand ("reserves) and commercial bank deposits on the other ("credit"). But this is far from arguing that bank reserves *produce* credit, that is, any increase in the supply of bank reserves leads to a compound increase in the supply of bank credit. This reserve-creation model is also referred to as exogenous theory of banking, contrary to the notion of endogenous bank money[127].

Now, for the reserve theory to hold, “the amount of reserves must be a binding constraint on lending (…)” (McLeay, Radia, & Ryland, 2014, p. 15). This notion, fundamental to understand the reserve theory, is akin to saying that the size an equity (mutual) fund is bound by its liquid cash reserves. The absurdity behind this idea should become directly apparent. The size of an open-end mutual fund is bound by the shares its current and prospective shareholders are willing to hold. If investors are willing to hold a greater number of shares, it can issue new shares and expand its portfolio size. As a general rule, mutual funds hold cash reserves to meet redemptions. However, to meet redemptions, it could also hold zero reserves and use bridge loans and the proceeds at liquidation to meet redemption demands. Likewise, a bank can hold zero reserves as long as it has other ways to redeem the deposits it issued: for instance by selling assets and paying out reserves, or by borrowing reserves.

Hence, what really limits a bank’s size (or, in aggregate, a banking system’s size) are the costs associated to each of these options, not the *quantity of reserves* it has parked on its balance sheet. In this sense, neither mutual funds are bounded by their cash position relative to shares outstanding, nor are banks bounded by their reserves relative to deposits outstanding.

5.1.3 The Term Structure, the Role of Maturity Mismatching and Financial Fragility

What is money and what is not is a continuum of degrees of liquidity (liquidity defined as how quick and how well an asset is or can be turned into present consumption). Milton Friedman (1977) seems to recognize the link between both. He argues that both short-term and long-term interest rates influence the demand for cash balances. However, he seems to glance over the fact that a cash balance (to which Friedman himself includes bank deposits) is credit and is part of the very term structure that he is analyzing. In other words, explaining

[127] This idea was, of course, also subject to harsh debates between the British Banking and Currency School in 19th century Great-Britain. Even within the British Banking School, which more or less started from the idea that bank money is endogenous, many differences can be observed. Likewise, many post-Keynesians adhere to the idea of endogenous money. However, some argue that the source of such endogeneity is the demand for financial (monetary) assets, while others argue that it is rather the demand for credit by entrepreneurs and businesses.

the demand for bank deposits (which is a form of short-term credit) by referring to the term structure is a form of circular logic, since the demand for bank deposits partly *determines* the term structure.

Friedman appears confused, as he concludes that “a rise in short rates and offsetting decline in long rates will tend to increase rather than *decrease* the quantity of money demanded”, which he presents as “perhaps the most intriguing theoretical result” (p. 413). Seemingly impressed, it should be obvious that an increase in the demand for liquid, short-term assets (money to Friedman) raises, all other things equal, the market rate of interest of these liquid assets. The apparent confusion can be easily explained with some historical context, however, as Friedman always made a broad divide between money and credit in his quantity theory of money which led him astray.

Curiously, Fama (2013) also shows that there is little evidence of, in this case, the Federal Reserve being able to “control” (long-term) interest rates. More than controlling interest rates, it appears that the Fed is “following” market interest rates. Moreover, Fama (2013) shows that variations in long-term interest rates is due to “time-varying” term premiums rather than forecasts of future short-term interest rates (this is well in line with our defended “habitat” theory of the term structure).

It seems, however, that empirical data falls short of being able to explain a central bank’s power over interest rates: interest rate data can be either interpreted by arguing that the Fed controls interest rates actively and that the market anticipates the Fed’s decisions (i.e., the market “follows” the Fed) *or* that the Fed follows market rates (i.e., the Fed “follows” the market). Hence, empirical data is of no use in solving this question. A wide variety of Austrian economists have opted for the former (the Fed “controls” interest rates), especially in light of their theoretical explanation of the business cycle (e.g., Huerta de Soto, 2006; Garrison, 1990). In the end, Fama (2013) concludes that the Fed exercises “some” control (a mere 17% of variance) over rates (using data from 1993 to 2013), but that such effect quickly disappears for longer maturities. Fama (2013) remains skeptical: “[T]he decline in short-term rates to near zero after 2008, despite massive injections of interest-bearing short-term debt by the Fed (and other central banks), **is a cautionary tale about how market forces can limit the power of central banks even with respect to the short-term rates that are supposed to be their special preserve**.” (p. 182) [emphasis mine].

Now, part of Friedman's initial confusion, was produced by his inability to understand the intricacies of banks' attempts to "arbitrage the yield curve." By borrowing short and lending long, they can optimize short-term gains[128] at the expense of future blow-up (Taleb & Tapiero, 2010).

Arbitraging the yield curve is the equivalent of maturity mismatching: the latter is the *cause*; the former is the *result*. A society's savings is equal to the market for financial assets. This market for financial assets consists of a wide variety of different assets: monetary versus non-monetary assets, short- versus long-maturity assets, debt versus equity, loans versus securities, fixed value (par) versus fluctuating value, et cetera. Maturity transformation, then, is a practice in which financial intermediaries finance themselves by issuing callable or short-term liabilities, while investing in assets with longer maturities. As a result, they narrow the spread between short- and long-term interest rates.

Despite large numbers of academic studies on maturity mismatching in the financial system, it is generally seen as "beneficial" by many academics, much in line with the proponents of shiftability in the 1920s. Greenwood and Vayanos (2008), for instance, imply that the "maturity structure could be used by the government as a tool to reduce financing costs and raise aggregate welfare" (p. 1). Many other authors also view maturity transformation as something desirable (e.g., Bahegot, 1873; Gorton & Penacchi, 1990; Lowe, 2015). According to this view, banks do society a great service by using liquid demand deposits to invest in illiquid long-term debt. For instance, Philip Lowe (2015) argues that: "The transformation of claims over fundamentally illiquid assets into claims that are highly liquid is one of the critical functions that the financial sector provides for the community. **Without such transformation, it is difficult to see how modern economies would work**. This transformation has been critical to the accumulation of physical capital in our societies as well as the operation of our modern payment systems." [emphasis mine]. Any bank run – the purest expression of a bank's fundamental illiquidity – is an exogenous event. As long as the

[128] Taleb (2012) calls this the "Bob Rubin trade." As many bankers earn annual bonuses over profits, without any downside when losses occur, they have the perverse incentive to optimize short-term gains by increasing long-term risk of ruin. Robert Rubin (Citi) earned over $120 million while creating hidden risks by increasingly borrowing short and lending long. When the whole scheme blew up, he defended himself by arguing that it was a "Black Swan" that nobody could predict and, moreover, that he did a great job in the previous years.

bank's clients and the public in general have complete confidence in the bank's ability to honor redemptions, no problem should arise. Hence, the task of government is to protect banks from ever being confronted with their fundamental illiquidity. Deposit insurance schemes (as well as legal tender laws protecting the market value of demand deposits) are encouraged to increase the general faith in a bank's ability to face redemptions[129]. Bank runs or, more generally, *flights to liquidity* are completely exogenous factors that befall society akin to how a bolt of lightning might befall a rather unlucky individual.

Nevertheless, there is a small subset of authors who recognize the broader implied dangers of maturity mismatching (e.g., Goodhart & Perotti, 2015). Most of these admissions are related to the notions of "financial instability" and/or potential bank runs, but mostly from an exogenous point of view. As a consequence, there is always an external or exogenous shock necessary to trigger a crisis.

Yet, the very possibility of maturity transformation *causing* future flights to liquidity is never considered. In this case, there is a mechanism through which maturity transformation leads to a liquidity crisis down the line. Flights to liquidity and bank runs would be "endogenous" events, caused by the initial act of maturity mismatching[130]. This dominant view (flights to liquidity are exogenous) is mostly the result of the application of the popular Modigliani-Miller theorem. This theorem states that the capital structure (the "liability side") of a firm (in this case, a bank) does, under certain assumptions, not affect a firm's value (Schilling, 2017). One of the assumptions is that bankruptcy (liquidation) costs equal zero. Asset illiquidity implies costly asset liquidations (Schilling, 2017). Hence, how a bank finances its assets, and as a consequence any asset-liability mismatches, are of no concern.

[129] In the Netherlands, a public debate was triggered by a bank run on a mid-sized bank, DSB Bank, in 2009, which eventually defaulted *despite* the presence of public deposit insurance covering the entire value of most deposits. As a consequence, former Minister of Finance Jan-Kees de Jager attempted to pass a law that would turn any public call for a "bank run" into a criminal offense, carrying a maximum penalty of 4 years in prison. Such measures are good examples of viewing "bank runs" and liquidity crises as exogenous, leading to desperate attempts to outlaw and discourage redemptions of demand deposits.

[130] The term "endogenous liquidity risk" has also been used to refer to an individual bank. The greater the maturity mismatch of a bank, the more it relies on outside or exogenous liquidity.

The Modigliani-Miller theorem is highly questionable. Besides the unrealistic assumption of no agency costs, not optimizing a business' debt-to-equity makes it susceptible to a hostile takeover or leveraged buyout. A firm's assets, regardless of how they are financed, have a present value that represents their ability to generate future cash flows. However, the return on equity can be arbitraged by leveraging a business. A business without debt is, therefore, more or less a high profit arbitrage opportunity. Yet, as soon as a business' optimal debt-equity ratio is reached, the difference in returns between debt and equity has been arbitraged away.

Now, apparently, asset and liabilities can be mismatched at two sides: (1) by financial intermediaries, (2) by the final users of capital (e.g., corporate businesses, etc.). While thus far this work has focused on the former, the latter remains a distinct possibility. Nevertheless, Silvers (1976) found that businesses tend to match asset and liability maturities (as measured by Macaulay duration). As Silvers (1976) concludes: "[M]anagers adjust liability maturities (and thus durations) to match the corresponding level of asset maturities, which in turn are dictated by product-market and production and organization decisions." (p. 61). Theoretically, capital markets act as a final judge on whether a firm's liability maturity structure is optimal. If a company is overdebted at short maturities, risks of losses and bankruptcy increase and the firm gets punished on capital markets by restricted access to external capital. If a business' liability maturities vastly exceed its asset structure, then the business gets punished by capital markets in much the same way (e.g., a lower stock price), but at the same time risks hostile takeovers, acquisitions and/or (leveraged) buy-outs. This contradicts Modigliani's thesis that the liability structure of a business is irrelevant to its value (Modigliani & Miller, 1958).

A maturity/illiquidity crisis, in this sense, is a prime example of Hayek's concept of "forced savings." While consumer-savers (or resource owners) express a higher time preference (i.e., they discount time at a higher rate), banks act as if a society's time preference is lower (i.e., as if they discount time at a lower rate). A recession is thus, in effect, a collective realization that liquid savings were used to invest in illiquid capital. In case of a suspension of payment (or any of its practical equivalents), the public was tricked into thinking that they could consume in the short-term while the structure of production is not able to accommodate such a wish. The public is, against its wishes, forced to "save" to allow the illiquid investments to be realized. The theory of the business cycle should therefore be characterized as a crisis of illiquidity (which, more often than not, results in the phenomenon of "forced savings").

The rather long-standing attempts to use public deposit insurance schemes to protect banks from their inherent illiquidity as a result of excessive maturity mismatching, has led some economists to discard maturity mismatching as a fundamental problem in financial intermediation. Demand deposits, short-term liabilities of banks, no longer suffer (or to a lesser degree at least) from bank runs. However, there still exist "collateral runs," which have been important in the wider financial sector (both banks and shadow banks), which are akin to past "liquidity runs", that is, the typical bank or deposit runs from the past (Brunnermeier *et al.*, 2012). A financial intermediary can be forced by other intermediaries to post more collateral if and when its (structural) liquidity is endangered. This would have practically the same consequences as a run on deposits.

Moreover, due to legal tender laws, the market value of a bank deposit is not allowed to fluctuate against the asset that serves as unit of account. As a result, any changes to reflect insolvency risk of the bank in question, must come through the interest rate channel. That is, if a bank would invest its funds under management badly and lose depositors, it would be forced to attract new deposits by raising interest rates or sell off asset and reduce its balance sheet. Interest rates would have to rise up till the point that it reflects the drop in the market value of the bank's liabilities if it were not for legal tender laws. Now, with credible and relatively broadly applicable federal deposit insurance, even the interest rate channel has become effectively sterile. Any signaling can now only come the bond market and/or the stock market.

Returning to the theory of the term structure, Fama (1981) posits and demonstrates that there exists a negative relation between changes in the consumer price index (CPI) and "real activity". "Real activity" is operationalized as the growth rate in "industrial production." Fama (1981) concludes: "**real stock returns** are positively related to **measures of real activity** like **capital expenditures**, the **average real rate of return on capital** and **output** which we hypothesize reflect variation in **the quantity of capital investment** with expected rates of return in excess of costs of capital" (p. 563) [emphasis mine].

Let us attempt to interpret Fama's conclusions. What Fama in effect shows, is that an increase in savings leads to greater investment (and with high rates of return on capital the demand for savings increases), and leads to greater capital expenditures. On the flipside, an increase in savings goes hand in hand with a reduction of present consumption, and thus puts a downward pressure on consumer prices. Stock market prices reflect the estimated net present value of firm's future cash flows (e.g., Spitznagel, 2012). With greater investment, this NPV tends to rise, which explains the increased returns on common stock, as stock prices are bid up to reflect the greater investment (and greater capital expenditures) at high rates of return on capital. Hence, both theoretically as empirically there exists a negative relationship between stock returns and inflation[131].

However, many theoreticians do not seem to recognize this relation, swayed by Fisher's analysis of interest rates in the 1930s, in which Fisher argues that nominal interest rates are a function of (1) a Wickselian natural rate of interest, or the real rate of interest and (2) the expected inflation rate. As Siegel (1998) summarizes: "Although higher expected inflation raises interest rates, inflation also raises the expected future cash flows available to stockholders. Stocks are claims on the earnings of real assets, whether these assets are the product of machines, labor, land, or ideas" (p. 161).

This line of reasoning, however, seems to confuse the NPV-concept with the concept of the stock market valuing the underlying assets with NPV. Yes, inflation would increase the nominal future cash flows that capital goods are able to generate. However, on stock markets, as Mises (1949) convincingly showed, all future net cash flows are accounted for in the present value of a stock. Otherwise, arbitrage opportunities would arise, which would either raise the present prices of the capital goods to reflect their higher replacement cost in the future or, more likely, higher stock prices. Either way, returns come down. However, as equity are equal to infinite maturity instruments, the higher discount rate will outweigh any marginal increase in the nominal future net cash flows. In other words, with a sudden and unexpected increase in the rate of inflation, the change in NPV would be *less* than the new rate of inflation. That is, the return on the stock would be *negative* in real terms.

Ceteris paribus, in equilibrium, the only result that a higher discount rate (due to a higher rate of inflation) would have, is a lower present value than at a lower discount rate. Any return on capital on top of the rate of interest, which might be due to a temporal imbalance (and

[131] This is in stark contrast to what many investors pretend in practice: stocks are often considered a solid hedge against inflation.

therefore arbitrage opportunities) in the price of underlying assets, would be arbitraged away quite quickly.

5.3 Fisher's "Pendulum of Returns"

One of the questions Knight (1944), for example, grappled with was: if marginal returns on capital are diminishing, then why have long-term interest rates remained stable for centuries. This is one of the questions our theory of the 'pendulum of returns' will be able to answer. Why does there seem to exist tendencies of "mean reversion" in a number of economic phenomena closely related to capital theory?

5.7.1 A Brief Explanation of Our Theory of the "Pendulum of Returns"

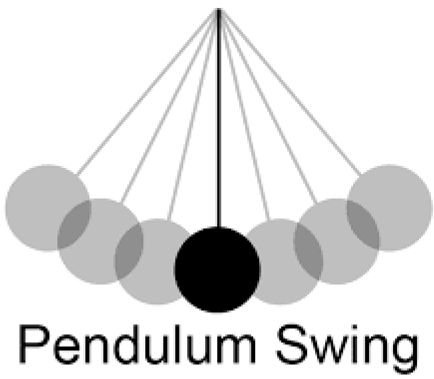


Figure 39: A visual representation of our theory of Fisher's "pendulum of returns."

Fisher's (1930) recognition in his theory of interest was close to brilliant. According to Fisher (1930), the origin of interest could be found on two sides:

1. **The "supply" side (of present goods) on the intertemporal market**
   The supply of present goods was fully captured and explained by the theory of time preference (or theory of abstinence), as we have seen before in our analysis of Fisher's work. However, the pure theory of time preference was not enough to explain the emergence of a *market* rate of interest.

2. **The "demand" side (of present goods) on the intertemporal market**
   Hence, Fisher (1930), who was highly critical of the productivity theories of interest that explain the phenomenon of interest with interest as some natural, unavoidable income on capital goods (that for mysterious reasons is never arbitraged away), offered a somewhat

different version grounded in the entrepreneur and the price structure. Not "physical productivity" would explain why demand for present goods existed (and thus a market rate of interest), but "value productivity."

Fisher (1930) referred to "investment opportunities," which as we have seen earlier, are found in the data of the price structure. Any profitable differential in prices between inputs and outputs, is an incentive of an eager and alert entrepreneur to obtain resources and start arbitraging away latent profit opportunities. Hence, the demand for present goods on the intertemporal market, is a product of the "availability" of profits (in other words, "investment opportunities").

Now, this is nothing that we have not seen earlier. The brilliant insight of Fisher (1930) revolved around the interaction between the both sides – the demand and supply side. The question that anyone active in the financial markets would inevitably have to ask himself is: why do long-term interest rates seem so stable?[132]

Fisher's (1930) recognition was extremely valuable. Possibly in his non-academic work, he recognized that a dynamic existed between both sides of the aisle: the "demanders" of present goods and "suppliers" of present goods on the *intertemporal* market. Fisher's reasoning was more or less as follows:

1. The more savings are available to be invested, the more resources entrepreneurs have at their disposal to arbitrage price differences, earn profits and thus reduce the number (and sizes) of profit opportunities.
2. *Ceteris paribus*, with a marginal decline in the attractiveness of investment opportunities, the demand for present goods (by entrepreneurs) is reduced, which leads to **lower market rates of interest** and, as such, discourage savers, on the margin, to continue saving income at the same pace or even a faster pace. This negative feedback mechanism is actually well-documented in the literature: many

[132] Van Winden (2002) shows that, in the case of the Netherlands, the long-term interest rate has moved very close to four and five percent ever since the 17th century: the only exceptions are the 1970s (due to double digit inflation) and the most recent post-recession period (with bond rates near zero).

studies concluded that there exists an empirical inverse relationship between interest rates and consumption (Balassa, 1989).

3. Now, when savers/capitalists begin to replace, on the margin, savings with consumption, they reduce the supply of present goods on the intertemporal market. This, in turn, leads to **an increase in the rate of interest** and to a situation in which profit-arbitraging entrepreneurs will have less resources at their disposal to bid up prices of inputs and bid down prices of outputs, which will lead to greater maladjustments in the price structure due to a (marginal) lack of arbitrage. Greater maladjustments in the price structure, effectively, **mean higher rates of profit**. This, in turn, encourages again the entry of entrepreneurs and an effective demand for present goods on the intertemporal markets, which eventually attracts or persuades savers/capitalists to start savings at higher rates of interest. This new flow of savings will, then, reduce again the market rate of interest.

This dynamic process on intertemporal markets might be visualized as a "pendulum," which oscillates within relatively narrow bounds, yet – due to the presence of natural feedback mechanisms – always gravitates back to its mean. Hence, the most typical characteristics of the pendulum theory is thus its reversal or regression to the mean.

We call this theory Fisher's (1930) pendulum of returns and applies even to other realms than merely the entrepreneur-saver/capitalist realm. The recent controversy on active versus passive investment, can be viewed from the same point of view (which we will do below).

The theorem of Fisher's pendulum of returns can thus be applied to many other fields. Take, for instance, the number of *financial* entrepreneurs and *nonfinancial* entrepreneurs[133]. Whenever returns on financial markets exceed the returns on nonfinancial markets, we will see resources (including entrepreneurs) exit the nonfinancial market and enter the financial

[133] When laymen colloquially refer to "Wall Street," they refer to *financial* entrepreneurs. There have been times that it was very rewarding to work in finance, while there also have been times that it was not very rewarding to work in finance. In the 1970s, for instance, finance was viewed as "boring" and not much of career choice for any bright fellow. Yet, the latter changed a few decades later. Now, Jim Rogers would say that the next decade farmers rather than investment bankers will drive Ferrari's, which echoes what we are trying to explain here.

market. As financial markets, then, become more competitive, driving down profits and reducing “investment opportunities,” entrepreneurs will on the margin abandon financial markets[134]. On the other hand, with an exodus of resources and human capital from nonfinancial markets, more maladjustments will result translating in more “profit opportunities.” Hence, the promise of higher returns on nonfinancial markets will then attract resources back to nonfinancial markets. In this sense, the resources invested in either one of both, follow the law of Fisher’s pendulum of returns.

There are many other applications, but for now we will elaborate on Fisher’s pendulum of returns in the context of entrepreneurs-versus-savers.

5.7.2 Entrepreneurs versus Capitalists and Tobin’s Q

Tobin (1963) formulated his famous quotient of asset valuation:

$$Tobin's\ q = \frac{Market\ valuation\ of\ capital}{Replacement\ cost\ of\ underlying\ assets}$$

As we have seen earlier (Spitznagel, 2011), this is akin to stating that:

$$Tobin's\ q = \frac{ROIC}{WACC}$$

Tobin (1963) would reformulate, essentially, Fisher’s pendulum of returns, in the context of asset valuation (specifically, business or firm valuation). Tobin’s q (q simply stands for quotient) represents on one side savers and on the other side entrepreneurs. Now, Tobin (1963) thought that reversals to the mean (mean q) would always come from the investment side, with entrepreneurs bidding up the prices of inputs. In Tobin’s words (1963):

> “Investment would not be related to q if instantaneous arbitrage could produce such floods of new capital goods as to keep market values and replacement costs continuously in line. For reason given [earlier], such arbitrage does not occur. Discrepancies between q and its normal value do arise. The speed with which investment eliminates such discrepancies depends on the costs of adjustment and

[134] It is probably not necessary to repeat here that for this feedback mechanism to work there is no need for financial entrepreneurs to “retrain” themselves and become business owners: substitution will more than suffice.

growth for individual enterprises, and for the economy as a whole on the short run marginal costs of producing investment goods." (p. 16)

Yet Spitznagel (2013) does not agree: "[T]here is absolutely no statistically significant or consistent relationship between [equity q] and subsequent aggregate capital investment—visible neither in changes in aggregate capital expenditures (as a percentage of invested capital) nor in changes in aggregate corporate net worth (the denominator of the [equity q])." (p. 151).

Smithers & Wright (2002) also chips in:

> "The effect of a high *q* [is] to encourage investment and discourage saving. If the USA were the only country in the world, and stocks were the only way companies could raise finance, the effects would be clear-cut. There would be a reduced demand for stocks, as consumers chose to save less, and liquidated existing stock holdings in order to consume more. There would simultaneously be an increased supply of stocks, as firms issued new equities in order to fund higher investment. The only way that the two tendencies could be reconciled would be by a reduction in stock prices. Once stock prices had fallen by enough to bring *q* back down to its normal level, incentives would have done their work, as usual." (p. 135)

Fundamentally, Smithers & Wright's (2002) argument can be summarized as:

> IF Return on capital (ROIC) > Cost of capital (WACC),
> THEN Δ Demand for savings AND Δ Savings

And, conversely:

> IF Cost of capital (WACC) > Return on capital (ROIC),
> THEN ∇ Demand for savings AND ∇ Savings

So, which of the authors is right? Let us first simply take a look at a chart that shows capital investment (as annual percentage increase/decrease) and the S&P Composite annual return. What we are looking for, is a pattern in which the S&P Composite returns *precede* fixed investment.

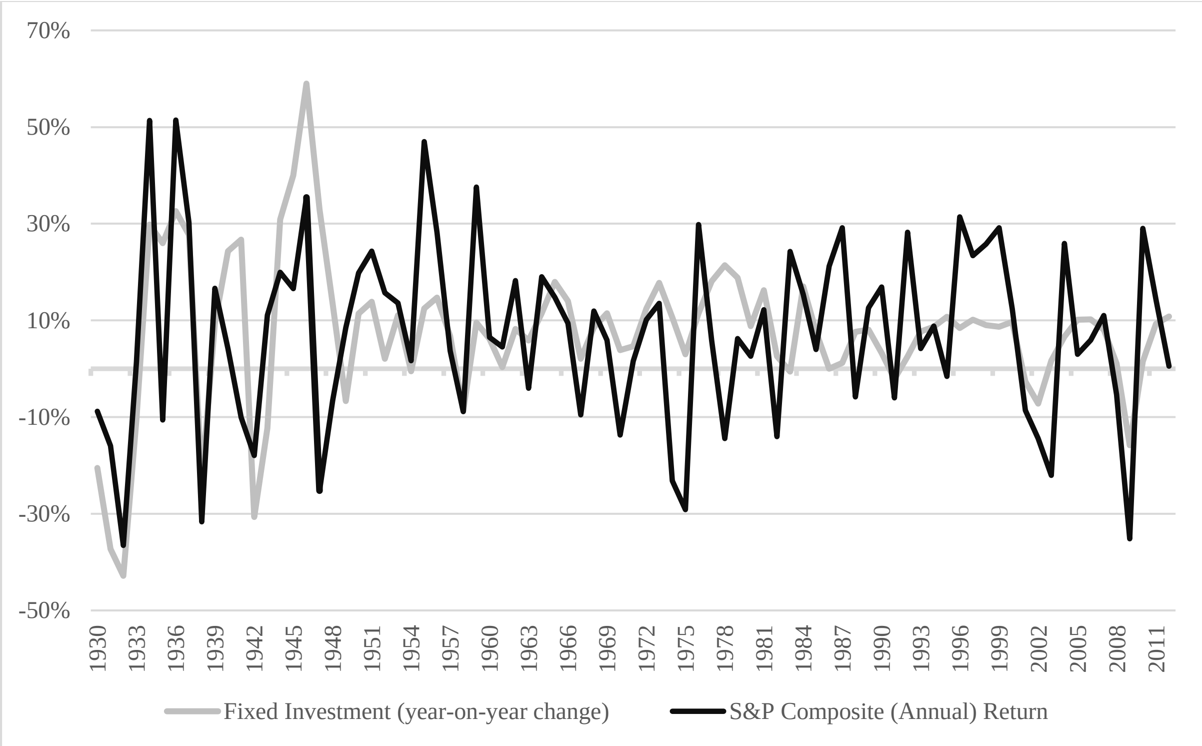


Figure 40: Fixed investment and S&P returns: is there a relationship? (Source: Shiller, St Louis Fed)

We could observe some pattern, but we could just as much be fooled with patterns that do not really exist.

Therefore, I decided to use a VAR model and, as we expect S&P Composite returns to predict *fixed* (capital) investment, we will first run a VAR lag selection model to identify the ideal lag between the two variables and whether such lag exists. We therefore specified a model with a maximum lag order of 5 (five years in this case), with the same data from Shiller (S&P Composite) and the St Louis Fed (Fixed Private Nonresidential Investment). Two of three information criteria (the Schwarz Bayesian criterion and the Hannan-Quinn criterion) indicated that the optimal lag is one year, so we began specifying models based on a one-year lag between S&P returns and investment. I also charted the coefficients across periods of the different variables (top-left returns on returns, top-right investment on returns, bottom-right investment on investment, and, as we can see in a beautiful fit, bottom-left returns on investment):

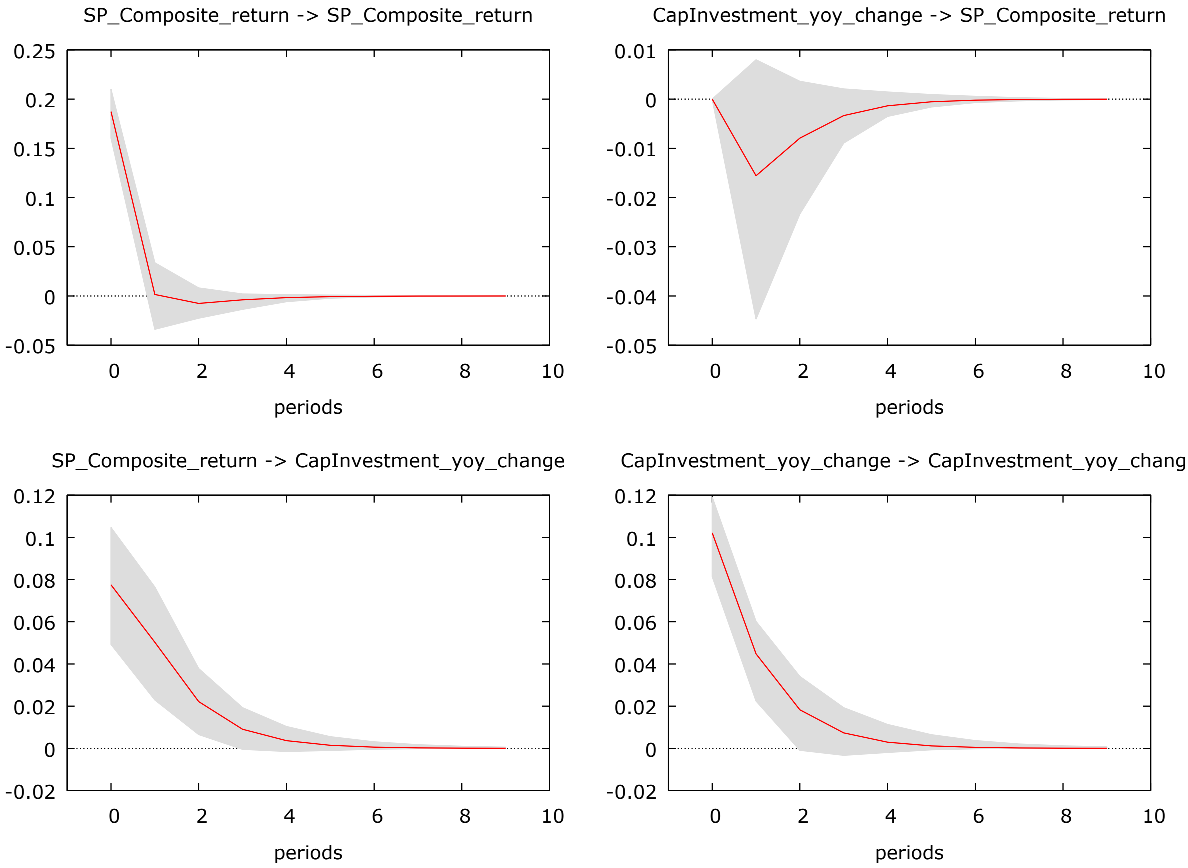


Figure 41: The relationship and time lag between S&P returns and capital investment.

The effect of S&P Composite returns on capital investment has a beautiful shape, an initial effect (at a coefficient of 0.087) which gradually diminishes as time goes by: after four or five years, no marginal increase in investment is left after an initial rise in S&P Composite returns. In the years prior, however, **S&P Composite returns do lead to more capital investment**.

However, Spitznagel (2013) is correct when he says that adjustments in equity *q* generally do not come from the replacement value side. In other words, Spitznagel might be wrong in arguing that a high equity q does not increase capital investment at all, yet he is completely right when he argues that regressions to the mean have always come from the market value side. Even though investment increases when returns and equity *q* are high, they only do so marginally.

To sum up, there exists a tendency for the cost of capital to approach or regress to the return on capital, whenever the cost of capital (WACC) exceeds the return on capital in any given industry or part of the economy. The numerator is simply bid up (stock prices in this case).

On the other side, whenever ROIC exceeds WACC, there is additional investment but not enough, contrary to what Tobin (1963) theorized, to bring *q* back to its mean. Adjustments always come from the market value side, where a large drawdown (a crash) brings prices back to their fair valuation.

Nevertheless, we can observe the same pendulum effect: when equity *q* is high (above its mean): (1) savers are discouraged and begin, on the margin, to consume and save less, while at the same time (2) long-term fixed capital investment is encouraged. Despite these two pendulum tendencies, they never reach full fruition: before they are able to bring *q* back to or near equilibrium, stock markets collapse and reduce the numerator, which does bring *q* swinging back to its pendulum starting point.

5.7.3 Active versus Passive Investment

Yet, we can also apply Fisher's pendulum of returns to the recent controversy on active versus passive investment (e.g., Malkiel, 2003). A passive investment strategy is simply buying the entire index (instead of actively trying to pick "winning" stocks) which typically outperforms a majority of active investment managers. Over the past decade, the popularity of passive investment strategies has grown manifold:

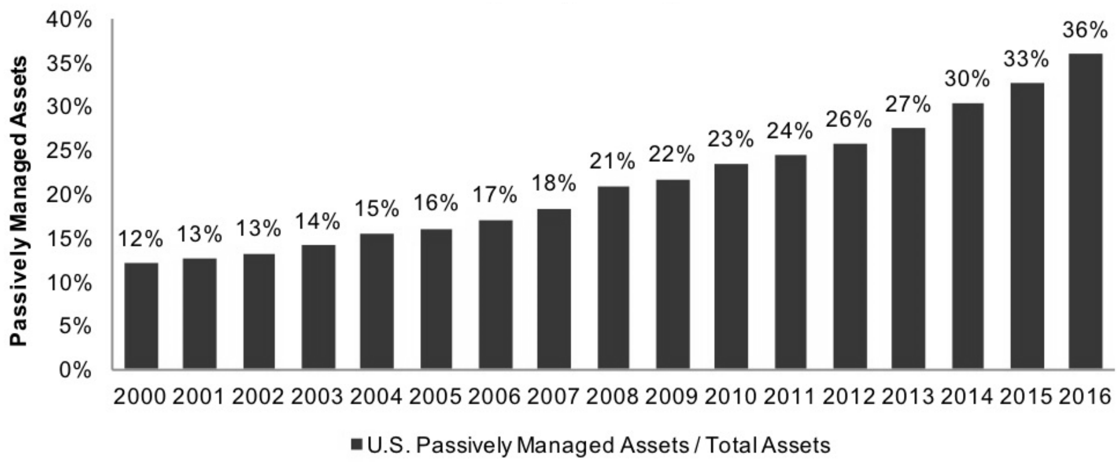


Figure 42: Percentage of passively managed (financial) assets to total (source: Forbes).

The argument here is surprisingly simple. In our original example, we could appreciate that if entrepreneurs are demanding less funds and remain on the proverbial sideline, prices would adjust less rapidly to changes in natural circumstances, consumer preferences or innovation in production processes. This, in turn, leads to higher potential profits and would raise the opportunity cost to *not* demand and invest savings.

So, how does the Fisher pendulum relate to the controversy on passive versus active investment strategies?

First, we should briefly explain what a passive investment strategy precisely is. A true passive strategy involves buying all stocks indiscriminately in an index. Yet, some of the businesses will flounder while others will succeed or surpass anybody's expectations. However, we will simply earn the average index return[135]. If we would have been able to select merely the winners and avoid the losers, we would of course, have outperformed the average. Yet, as we should be skeptical of fund managers having the rare skill to consistently outperform other investors, we could settle for simply having an average return. The opportunity cost, in this case, is the difference between what we could have earned and what we have earned by holding the index (say, for instance, 5% to an opportunity cost of 15%). However, as long as there are non-passive active investors, I can count on the fact that they will arbitrage stock market prices and bid up prices of the beauty contest winners and bid down prices of the rotten apples. As a passive investor, I am in a certain sense, partly free-riding on the price arbitrage done by other nonpassive financial entrepreneurs.

Now, in a second step, let us engage in a brief thought experiment. What would happen with stock market prices if all investor were passive investors? Simply put, there would be no effective price mechanism. Any stock would move up and down in a completely random and disconnected way from its fundamentals: businesses that are doing extremely well see their stock prices rise just as much (or go down as much) as businesses that are rapidly nearing bankruptcy and are losing customers by the droves. In this hypothetical world, stock prices would be so ridiculously priced that huge arbitrage profits could be earned by picking the most obvious stocks.

And here we stumble upon the key to Fisher's pendulum of returns in the context of the controversy on passive versus active investment strategies. To the degree that *more* investment is passive, the greater the price maladjustments and, thus, the greater the potential profits to be earned by buying the undervalued stocks and selling the overvalued stocks. Hence, as it gets more profitable to actively pick stocks, since the margin between the good

[135] Not entirely, due to transaction costs and management fees.

and the bad are becoming so ridiculously small, more active investors will get into the market and begin arbitraging, yet again, stock prices, as long as (high) excess profits are earned.

However, as soon as too many active investors try to arbitrage stock market prices, looking at the fundamental value of the businesses, the lower the returns. This occurs up till the point that a majority of the investors actually *do worse* than the average. Hence, on the margin, passive investment will be favored over actively managed investment which will then, in due time, begin balancing returns when the growth of passive investors surpasses its unobservable and everchanging optimum again.

This is how Fisher's pendulum of returns helps us to give a logical treatment of a recent controversy on active versus passive investment strategies.

5.4 Capital, Poverty and Inequality

5.4.1 A Brief Comment on Hernando de Soto's Mystery of Capital

The main question Hernando de Soto (2000) posits at the beginning of his treaty is very similar to the convergence paradox: why does capitalism and capital not flow into poor countries? In De Soto's (2000) words: "Why does capitalism thrive only in the West, as if enclosed in a bell jar?" (p. 5). De Soto (2000) contributes this phenomenon to the poor country's "inability to produce capital." As De Soto (2000) explains: "Capital is the force that raises the productivity of labor and creates the wealth of nations. It is the lifeblood of the capitalist system, the foundation of progress, and the one thing that the poor countries of the world cannot seem to produce for themselves, no matter how eagerly their peoples engage in all the other activities that characterize a capitalist economy." (p. 5).

De Soto (2000), effectively, contributes the "convergence" problem primarily to what he calls "dead capital." While poor countries *do* save, most of it is what De Soto calls "dead capital." This is property that while it is held, it is not legally recognized as such (e.g., structures on government-owned land). Therefore, there exist immense obstacles to exchange. Informal economies serve as great impediment to progress. This lines up well with the observation that in low-income economies the size of the informal economy is negatively correlated with growth: more informality leads to lower growth (Elgin & Birinci, 2016). Any asset in an informal economy possesses, per definition, no legal rights and can therefore not be transferred much less used as collateral for, for example, loans.

We can also conclude, from our earlier discussion on the term structure, that risk premia due to uncertainty are higher in many present-day poorer countries than in developed countries. This is a clear impediment to economic progress as long-term investments are discouraged by the institutional setting. To sum up, without property rights the theory of capital becomes meaningless, since without well-defined and enforceable property rights no economic calculation can be used and thus no profit-and-loss capital accounting can be applied.

5.4.2 The Lucas Capital Paradox and Development Economics

(Neoclassical) development economics came on the back of the controversies on capital theory. Despite the devastating critiques on the neoclassical notion of capital and, specifically, the production function, neoclassical economics and their growth models based on a material "capital as input" theory persevered.

In 1990, Robert Lucas published a paper on why capital does not flow from rich countries to poor countries. After all, "(…) the Law of Diminishing Returns implies that the marginal product of capital is higher in the less productive (i.e., in the poorer) economy (…). [Therefore] it is clear that, in the face of return differentials of this magnitude, investment goods would flow rapidly from the United States and other wealthy countries to India and other poor countries" (Lucas, 1990, p. 92).

We can see how the neoclassical legacy of the production function blinds Lucas from reaching the right conclusion: the neoclassical idea is that with an equal supply of labor (L), an increase in capital (K) increases economic output (y). However, there are diminishing returns on capital, such that an increase in K gives less bang for the buck in "rich countries" (where the level of the capital stock is rather high) than in "poor countries" (where the level of capital is lower).

Unfortunately, Lucas (1990) seems to have forgotten his own Lucas critique (Lucas, 1976), ignoring the very microfoundations of K. You cannot aggregate all capital in an economy and then assume an aggregate diminishing marginal return of the whole, since capital in this case is not a homogeneous quantity but rather a sum of many firms with different capital structures and returns. We can sum up the parts of the whole, of course, just as we can average the rate of return on capital for every one of these parts. However, it is a fallacy of composition to then apply the concept of diminishing returns to the aggregate[136]. As we have argued elsewhere, it is moreover the individual deviations from the average rate of return (both positive and negative) that matter to an individual entrepreneur. Lucas's (1990) starting point is, therefore, wrong to begin with.

In short, yes, there exists a tendency for capital to move to the countries where real risk-adjusted returns on capital are highest. Entrepreneurial arbitrage and the profit motive ensures so. However, the risk-adjusted[137] return on capital depends on much "softer" and institutional variables than is often taken for granted. Nevertheless, as Taleb (2012) points out, those risks might not surface for an extensive period of time. Hence, in a static view, where returns are either not at all adjusted for risk or adjusted for risk by some illusionary "country risk premium[138]," the difference in returns among (developed and developing) countries is overstated. Moreover, as we have seen above, average returns are not the relevant metric, but rather individual returns: returns differ across industries, assets, cities, methods, raw materials, et cetera.

---

[136] We have more severe objections to the theory of diminishing marginal returns on capital, since it is a *physical* or *material* theory. A *financial* reformulation of the theory of diminishing marginal returns on capital refers to the fact that there is a tendency for profits to disappear as entrepreneurs bid up input prices and increase supply / decrease prices of the output products or services. Let us assume that Entrepreneur X found a great arbitrage opportunity between markets Y and Z. This arbitrage opportunity yields him an easy annualized return of 80%, with the rate of interest at 5%. Every additional act of arbitrage (in which he uses capital to buy and bid up prices of certain goods and services and then sell those same goods or services against a profit) lowers his returns and represents a better adjustment between the means and ends of the people involved. The value of the capital stock has no *a priori* bearing on the average return on capital, which can be high despite a comparably large capital stock if many maladjustments exist (for instance because of a new technological innovation and/or a change in consumer preferences).

[137] We generally follow Taleb's notion of risk, see p. 267.

[138] In CAPM, it is assumed that country-related risks can be captured by the spread between a country's sovereign debt compared to the risk-free rate (generally U.S. sovereign debt). However, this premium that might reflect the relative default risk of any given foreign government, will never demonstrate the real 'on the ground' risks: the risk of confiscation, regulations, strikes, extorsion, corruption, delays, adverse weather and other risks are never captured by the country spread since these types of risks simply do not exist in *simple* bond payments from government to bondholder.

In addition, Lucas (1990) was blinded by the neoclassical model by considering differences in human capital. In the neoclassical production function, labor merely shows up as a material input that yields, in combination with capital, a physical output. However, measuring labor hours as some homogenous material unit makes no sense when the value-added across different laborers is extremely disparate, as is always the case in the real world.

Yet, most importantly, Lucas (1990) simply does not understand that the return on capital is produced by any given state of the *price structure*, rather than some physical input combined with a labor effort. Therefore, simply importing material capital goods does not automatically lead to excess returns, as Lucas (1990) seems to assume.

5.4.3 Piketty's Capital in the 21st Century as a Neo-Ricardian Legacy

For some apparent reason[139], Piketty's work on capital and primarily the "concentration" of capital became immensely popular ever since the publication of *Capital in the 21st Century* (2014). Piketty takes a regular production function and tortures it until it fits his inequality mold. For instance, Piketty & Zucman (2014) take a two-factor production function with exogenous productivity growth *g*:

$$Y_t = F(K_t, e^{gt} L_t)$$

Where:

$Y_t \ = national\ income$
$K_t = physical\ (non - human)\ capital$
$g = exogeneous\ labor\ productivity\ growth\ rate\ \ e^{gt}L = efficient\ labor\ supply$

[139] We can only speculate about the sociological reasons behind Piketty's success, but political expedience should come to mind. Piketty's theory on the concentration of capital justifies higher taxes on the wealthy, maximizing short-term tax revenue at the expense of greatly impairing the generation of future wealth. A sense of injustice regarding wealth distribution and resentment among the general public is another often-cited motive. This, however, seems less likely since Piketty's conclusions failed to resonate with a majority of voters, for instance in the 2016 United States presidential election. Piketty's work *did* resonate with the political class, which Taleb (2016) refers to as *Intellectuals Yet Idiots* (IYIs): "He speaks of "equality of races" and "economic equality" but never went out drinking with a minority cab driver."

Curiously, Piketty (2014) uses the neoclassical production function that we criticized earlier. What Piketty in effect assumes, is that, in $r = g$, the return on capital (r) exceeds output (g) consistently. This leads to increasing wealth concentration and a "rich get richer" effect.

As we have attempted to demonstrate the incoherence of excluding human capital from the concept of capital in earlier debates[140], Piketty (2014) seems to largely ignore the theoretical arguments made in past debates. Indeed, he mysteriously asserts are "many reasons for excluding human capital from our definition of capital" (p. 46). But he offers only one: "human capital cannot be owned by any other person" (ibid). The same argument has been made over and over again, but it is valuable to mention this fact as a demonstration of how modern-day concepts of capital lack severely in coherence.

Beyond Piketty's misstep regarding *human* capital, Piketty's argument is, in fact, a restatement of Ricardo's "principle of scarcity" (1817). A recent paper by Knoll, Schularick and Steger (2017) actually concludes:

> "The importance of land prices for wealth brings Ricardo's famous principle of scarcity to mind. Ricardo (1817) reasoned that economic growth disproportionately benefits the owners of the fixed factor land. Writing in the 19th century, Ricardo was mainly concerned that population growth would push up the price of corn so that the land rent and the land price would continuously increase. In the 21st century, we may be more concerned with the price of residential land, but the underlying mechanism remains the same." (p. 33)

However, this is pure nonsense. Piketty's statements would imply that for over a hundred years investors are consistently and persistently doing a terrible job in estimating the future yields on land. The price of land tends toward, just as any other production factor, the net present value of its future services. Moreover, land is bought and sold all the time by actors that at $t = 0$ have a better (more profitable) marginal use for the land and are therefore able to make a compelling offer. The (up to a point, unexpected) increase in land prices and, as a result, economic concentration of land wealth is a result of pure entrepreneurial profit (as buyers bought the right pieces of lands in the right areas), zoning laws (which have created an artificial scarcity of, for instance, residential land), and building restrictions (especially when

[140] See p. 183 for Knight's and Machlup's defense of human capital as capital.

it comes to the construction of high-rises; this effect is particularly disturbing in European capitals).

Curiously, Piketty explains the concentration of wealth is the “inevitable” result of a capitalist economy due to this flawed theory of capital, and then asks for intervention to fix this market failure. A clear empirical refutation of Piketty’s and Ricardo’s fatalistic views can be clearly observed in Japan. In Japan, land prices were bid up to such heights that the annual return on residential (both urban and nationwide) land since 1990 has been without exception negative. Here we can also see that risk is not a function of mere probability, but that risk increases as relative prices increase, since risk-pay off distributions got greater left tails (downside) and reduced right tails (upside). See below a chart of historical land prices in Japan:

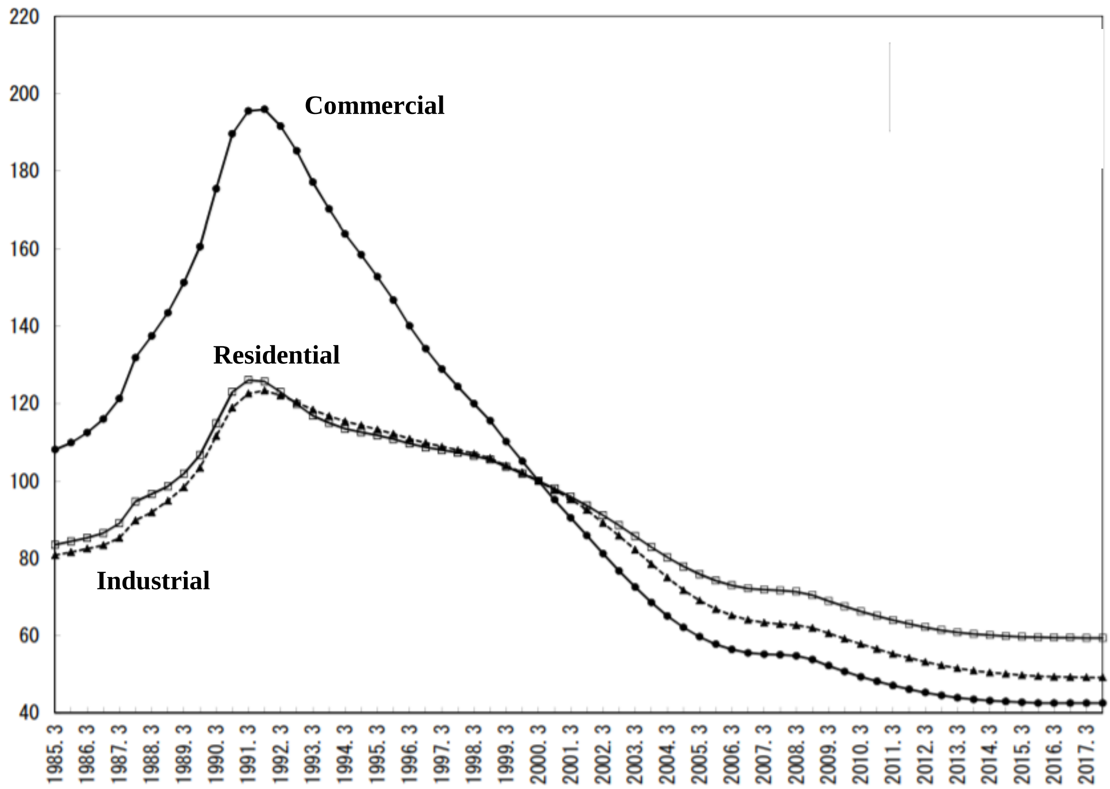


Figure 43: Japanese land price index (nationwide) from 1985 to 2017 (end of March 2000 = 100). Source: Japan Real Estate Institute (JREI)

Moreover, if land owners *would* earn a higher return on their land in perpetuity, they would end up owning everything, which is clearly a fallacy.

Hence, we can conclude that Piketty’s (2014) theory on equality might be politically embraced, but factually and theoretically incorrect. It is a mere restatement of century-old Ricardian fallacies. Policy makers should focus on creating wealth, rather than equality.

Chapter 6: Summary and Conclusions from Literature Review

6.1 Capital

The tragic legacy of the classical thinkers (e.g., Smith, 1776; Ricardo, 1817) was to think of production as something physical: capital, understood as material capital goods, were inputs next to labor and land. By combining some physical capital goods (e.g., machines), with labor and original or permanent resources such as land, a given amount of physical output can be produced. Mankind gets richer as physical goods become more abundant. Moreover, every "input" (as stock) would have its own "income" (as flow): capital begets interest as the inevitable result of its physical productivity (akin to how an apple tree inevitably grows apples), land gives rent and labor earns wages. This classical triad of material production was here to stay.

This regrettable view took a foothold in economics when thinkers such as Böhm-Bawerk adopted the same view. Capital was a combination of heterogenous goods, mere inputs, next to labor and land. Cobb-Douglas (1928) would later turn this notion into their production function, which represented a theory of *physical* production. Capital were mere machines and equipment, nothing more.

This unfortunate material inheritance affected many thinkers even in more recent times. Modern-day Nobel Prize winners such as Solow (1956) and Swan (1956) still follow the same material theory of capital as their 18th century peers. They took the Cobb-Douglas theory of production to its extreme. Capital is a physical input that, together with labor, yields a physical output. Since capital as a physical quantity is immeasurable, we will settle with the next best: taking the physical quantity at a monetary cost depreciated at an equal annual rate. This pragmatism, however, led Solow, Swan and Samuelson to defeat in the third round of capital controversies, the famous Cambridge Controversies.

However, these views of capital are effectively *not* economics. They are technical theories of production, badly applied to the whole of an economic complex. Economics is about coordination, not physical production (e.g., Kirzner, 1960). The economic problem is how we can use the resources we have at our disposal to satisfy, to the best of our ability's, consumer preferences. Moreover, such resources are not given (Huerta de Soto, 2010) and, more often than not, intangible instead of tangible. Hence, a large part of what is considered "capital theory" or even "*applied* capital theory," has nothing to do with economics.

Contrary to the physical theories of capital, Menger (1888) made, however, a first step toward a truly subjectivist theory of capital. Fisher (1930) and (partially) Knight and later Mises (1949) followed in his footsteps. The Mengerian theory of capital was largely forgotten in economics at the start of the 1960s the Cambridge controversy, but discussed and practiced in finance at large especially in explaining financial asset returns (CAPM, et cetera). In Menger's (1888) subjectivist view of capital (in which capital equals *net worth*), labor is simply a subset of capital, that is, human capital.

Capital, as such, is essentially forward-looking: it is concerned with value, specifically, the *present* value of the future income (and consumption) that a capital structure is able to provide to its consumers. It has no room for physical or material, arbitrary classifications of economics goods: it puts the entrepreneur central in its conception. Capital, as such, is a tool of economic calculation, of capital accounting (profits and losses), and is far from being a material *input* to a production process, an *outcome* of the present-state of a structure of underlying productive assets. Classifying economic goods according to their historical origin (for instance, capital goods as a "produced" means of production, that is, a prior combination of land and labor) is nonsense: such a backward-looking view of capital has no room in an economics concerned with the subjective wants of its contributors.

In some cases, capital was analyzed and defined as a "subsistence fund" (typically a "wage" or "wage-rent" fund). While such attempts were futile, they did show an important link between the income that a given capital structure is able to yield and the intertemporal consumption patterns of households.

To briefly summarize and visualize the different theories of capital we have discussed in the past few chapters, we will assess them according to different tenets:

(1) Is the theory forward-looking or backward-looking?
(2) Is the theory physical/material ("capital as an input") or subjective/financial?
(3) Does the theory separate capital from labor, land and other resources (which, as long as capital is seen as input, is often inevitable)?
(4) Does the theory make any distinction between fixed and working (circulating) capital?
(5) How does the theory explain and incorporate the phenomenon of interest?

(6) What are the contributions, if any, of the author (or authors) to the broader discussion on capital theory?

Our overview, which is based on specific authors rather than specific schools or lines of thought, can be found below:

Author(s)	**Capital**	**Circulating (working) / fixed capital**	**Value of capital**	**Interest**	**Labor**	**Contribution(s)**
Böhm-Bawerk (1888)	Physical concept, backward-looking	No, but capital as "subsistence fund"	Physical productivity (roundaboutness)	Time preference, physical productivity of "waiting"	Separate factor of production, prerequisite to capital, capital as "subsistence or wage fund"	Critique of naïve productivity theories of interest
Clark (1899), Knight (1934)	Finance concept, forward-looking	Unimportant	Market value of productive assets in terms of money	Rate of profit, physical productivity, complete lack of time	Labor is human capital	Critique of material theories of capital, human capital
Fisher (1930)	Finance concept, forward-looking	Unimportant, crude 'expectations' theory of the term structure	Market value of productive assets in terms of money	Time preference (supply), "investment opportunities" (demand)	Labor is human capital	Theory of interest, elaboration of capital as financial concept
Menger (1888), Mises (1949)	Finance concept, forward-looking	Unimportant	Market value of productive assets in terms of money (*net worth* approach)	Time preference	Allows for human capital	Capital as financial concept, role of entrepreneur
Hayek (1941), Lachmann (1956)	Physical concept, separate from "permanent" resources	Question of degree, ability to convert into consumer good (liquidity)	Misleading, focus on capital as complex of heterogenous goods of different "orders" and degrees of specificity	Time preference	Separate factor of production	Lachmann's "capital structure," liquidity of assets (money along a "liquidity continuum"
Solow (1956), Swan (1956), Samuelson (1960)	Physical concept (in dollar terms), backward looking	Assumes no difference between circulating and/or fixed capital	Rate of profit	Rate of profit, physical productivity	Separate factor of production, with different unit of measurement (labor hours)	None

Robinson (1953), Sraffa (1960)	Physical concept, backward looking (past labor inputs)	-	Wage rates (labor theory of value)	Rate of profit, surplus that accrues to capitalist at the expense of workers	Separate (yet important) factor of production	Critique of neoclassical contradictions on capital

Figure 44: An overview of the most prominent economic thinkers' theories on capital and their most important elements.

6.2 Interest

The fallacies of what Mises (1949) called the "naïve" productivity theories of interest were refuted by Böhm-Bawerk (1888). Böhm-Bawerk would later replace it with a "less naïve" productivity theory of interest, in which the rate of interest was determined by 'time preferences,' yet one of the causes of 'time preference' was the superior material productivity of more "roundabout" production processes. Unfortunately, despite Böhm-Bawerk's devastating critique many economists persisted in this fallacious theory of interest: Knight (1934), for starters, attempted to explain interest as some yield inherent to capital for the sake of being capital. This position was later adopted, in the famous "Cambridge Controversies," with little consideration by the neoclassical economists swayed by the Cobb-Douglas production function in all its colors.

Others, largely expounded crude versions of the 'pure time preference' theory of interest, which is grounded in the subjective valuations regarding time and consumption over time of consumers. This was, for instance, the theory of interested proposed by Mises (1949) and Hayek (1932).

It was eventually Fisher (1930) who completed the theory of interest by showing that while the theory of 'time preference' (or abstinence) explains a great part of the phenomenon, an intertemporal demand side was lacking. As such, he integrated a non-material, financial theory of "investment opportunities" with the theory of time preference to replace the naïve physical and material theory of (marginal) "productivity." This was one of the main contributions of Fisher (1930) to the theories of capital and interest.

Later contributions were made by, for instance, Culbertson (1957), who first laid out the principles of a theory of the term structure of interest rates as separate, "segmented" markets. Any maturity cannot be readily substituted with another maturity. Therefore, there does not exist *one* single rate of interest, but multiple rates of interest at different maturities. In this light, we could reinterpret Keynes's theory of interest largely based on liquidity preference; Keynes's theory is an incomplete explanation of interest since it only explains the isolated phenomenon of long-term rates as a *cause* of short-term rates without taking into account the true dynamics that underlie the term structure.

6.3 Profits and Losses

Few authors have been able to explain the *origin* of profits and losses. Most theories of capital (e.g., Clark, 1899; Knight, 1934) incorporate a theory of interest in which the rate of interest simply equals the rate of profit. Interest, following the classical classification of interest as income, is the income a capital goods yields. Merely because of its (physical) productivity, such a profit exists. The rate of profit, in many of these interpretations (e.g., Knight, 1934; Ricardo, 1817) are cataclysmic, since an increase in capital (again, as input) leads to diminishing returns. The rate of profit is thus destined to become lower and lower and lower, even leading to a "crisis" of capitalism. Profits, in this sense, are just as natural as the apples that fall off an apple tree; they exist merely because of the inherent physical productivity of the capital good.

This myth has, of course, largely been refuted by Böhm-Bawerk, who showed that if this were true, there would exist an arbitrage profit to be earned by alert entrepreneurs who would eventually bid up the price of the capital good until no profit remains. Hence, physical productivity could not explain the rate or existence of profits. Profits are not "natural."

The explanation that prevails amounts to stating that the rate of profit (put differently, the return on capital) is exactly zero in equilibrium. Whatever is left at that point is interest, but as the price of asymmetric time preferences rather than profits due to physical productivity (e.g., Mises, 1949). Profits are, then, a temporary phenomenon that prove a maladjustment exist between prices, that is, between input and output prices, which by a process of entrepreneurial discovery and competition, disappear as the system tends toward equilibrium (e.g., Huerta de Soto, 2010; Kirzner, 1960). Profits are an *ex post* "reward" of taking advantage of an (expected) price differential. Moreover, what matters is not the *average* rate of profit, as many (e.g., Samuelson, 1966) seem to think, but the variability of profits and losses (for instance, high positive returns in one sector against negative returns in various others), even though the *average* rate of profit is zero.

There is a reason, however, why a rate of profit persists, as we have explained with our theory that we have labeled Fisher's pendulum of returns. Entrepreneurs need resources to be able to arbitrage prices (and thus earn profits); the more resources available, the lower profits tend to be as they are arbitraged away. This, however, discourages savings which limits resources entrepreneurs can use to arbitrage away profits, creating a countermovement. This is the origin of Fisher's pendulum of returns. The rate of profit, thus, explains *why* entrepreneurs demand present goods on intertemporal markets. As we have seen before, Fisher's theory of "investment opportunities" is necessary to supplement the "pure" theory of time preference. Without prospective returns to be earned, little to no demand would exist on intertemporal markets.

We should, at this point, that in a closed accounting system, no aggregate profits or losses can actually exist. Ludwig von Mises (1949) commits an error when he compares a growing and a contracting economy[141]. He argues that in a growing economy a greater amount of capital is invested (per capita) and that *therefore*, in the aggregate, profits exceed losses. In other words, in a growing economy as Mises (1949) defines it, the return on capital or aggregate profits would be per definition positive. A contracting economy the exact opposite would occur, as the (average) return on capital or aggregate profits would be negative. Put differently, in a growing economy (with more capital invested) the aggregate average rate of profit is positive. Profits are, then, a temporary phenomenon to the degree that capital is accumulated. However, this notion is false, since profits can exist with or without increases in capital. Better yet, we could expect profits to fall to the extent that capital is accumulated, as entrepreneurs have access to more resources to arbitrage away profits.

Profits are, moreover, not the "income" earned on a given set of physical (capital) inputs that are left after paying wages to labor inputs. Solow (1956) and Swan (1956), according to their own model, view profits as a bounty, a share of physical output, that accrues to the owner of a capital good after having paid off the labor-side of the equation. This is false, since human capital is a part of capital and income that accrues to owners of human capital is very much equal to the income that accrues to owners of nonhuman capital. Moreover, profits are not an automatic or "spontaneous" result of mere ownership, as Cambridge, U.K. seemed to think (e.g., Robinson, 1953). Capital does not beget profits and wealth does not beget additional

[141] Ludwig von Mises uses the term "progressing economy" and defines it as "an economy in which the per capita quota of capital invested is increasing"

wealth. Capital equals simply the present value of all future consumption, represented by the many individual income streams that are earned, bought and sold by entrepreneurs of different stripes.

6.4 The Entrepreneur

Early formulations of the entrepreneur that we have highlighted throughout our literature review were focused on the entrepreneur as an equilibrating force, that is, a profit arbitrager that identifies an *ex ante* profit opportunity (a certain excess return over interest) and acts upon such a profit opportunity, by buying inputs and selling outputs, to realize an *ex post* profit. As such, entrepreneurs are the force that lead to an unassailable equilibrium in which the rate of profit is zero and only the rate of interest remains. This is the theory of entrepreneurship espoused by J.B. Clark (1888), of Böhm-Bawerk (1888), of Menger (1871), of Fisher (1930), of Mises (1949) and was later even further developed by Kirzner (1960) and Huerta de Soto (2010), but later completely forgotten in the context of capital in later controversies on capital theory. The participants in the Cambridge controversies of the 1950s and 1960s, for example, have never considered the theory of entrepreneurship in the context of their debates on capital. Therefore, unfortunately, many preventable mistakes have been committed; yet it also simplifies our present summary of the theory of the entrepreneur.

To summarize, the entrepreneurial profit-arbitrage theory is "objective," in the sense that the prices of all economic goods at any given instance are given. To the degree that by buying certain goods and subsequently selling them, either after a more or less explicit process of production, the entrepreneur bids up the prices of the inputs and bids down the prices of the output, which — if his endeavor is a profitable one — will marginally lower the rate of profit on his venture *until* no longer any net profit is left (net of interest, that is). As Mises (1949) adds: "Production is not something physical, natural, and external; it is a spiritual and intellectual phenomenon. Its essential requisites are not human labor and external natural forces and things, but the decision of the mind to use these factors as means for the attainment of ends." (p. 141).

The most contrasting theory of the entrepreneur can be found in Knight (1921). The Knightian entrepreneur is a "bearer of uncertainty." The future is inherently uncertain and incalculable, argues Knight (1921). Entrepreneurs shoulder the burden of such uncertainty. Yet, in the theory of the profit-arbitraging entrepreneur, which is a role rather than a person, the entrepreneur is essentially *resource-less*. Therefore, the uncertainty is shouldered not by the entrepreneur, but by the resource provider. Ultimately, savers themselves (also called *capitalists*) will bear the ultimate burden of uncertainty. Losses will ultimately fall on (a given subset) of savers, *as they are the resource providers* that allow entrepreneurs to engage in profit (return) arbitrage.

Hence, even though the *ideal type* of the entrepreneur is *resource-less*, that does not imply that they will not control resources, just that they will control *other people's resources*. The Kirznerian theory of the entrepreneur, in which the entrepreneur arbitrages away profits in a competitive setting, accounts for said "resourcelessness."

Section III:
Proposing a New Theory of Capital: Toward a Truly Subjectivist Theory of Capital

Chapter 7: A Proposed Definition of Capital, Interest, Consumption and Savings

Section II was wholly dedicated to breaking down the, what we call, material or objective theories of capital and highlight breakthroughs and advances of subjective (and financial) theories of capital. In this chapter, I will put forward my own proposal for a coherent, in subjective terms, theory of capital. A capital theory previously elaborated by Menger (1888) is revived and extended based on the concepts of the individually acting entrepreneur and his monetary calculation. This introductory chapter will only recollect and outline my own definitions of the relevant concepts in capital theory (that is, capital, interest, consumption and savings), each with an explanation of how we arrived at the proposed definitions and the relevance of each definition. If deemed necessary, concepts will be linked to (simplified) balance sheets or finance concepts. This chapter thus serves as a way of understanding the world.

7.1 Definition of Capital

Capital equals the *net worth* of an individual, a business or country. Put differently, *net worth* equals the value of assets net of debt (the market value of assets minus the market value of debt). This allows us to avoid any trouble with double counting asset values. The *market value of net assets* reflects, in equilibrium, the present value of the future income they are able to generate. What in fact matters, is not the physical origin or objective characteristics of assets, but rather how the entrepreneur views and acts upon them.

Take a given financial asset that might be part of someone's net worth. In the aggregate, debts will be cancelled out against each other. If a debt is part of someone's asset column and thus his net worth, it is subtracted from the counterparty's net worth since it is net of debt. Imagine:

Balance sheet X

Assets	**Liabilities**
IOU from Y 10	-
Other assets 90	**Net worth 100**

Balance sheet Y

Assets	**Liabilities**
Other assets 100	IOU to X 10
	Net worth 90

Figure 45: Debts are automatically cancelled out against each other when we aggregate net worth and we thus avoid double counting of assets in capital accounting.

In the above example we see that the "other assets" account (let us call these assets "productive assets") adds up to 190. Total assets, however, equal 200. Given the fact that there exists an IOU of 10, we can subtract 10 from 200 to get to a *consolidated net worth*. The result, 190, is exactly equal to total nonfinancial, productive assets, since any financial asset has two sides to the equation (a financial asset is someone's asset and someone's liability).

Nevertheless, in case we use an *aggregate net worth* of a specific country or society, we run into trouble regarding double counting since we count equity twice: once as asset (as shareholdings of a business) and once as net worth on the liability side. Imagine:

Balance sheet X

Assets	**Liabilities**
Equity from Y 10	-
Other assets 90	**Net worth 100**

Balance sheet Y

Assets	**Liabilities**
Other assets 100	-
	Net worth 100

Figure 46: Equity holdings (as assets) are not automatically cancelled out against each other in our net worth *approach to capital theory.*

In this case, aggregate net worth adds up to 200, even though (nonfinancial) productive assets only equal 190. In case of *private net worth* (either of an individual, a household or a business) no such aggregation problem exists. However, even if we aggregate net worth across businesses (as we would on, for example, country level) this case is of little consequence. In fact, in practice, businesses aggregate net worth all the time in consolidated balance sheets of a myriad of separate legal entities. A *consolidated balance sheet* of the parent company includes all the individual balance sheets of its subsidiaries. Several corrections are made so that the consolidated balance sheet truly reflects all its components:

(1) Any revenue of the parent that is an expense of a subsidiary is subtracted from the consolidated account;
(2) Any account payables or receivables between parent and subsidiary is subtracted from the consolidated account;
(3) Any equity (in case of full ownership) to avoid double counting[142].

In effect, we see that modern accounting and finance practice allows us to operationalize Menger's (1888) original theory of capital and do away with Mises' (1949) objections to aggregating capital. That said, we should never forget that net worth is privately calculated and has no objective existence outside of the entrepreneurial mind. Different entrepreneurs can reach different conclusions about the value of the assets at their disposal. Nevertheless, the cost of being wrong is borne by the entrepreneurs themselves. Hence, entrepreneurial initiative is inescapably connected with *net worth*. In the absence of entrepreneurs, no capital (*net worth*) would exist. The idea that an aggregate net worth allows top-down central planning should therefore be rejected. What it does suggest, however, is that governments can devise policies that help entrepreneurs to effectively apply economic calculation. Economic calculation is a society's compass that allows entrepreneurs to steer production into the lines that are most beneficial according to consumer preferences. To allow for better coordination between means and ends, policy should be geared toward improving entrepreneurs' economic calculation and reduce institutional uncertainty, for instance by assuring stable rules and efficiently protecting property rights.

[142] For a more detailed description of how aggregate net worth is or should be consolidated, consult PWC's guide *Consolidation and equity method of accounting* (2015).

Curiously, we can also see clearly that any capital (defined as *net worth*) ultimately must be part of an individual's estate or balance sheet. Any *net worth* held in corporations, must ultimately be owned by individuals[143]:

Figure 47: Net worth of businesses is ultimately net worth of households.

Hence, capital has another side to it. Since capital equals *net worth*, it has an accounting existence on the right side of the balance sheet:

Figure 48: The basic accounting existence of capital

But capital, in this sense, is backed by an equal amount (in money terms) of underlying productive and specific, heterogeneous assets. However, the exact combination of heterogenous assets lies beyond the scope of economics and is, in fact, an entrepreneurial field. Only entrepreneurial theories can be proposed with regard to their exact composition. Whether they consist of apples or pears is irrelevant to the economist. Moreover, such entrepreneurial theories will never be "economic laws," but rather weak conjectures bound to change over time.

[143] Nowadays, it should be clear to any observer that a great majority of capital is ultimately concentrated in the hands of middle class working people in developed countries through pension funds, banks and other types of financial intermediaries.

The value of such assets *is*, however, a key object of study within economics. In fact, we could say that the (market) value of the underlying assets is the *cost of capital*:

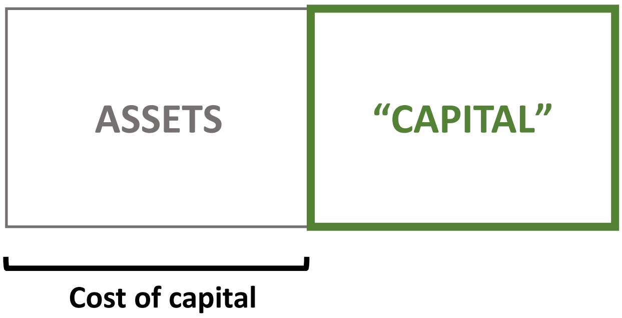


Figure 49: The underlying productive assets as "cost of capital"

The value of the underlying assets equals the "replacement cost" of capital[144]. If the value of capital would differ from the value of the underlying assets, a disequilibrium exists and thus a profit opportunity emerges (or, in other words, a return on capital that exceeds the cost of capital emerges).

We could express the cost of capital as a percentage cost and the return on capital as a percentage return. In equilibrium, the cost of capital would equal the return on capital (WACC = ROIC). But this is the same fact expressed in a different way (Spitznagel, 2011). Hence, the value of capital is equal to the value of the underlying productive assets or $Value/Invested\ Capital = ROIC/WACC$.

What determines the value of capital? Capital simply equals:

[144] The cost of capital is normally expressed as the rate of return that could be earned by investing in another project. However, in aggregate and in equilibrium, all new investments would equal a zero net present value of capital goods at the going rate of interest. All prices of the underlying assets are bid up to such heights that they reflect said rate of interest, with zero surplus profits left. Hence, the return on the underlying assets on top of the rate of interest is zero at their given equilibrium prices and the cost of capital would equal the return on capital. Put differently, it makes no difference whether we express equilibrium as a percentage cost of capital that equals a percentage return on capital or express equilibrium as absolute price of the underlying assets (cost of capital) against the price of the claim on these underlying assets (return on capital).

$$NPV = \sum_{t=0}^{n} \frac{CF_t}{(1 + r_t)^t}$$

where

CF_t = cash flow at time *t*

r_t = rate of interest at time *t*

In other words, capital equals the net present value (NPV) of the sum of the current and future cash outflows (expenditures) and inflows (income) discounted by the term structure of interest rates[145]. The discount rate is an opportunity cost of the funds invested in any given asset or venture (Stowe & Gagné, 2012).

Put differently, capital as *net worth* is a claim on an underlying asset or underlying assets. In equilibrium, the price of such claims is equal to the net present value of the underlying assets. As a result, no profits exist. As long there is no equilibrium, entrepreneurs can buy and bid up the prices of the underlying assets or financial entrepreneurs can sell and bid down the prices of the claims on the underlying assets. Conversely, underlying assets can be worth *more* than the claims on the underlying (productive) assets; entrepreneur can sell and bid down the prices of the underlying assets or financial entrepreneurs can buy and bid up the prices of the claims on the underlying assets. We can call this economic law *the first fundamental law of capitalism*:

$$K = \sum_{t=0}^{n} \frac{CF_t}{(1 + r_t)^t}$$

This implies that capital per definition *cannot* be an input, as is the case in the conventional neoclassical production function ($y = K, L$). Rather, it is an intellectual conception of the goods that *are* inputs. If society invests in more durable consumer goods (consumer goods with longer effective durations), it is in effect *accumulating* capital. Indeed, capital is more an outcome of *output* (of value, not physical goods) than a *direct cause* of output.

Moreover, this implies that, although strictly true, the law of diminishing marginal returns make little sense in capital theory. It is true that whenever equilibrium exists and no further

[145] For more on the term structure instead of a single rate of interest, see p. 121, p. 261 and p. 288.

changes are induced that profits tend to decrease. However, as continuous changes create new profit opportunities it would be foolish to argue that the law applies *in general*. It is not that "adding capital" (which is actually correctly expressed as "adding savings") leads to a diminishing rate of return on capital. The "amount of capital" bears no relation to the return on capital, since the return on capital is an expression of economic maladjustments rather than some type of physical return on a given amount of inputs. If entrepreneurs demand savings and invest them profitably, the *average* return on capital diminishes. Nevertheless, since the very act of entrepreneurial intervention generates new information and per definition leads to new changes, new profit opportunities arise which would *augment* the *average* return on capital (Huerta de Soto, 2010).

Capital, then, is the act of *capitalizing* income streams or yields at the given rate of interest. Since we can *capitalize* at different levels – different scales – which are subjectively identifiable by the acting entrepreneur, arbitrage opportunities arise at the different scales of capital. We can capitalize the income that a firm, for instance, is able to generate from here to perpetuity and compare that *capital* to the capital we get when we take the individual, underlying assets of the firm, and *capitalize* the individual income streams or yields of all individual assets. It is fundamentally the same concept of capital, yet at a different scale.

7.2 Capital and Savings

Savings is income that is, in the most literal sense of the term, not consumed. As it is not consumed, it remains *on* the balance sheet of the respective saver. Savings can therefore be viewed as a portfolio, made up of different *financial assets*, which are simply claims on the *productive assets* of a society. We get the same double-sided equation, our *first law of capitalism*, as in our earlier definition of capital:

$$K = \sum_{t=0}^{n} \frac{CF_t}{(1 + r_t)^t}$$

The ratio between *K* (capital) and the net present value (NPV) of the underlying productive assets can be called *q* (Tobin & Brainard, 1969). Now, the ratio (*q*) *impacts* savings and investment:

> "The effect of a high *q* [is] to encourage investment and discourage saving. If the USA were the only country in the world, and stocks were the only way companies

could raise finance, the effects would be clear-cut. There would be a reduced demand for stocks, as consumers chose to save less, and liquidated existing stock holdings in order to consume more. There would simultaneously be an increased supply of stocks, as firms issued new equities in order to fund higher investment. The only way that the two tendencies could be reconciled would be by a reduction in stock prices. Once stock prices had fallen by enough to bring *q* back down to its normal level, incentives would have done their work, as usual." (Smithers & Wright, Valuing Wall Street, 2002, p. 135)

Hence, there exists a close dynamic between the supply of savings (as determined by individuals' time preferences) and the demand for savings (as determined by the available returns on capital[146]).

Investment is the application of prior savings. It is the allocation of savings to particular ventures, assets or holdings. In contrary to various authors (e.g., Huerta de Soto, 2009), it is possible to invest in cash holdings[147]. Ultimately, if such cash holdings are held in the form of liabilities of some type of financial intermediary, there is corollary in some type of financial asset (debt or equity) or nonfinancial asset (e.g., monetary gold). There is, in principal, no difference between household savings and retained earnings, since retained earnings are simply savings to the extent that they are not returned to shareholders.

[146] As of writing, record low interest rates have often been used as a justification of historically high stock market prices. However, the unprecedented low rates of interest are to be explained through the demand side: returns on capital are low and demand is weak. However, if stocks are the present value of all discounted future cash flows of an underlying business, and if low interest rates are a result of weakness, then lower interest rates *should* be offset by a lower terminal growth rate (*g*). After all, the present value of a perpetuity equals $CF_1 / r - g$. If this is true, then lower rates of interest cannot justify higher stock prices since any increase in NPV due to a lower rate of discount (*r*) would be compensated by an equally lower terminal growth rate (*g*). John Hussman, for instance, writes: "It is essential to recognize that *if interest rates are lower because likely future growth in deliverable cash flows is also lower, then no valuation premium is justified at all*." (retrieved from Hussmanfunds.com)
[147] We will elaborate on the erroneous distinction between "saving" and "hoarding" below, on p. 345.

Hence, money and saving are intrinsically connected[148]. In fact, under a less sophisticated classic gold standard, the mere hoarding of gold coins (or "outside money"[149]) is different from what many economists (e.g., Hoppe, 1994) consider an act of saving, that is, postponing present consumption. Nonetheless, if we analyze the hoarding of gold from the standpoint of economic efficacy, the inescapable conclusion is that the hoarding of gold is an extremely inefficient way of saving and putting – temporarily – resources at the disposal of entrepreneurs.

Let us compare the consequences of saving by hoarding in comparison to saving through financial intermediaries. In the former, the demand for (outside) money increases, which leads to an increase in the price of the monetary unit (for clarity, we can assume that the hoarding occurs in gold). Since the gold is hoarded, it is not spent on other consumption or production goods. Therefore, all other things equal, the prices of consumption and production goods go down. This is equal to an increase in the *real* rate of interest. The act of hoarding a nonfinancial monetary asset like gold would lead, after a gradual and prolonged adjustment period, to a fall in the general price level. As we have outlined above, such a fall in the general price level, in essence, is a way of transmitting to society that resources have been saved, not consumed and "freed up". In fact, businesses can now afford spending more on production goods than before, which allows them to expand their business. Moreover, and more importantly, *real interest rates drop*, which creates an incentive to increase (new) investment in more capital-intensive projects with longer durations (or cash flows pushed out further in the future). The effects – despite a delay and an inefficient allocation – are equal to other, more efficient ways of transmitting savings or converting savings into investment: the value of capital tends to increase (as the *real* rate of interest goes down) while present spending is reduced.

If this is the case, any increase in cash balances (either nonfinancial or financial liquid assets) represents genuine savings just as much as a *direct* investment in corporate bonds, equity or a long-term time deposit. And, as we have explained above, savings are more efficiently

---

[148] There is substantial non-monetary saving. The hoarding of resources or consumption goods, or commodities, are ways of non-monetary saving.

[149] Among other authors, Selgin (1988) refers to base money as "outside money"; currency that can be completely withdrawn from the financial sector and, metaphorically, stuffed under the matrass.

converted into investment through financial intermediaries, rather than through changes in the general price level that temporarily upset the entire economic structure.

The “ideal type” related to saving, to which we have often referred as “saver,” is the capitalist. As Schumpeter (1954) nicely sum up: “No doubt the entrepreneur was being distinguished from the capitalist, and his profit from interest, with ever-increasing clearness as time went on.” (p. 860).

7.3 Capital and Interest

The link between interest and capital has often been misunderstood, as we have seen in our extensive literature review. Interest is far from equal to the (net) return on capital. The net return on capital is explained by maladjustments between prices that represent (latent) profit opportunities. Interest is, instead, the market price of time. As Spitznagel (2011) expresses it: “[I]nterest as the cost of time has an entirely different meaning; it is the inherent price one must pay to access capital sooner rather than later, which in turn becomes the threshold for determining one’s return and the prudence of making an investment.” (p. 89). Interest rates do not directly reflect time preference (as strawberry prices do not directly reflect the subjective valuations of consumer preferences or marginal utility of strawberries), but the *opportunity cost* of consuming in the present versus consuming at some future point in time.

As a consequence, theories that explain the rate of interest *as if* interest is a phenomenon exclusively dependent upon subjective time preferences are mistaken. Kirzner (1960), for instance, argues that: “(…) interest expresses the universal (“categorial”) phenomenon of time preference and will therefore inevitably emerge also in a pure exchange economy without production.” (p. 53). Kirzner’s reference to a “pure exchange” economy “without production” is paradoxical to say the least. Without production, nothing can be exchanged. Kirzner (1976) appears to forget Mises’s (1949) insight that production is *not* something physical and that, our previously elaborated, concept of capital *includes* human capital. If the return on capital is across the board below zero, no demand for savings will arise or, in other words, no one would possess the low time preference that is necessary to produce the required asymmetry for a market rate of interest to emerge. One of the issues with Kirzner’s (1976) and Mises’s (1949) statement, is that the theory explains the *origin* of credit (the intertemporal exchange of resources or capital), but not the *cause* of a large majority of the present-day demand for savings or present resources. That cause consists precisely of the rate of return entrepreneurs expect to make when demanding funds.

Moreover, this implies that interest rates and discount rates do not directly *reflect time preference*, just as other prices do not directly *reflect subjective valuations*. Market interest rates are rather the opportunity cost of consuming in the present or at some future point in time. Kirzner (1976) is wrong: "Interest is not the specific income derived from using capital goods; nor is it 'the price paid for the services of capital.' Instead, interest expresses the universal ("categorial") phenomenon of time preference and will therefore inevitably emerge also in a pure exchange economy without production." (p. 53)"

We should therefore reject any reference to a Wickselian "natural rate of interest" (e.g., Wicksell, 1898; Mises, 1949; Hayek, 1932). A natural rate of interest, which per definition can never be observed, lacks any real-world relevance and is a theoretical dead-end.

Moreover, as Gunning (2004) correctly argues, time preference (that is, individuals distinguish between sooner or later) has often been confused with "positive" time preference (that is, consumers *always* prefer sooner rather than later), for instance by Caplan (2005). Indeed, returning to Kirzner's earlier take on the "universal" phenomenon of interest, if consumers would *always* prefer sooner rather than later no asymmetry in time preferences would exist and thus no rate of interest could possibly arise. Here we notice the incoherence in some of the definitions.

Hence, the *rate of interest* finds its origin in, on the supply side, time or saving preference and, on the demand side, the expected number of (latent) profit opportunities, that is, expected returns. Some define the latter as demand by entrepreneurs (e.g., Mises, 1949), others would include borrowing for consumption purposes. Both are, however, different sides of the same coin. Any individual (or any individual household) can only trade equity (net worth) for debt. Hence, any borrowing for consumption purposes is limited by (a) previous borrowing and (b) the market value of their human capital (the NPV of future wages). A high time-preference individual brings its future earnings to the present at a cost. If the net present value (NPV) of an individual's assets (including his or her human capital) would equal the amount of previous borrowing, no margin exists to extend more credit[150].

Now, more importantly, there is no single rate of interest, but various rates of interest for various maturities. As we have seen in our literature review, the naïve expectations hypothesis of the term structure of interest (which, briefly summarized, refers to the idea that long-term rates of interest are merely expected future short-term interest rates and inflation premia) should be rejected. As we will see below, when we discuss the portfolio approach to the demand for money (and, more broadly, the demand for financial assets), we will see that the *segmented market hypothesis* (or the weaker *preferred habitat hypothesis*) of interest rates aligns better with a priori theoretical reasoning and empirical observation. Hence, capitalists and thus capitalist-entrepreneurs (due to the risk of reflux or blow-up) have a preferred range of maturities (or *duration*) and will only prefer different maturities *if* the reward is sufficiently high[151]. Hence, any given chosen maturity **is a direct expression of time preference**.

Such arbitrage between maturities (*habitats*) should ultimately come from capitalists. The reason is simple: a capitalist-entrepreneur (that is, a financial intermediary) cannot directly arbitrage the yield curve by using short-term funding and investing in long-term credit, since his level of liquidity would not suffice and would eventually eliminate him from the market when a run on liquidity occurs[152]. He is only able to offer higher yields to capitalists (savers) at certain maturities to entice them to opt for longer maturities rather than shorter, more liquid

[150] Individuals generally cannot sell a share of their equity, not since the abolishment of slavery at least. The recent creation of "Human Capital Contracts" (HCCs) are equity-like but structured as debt.

[151] Sufficiently high is entirely subjective, as a rather ironic historical episode proves. Long-Term Capital Management (LTCM) tried to arbitrage a minimal 5 basis point spread between the 29-year US Treasury and 30-year US Treasury. However, when the Russian government defaulted, there was a run on 30-year US Treasuries, pushing the spread up to 15 basis points. LTCM ended up losing $215 million on this (leveraged) trade.

[152] We will see later that such a run on liquidity is inevitable if arbitrage of the yield curve originates from the capitalist-entrepreneur (i.e., intermediaries), rather than the capitalist (i.e., saver). See p. 406.

maturities and thus to reduce the intermediation spread by offering higher yields to savers on longer maturities[153].

A change in the rate of interest is directly reflected in financial asset prices, but not directly in prices of the underlying productive assets. Which not only encourages entrepreneurs, as Cachanosky & Lewin (2014) explain, to embark upon new investment projects with longer durations that are more capital-intensive, but also leads to changes in financial asset prices (e.g., bond and equity prices). What we *can* establish, thus, is the link between the *duration* of the (combination of the) underlying assets and the maturity of the liabilities (financial assets) that finance the underlying productive assets.

Let us illustrate how this dynamic affects the structure of production of a market economy. Let us assume that a change in interest rates occurs. Long-term interest rates go down as a result of a greater supply of savings. This change in interest rates opens up arbitrage opportunities in bonds and stocks. Capitalist-entrepreneurs (colloquially called traders in financial markets) buy bonds with higher (nominal) interest rates than the current interest rate and are willing to buy the bond at a discount because the total return is still higher. Interest rates on ten-year bonds drop and bond prices in the secondary market go up. On the other hand, stock prices of capital-intensive companies, all other things equal, are also bid up to reflect the new interest rate level. This produces two effects: (a) it incentivizes businesses to issue long-term bonds in the bond market, (b) allows businesses that are to a greater degree affected by the lower interest rate to expand relative to businesses that are less affected by the lower interest rate. As the stock prices of the former are bid up to a greater degree than of the latter, it is easier for them to issue new shares to obtain financing or borrow against more favorable conditions.

7.4 Capital, Consumption and Liquidity

What the field of marketing has already discovered is largely neglected by economists (Vargo & Lusch, 2004; Xie, Bagozzi, & Troye, 2008). Production in many cases coincides with

[153] Another way is, of course, contractually. Demand deposits or short-term credit instruments might contain “liquidity” clauses than can be evoked whenever a run on liquidity occurs. This allows the issuer to (temporarily) suspend payment and thus, in effect, align its *de facto* asset maturities and its liability maturities. This is what free bankers propose (e.g., Selgin, 1988), but seems a rather counterproductive way of aligning the maturities of savings with the maturities of investment.

consumption, a phenomenon that marketing scholars label "prosumption", a *portmanteau* of *pro*duction and con*sumption*. Many services are characterized by this interactive consumption: the fact that the consumer making part of the production process is a necessary condition to enable the actual consumption of the good. For instance, the services provided by the barber requires you to be physically present, forming part of the production process, but it also refers to more complex forms of prosumption that are nowadays becoming increasingly widespread due to the evolution of the Internet.

Production is a far broader phenomenon than some physical output that ends up in the hands of what economists designate “the consumer.” It ignores the fact that every individual is a producer and a consumer or, in other words, that the *consumer* is nothing more than an economic role. As Mises (1949) reminds us: “Production is not something physical, natural, and external; it is a spiritual and intellectual phenomenon.” (p. 141). In fact, consumption is a concept that only makes sense from the perspective of the individual. Whenever we purchase a house, we have not "consumed" the house. One "consumes" the house by living in it, by "extracting" its services over time. One who buys a car but fails to use it is not consuming. He merely assumed ownership of the car. Ownership, however, is not consumption.

This recognition is clarified when we consider the existence of stocks or inventories. Many economists have attempted to distinguish stocks of goods from capital and consumption. This, however, implies a material or historical conception of a particular good. According to the economist, an inventory is not consumption because it is not consumed, which is correct, nor production because it is not used in any productive process, which is incorrect. The error which is implicit in this reasoning is obvious. It insinuates that production is a physical phenomenon, rather than a spiritual phenomenon (Mises, 1949). As argued before, ownership does not equal consumption. An inventory produces a certain service that is highly valued by the acting individual, equivalent to a house rendering services over time. It is not the physical good that we value, but the good in given circumstances at a given moment in time and the benefits it conveys to us in that instance. The product x at time a is not the same as product x at time b. They refer to different economic goods.

Consumption only attains theoretical consistency when considered from the point of view of the individual. A good becomes a consumer good when it is viewed as such by the acting individual. Hence, a consumer good is an economic good that is utilized as a direct means to reach a goal, i.e., to alleviate a felt uneasiness, as perceived by the individual. A consumer good can also be called a good of the first order (Menger, 1871). In that case a production good becomes a good of a higher order. Consequently, a higher-order good is an indirect means to achieve individual ends. This redefinition allows us to make apparent a previously hidden factor of time. All production is made over time and passes through certain productive stages, until it reaches its final stage and becomes a good of the first order (Mises, 1949). We should note, however, that these stages are not given. They have no definable start nor end and we cannot discover the input nor output of a particular stage. However, they serve an important role as an analytical tool.

As way of visualizing this subject view of consumption, we assume that the *contractual rights* of a good are simply transferred to a final consumer. An ordinary wage earner could have a balance sheet, even though it is not customary to articulate one, as the following:

Figure 50: Simplified balance sheet of an ordinary consumer c *at* t *= 0*

Assets		**Liabilities + equity**	
Human capital	1,000,000	Credit card loan	2,000
House	400,000		
Car	20,000	**Net worth (capital)**	**1,419,000**
Checkable bank deposit	1,000		

Where:

Human capital = the net present value (NPV) of all *expected* future wages
House = the acquisition cost of a house
Car = the acquisition cost of a car
Checkable bank deposit = units deposited in a checkable (liquid) bank account
Credit card loan = outstanding debt owed to a bank
Net worth = assets minus liabilities

At $t = 0$, no consumption has occurred. The basic idea is that consumers consume their assets gradually over time. Hence, our definition does include stocks of consumer goods, both durable and nondurable. We can also see the logic behind this reasoning due to the consequences of our model. If we buy a year worth of canned tuna in bulk, we are “saving” and not consuming. This is evidenced by the consequent saving we achieve by not redirecting part of our income to the purchase of a week worth of canned tuna. If we would not have a stock of canned tuna, we would have to dedicate a greater portion of our income to the purchase of canned tuna to be consumed. Hence, consumer decisions with regard to the durability of the goods they purchase, as well as the hoarding of consumables, impact the market of intertemporal exchange through their secondary consequences. The hoarding of canned tuna in $t = 1$ is, in a certain sense, a speculative act by the consumer (both with regard to the goods, their present and future prices, and his own preferences). This speculative act could lead to higher savings elsewhere in $t = 2$, $t = 3$ and $t = 4$. If we would simply omit the hoarding of consumer goods as savings, we could ignore important economic flows that contribute to the formation and maintenance of capital.

Curiously, the consumer would incur a loss if he bought canned tuna in bulk for a year at $1/can, only to find out that the price of a can of tuna falls to $0.50/can. This fall in canned tuna prices would have increased his remaining purchasing power *ex post*, yet this is not the case. On the other end of the spectrum, a producer sold more cans at $1/can and thus benefits from the entrepreneurial error of the consumer.

This allows for a conceptual foundation behind justifying education spending (primarily public) as a present cost and its ability to augment the net present value of a student's human capital[154]. Indeed, if we assume an active working life of 50 years, a rate of interest of 8%, and our one million net present value of future labor (that is, human capital), then we get an annual wage of approximately $21,321 or divided by twelve months a monthly wage of

[154] This would avoid public "malinvestment" in education, especially in studies that are net losses to society since the skills that are being educated have no economic value in real life. The most unusual college majors in the US and UK, for instance, include majors such as "Bagpiping" (Carnegie Mellon University), "Beatles, Popular Music and Society" (Liverpool Hope University, UK), "Bowling Industry Management and Technology" (Vincennes University) and "Surf Science and Technology" (Cornwall College, UK). Indeed, in many European public educational systems, we can observe an alarming rise in "fun studies," as college tuition is subsidized or even free. Such studies do little in raising post-education income (and are therefore a net impairment of human capital), but do carry a significant cost over the education's timespan. Moreover, the issue of study period extensions has become an issue since education in most (European) nations is subsidized. Students, as a rule, take various years longer to complete the study than its intended duration. By comparing cost with increases in human capital, there is a basis to reject extensions on societal grounds. Students would base their choice of study on the net economic value of a study. Note also how the Solow-Swan growth model assumes "human capital" as a single homogenous mass. Therefore, *any* investment in education is positive by definition. This framework challenges this assumption.

$1,777. If we wish to know whether student loan debts of $100,000 to $200,000, in case of the U.S., are justified investments, we would have to compare the NPV of our human capital in absence of the study to having completed the study. In most cases, such educational costs and student debts are net impairments to human capital. In some cases, education would have a positive return on human capital.

Liquidity is the cost an individual incurs when he attempts to convert assets (capital) into present consumption. By converting assets into an intermediate good (a so-called medium of exchange), he can easily obtain what he wants. However, this definition of liquidity does not *depend* on money and is inherently subjective of nature. If an individual has an incredible appetite for bananas, liquidity would equal bananas. However, since it would be difficult to sell in the context of a modern-market most (financial) assets to a banana producer who happens to have an excess inventory of bananas, in general liquidity is the most accepted and least volatile (in terms of value) good over time. Monetary assets such as a demand deposit are not per definition more liquid than, for instance, consumer goods, but they give a greater degree of *optionality*.

Liquidation is derived from liquidity and simply is "the act of exchange of an asset of lesser liquidity with a more liquid one." A more liquid asset would imply, in terms of exchange, being nearer to the subjective end of a respective consumer. Liquidation, in case of bankruptcy, can therefore not refer to "distributing" a company's assets, but rather to distributing the proceeds of the sales of a company's assets. Liquidity is not so much about marketability in the sense of being able to exchange it for other goods, but rather to "liquidate", which would ultimately refer to "consume." We "liquidate" our homes by consuming the services it renders (and the more durable our home, the longer it takes to liquidate it). We "liquidate" our stack of apples by eating them. It would be a stretch to say that possessing apples, which are less "marketable" than cash, is less liquid in the eyes of a consumer who craves for apples. We "liquidate" the human capital of our barber by going for a haircut. Such "liquidation" of stocks of goods can be visualized through a simplified balance sheet:

Balance sheet of consumer X at t = 0

Assets	**Liabilities**
House 90	-

Balance sheet of consumer X at t = 1

Assets	**Liabilities**
House 81	-

Apples	**Net worth**
10	**100**

Apples	**Net worth**
8	**89**

Figure 51: Durable goods (such as houses) and hoarded stockpiles of consumer goods (such as apples) are "liquidated" over time, that is, consumed.[155]

Palyi (1936) defined "perfect liquidity" as a situation in which, "(…) for any length of time, all financial obligations are fulfilled **without net liquidation of capital**." (p. 5) [emphasis mine]. Palyi's definition aligns with our broader definition above. If a consumer is planning to spend part of his savings on a luxury doll in one year from now, then this consumer has to make sure he saves at a maturity of one year and that, at repayment, the luxury doll is up for sale. If, on the contrary, our doll aficionado saves at a ten-year maturity and the (capitalist-)entrepreneur subsequently invests in underlying productive assets with approximately the same duration profile, a net liquidation of capital will have to occur to allow the consumer to *liquidate* and *consume* his savings. Moreover, the doll might not have been produced, since entrepreneurs were convinced that consumers were saving at a ten-year maturity. Instead, they – for example – began building a giant doll factory that spits out high-quality dolls at twice the rate as before but has not come to completion at the end of year one.

Therefore, the preferred maturities of consumers (that is, their time preferences) should be aligned with the effective *duration* of producers' investments. Intertemporal coordination between consumption and capital (the value of production goods in terms of future consumption) is of key importance. When time preferences of consumers (i.e., the maturity of their expected future consumption) are aligned with the maturities of producers (i.e., the maturity of their expected future production), an economy is *structurally* liquid. When both sets of maturities are misaligned, an economy is *structurally* illiquid. This is perhaps best expressed by Howden & Bagus (2010): "There is a term structure of savings and a subsequent term structure of investing that align, optimally, with consumers' plans." (p. 65)

Capital, consumption and liquidity are thus intimately related to each other. Liquidity ultimately refers to approaching consumption, while consumption is always of services, not

[155] The physical decay of such (perishable) consumer goods is, for our purposes, equal to consuming them, unless someone has a (valuable) use for perished goods and is thus willing to pay for them ("one man's trash is another man's treasure"). In that case, decay leads to a partial but not complete loss (as consumption, in a strict accounting sense, is).

physical goods. Taking ownership of an alleged “consumers’ good” does not equal consuming. Consuming the services that a good provides, is consuming (Vargo & Lusch, 2004). Durable consumer goods, such as cars and houses, are thus consumed as we consume their “value” over time.

7.5 Capital and Money

7.5.1 The Relationship Between Capital and Money

Not only is capital expressed and estimated in money terms, it lies at the opposite side of financial assets. **Any productive asset[156] finds its corollary in a financial asset[157]**. And part of a society’s financial asset base consists of financial assets with monetary characteristics (checkable, transferable or any equivalent).

Thus, any *capital* held, is held through some type of financial asset. If an individual owns capital (that is, in this case, legal title to underlying assets and thus part of its *net worth*), he either owns equity or debt instruments[158] in some underlying asset(s). Put differently, part of a society’s investment depends on the amount of bank money households and businesses hold (the other part depends on the amount of other non-bank financial assets that households and businesses hold[159]).

[156] Productive assets are a universal category: even government debts are part of these productive assets. Government debt, essentially, commands a market price as a claim on future taxes on human capital.

[157] Underlying (productive) assets purchased with retained earnings also have a logical counterpart: the net worth of the company.

[158] Of course, there exist many hybrids between debt and equity. The greyest between the black and white of debt and equity are perhaps the debts convertible into equity. The practical and theoretical distinction between debt and equity is, however, undeniable.

[159] The dichotomy between households and businesses when it comes to “holding financial assets” is largely illusory: if a business holds financial assets, ultimately, its shareholders will be households. We could, therefore, say that only households hold financial assets, since in this case, by holding equity of a business that, subsequently, holds a portfolio of financial assets, households *indirectly* hold these very same financial assets. We sometimes get confused by questions of (indirect) ownership and intermediation in the broader sense of the word.

Hence, the backing theory shows how the value of financial assets (among which means of payment) depends on the assets (capital) backing the (our q, or $K = NPV$, continues to apply). On the contrary, a liquidity run (reflux) would occur which would force down the value of the liabilities issued to the value of the underlying assets (we will call this *banking q*).

Let us assume, for instance, that two banks that issued ten equal-amount deposits (worth 10 each) backed by different assets (loans) with equal maturity, yet Bank X's loans has an adjustable rate whereas Bank Y issued a loan with a fixed rate[160]:

Bank X at $t = 0$

Assets	**(4%)**	**Liabilities**	**(3%)**
Loan to A	100	Deposits	100
Cash	5	**Net worth**	**5**

Bank Y at $t = 0$

Assets	**(4%)**	**Liabilities**	**(3%)**
Loan to B	100	Deposits	100
Cash	5	**Net worth**	**5**

Figure 52: Example balance sheet of bank X and Y (1)

Now, let us assume that the (market) rate of interest changes. This makes Bank X more profitable than Bank Y, since Bank X is now earning 8% over its interest-bearing assets, whereas Bank Y still earns 4%. Moreover, the market value of Bank Y's Loan to B drops to the extent that market interest rates exceed the contractual interest rate of the loan (in this case, the price falls from 100 to 93 to reflect the change in interest rates):

Bank X at $t = 1$

Assets	**(8%)**	**Liabilities**	**(3%)**
Loan to A	100	Deposits	100
Cash	5	**Net worth**	**5**

Bank Y at $t = 1$

Assets	**(4%)**	**Liabilities**	**(3%)**
Loan to B	93	Deposits	100
Cash	5	**Net worth**	**-2**

Figure 53: Example balance sheet of bank X and Y (2)

This allows Bank X to offer higher interest rates to its depositors, luring depositors away from Bank Y. For example, Bank X raises its rate from 3% to 4%, which means it remains more profitable compared to Bank Y (it reduces its net interest margin from 5% to 4%,

[160] This phenomenon can easily be observed in the real world: take the troubled German bank Deutsche Bank with a comparatively weak asset portfolio with low returns. Deutsche Bank offered a mere 0.01% yield on a savings account. Comparable rates on savings accounts of rival banks ranged from 0.75% (ING-DiBa AG) to 0.03% (Targobank AG) with an average of 0.04% per annum (substantially higher than Deutsche Bank's offer).

whereas Bank Y has a net interest margin of 1%). As a result, a *reflux* (loss of liquidity) occurs since depositors move their deposits, on the margin, to Bank X, since Bank's yields are higher and its assets of higher quality (assuming zero risk). This implies that, for Bank Y, essentially the value of the assets that back the deposits it issued no longer suffices if the reflux would include *all* (or a majority of) deposits (if it sells its assets, it will raise 93 in cash for 100 in deposits). Moreover, there is an incentive to redeem sooner rather than later, since if redemption is not according to the underlying net asset value but a nominal value (in this case my deposit equal to 1o), the loss in backing will be shouldered by the remaining deposit holders or the deposit holders who try to redeem at par later rather than sooner:

Bank X at $t = 2$

Assets	**(8%)**	**Liabilities**	**(4%)**
Loan to A	100	Deposits	100
Loan to C	+20		+20
Cash	5	**Net worth**	**5**

Bank Y at $t = 2$

Assets	**(4%)**	**Liabilities**	**(3%)**
Loan to B	74	Deposits	80
Cash	4	**Net worth**	**-2**

Figure 54: Example balance sheet of bank X and Y (3)

Bank Y has several options: (1) it could wait until its loan to B reached maturity and is repaid in full, (2) it could stop redeeming at par, or (3) it could issue new shares and dilute its current shareholders.

Now, the first option is problematic if repayment is in 30 years, rather than in a week or a month (in calculating the price of Bank Y's assets we have assumed a two-year maturity). Here we return, again, to a critique on modern free banking theory (e.g., Selgin, 1988; White, 2012). Temporary suspension of payment only makes sense *if* the underlying loans are repaid in full in a more or less foreseeable future (hence, not in 30 years) or *if* interest rates decline to their prior level or lower in a more or less foreseeable future. The value and quality of the underlying assets of a financial intermediary is a *key determinant* in being able to redeem at par or face a reflux without trouble. In this case, material losses will be largely avoided. At most, the shareholders and/or some deposit holders will lose some type of opportunity cost, yet will not lose any principal.

The second option is a possibility. In this case, losses will fall entirely on the remaining deposit holders (in our example, the two deposit holders that moved 20 to Bank X will not suffer any material loss). Historically, the market price of bank deposits would fluctuate (and seldom traded at par), effectively taking into account the value of a bank's underlying assets.

The third option is completely dependent upon some third party's liquidity. That is, new shareholders will have to be liquid to be able to invest in Bank Y. Hence, we see yet again the difference between *individual liquidity* and *systemic liquidity*. If some potential equity investor possesses liquid assets, the third option is a possibility. In this case, the losses would fall exclusively on Bank Y's existing shareholders.

In sum, there are arbitrage forces that will move the value of the underlying assets (the backing) toward the value of the corresponding liabilities (debt and equity, among which, importantly, demand deposits).

At any rate, what this example shows is that, **given the law of reflux, the value of Bank Y's (monetary) liabilities depends on the value of its assets**.

This is, of course, completely accepted when it comes to the value of non-monetary financial assets. If a company's assets are worth $50 million (given $q = 1$), and a million shares were issued, each share would be worth $50. Now, if this company would issue another million shares to buy another $50 million worth of assets, there would be 2 million shares in circulation backed by a total of $100 million in assets. Individual shares thus continue to be worth $50, yet the quantity of shares has doubled.

What would happen if a bank has a hole in its balance sheet in the sense that a third of its assets are not providing *any* return? This simply spells disaster for the bank in question. They will be unable to compete for funding (depositors) without incurring losses. Bank profitability is, thus, a measure of underlying asset quality, although high bank profitability does not necessarily equal high underlying asset quality. This is due to the nonlinear loss function; profitability (benefits) can be temporarily augmented by increasing the risk of future losses[161].

Losses *were* incurred in the 2008 crisis as a result of loss of backing, although largely indirectly since the losses were absorbed by governments and public budgets.

Moreover, **reflux occurs whenever a channel exists that allows liabilities to be extinguished**. We will continue by listing the possible channels of reflux:

(1) Liabilities are directly redeemable or convertible into the underlying assets

This is the modern-day reality for commercial banks (and other financial intermediaries, such as mutual funds[162]). Demand deposits are directly convertible into cash (currency issued by the central bank) or can be used in wire transfers to competing banks (which results in an adverse compensation and, likewise, a loss of reserves/assets). This is what Fullarton (1845) had in mind when he first formulated the law of reflux.

(2) Assets have a (fixed) maturity and/or are sold

This is the modern-day reality for many central banks (but generally applies to any banking institution, including commercial banks). If a bank, for instance the Federal Reserve, owns 1-year Treasuries, it will experience a reflux at date of repayment. The Treasury has a bank account at the Federal Reserve and, at maturity, repays its debt by reducing its balance at the Fed:

[161] A point in case is the Icesave bank collapse. Many Dutch and British savers moved their savings to the Icelandic bank Landsbanki (Icesave) since it was offering a higher rate of interest on savings accounts. However, Landsbanki was attracting depositors by offering high rates of interest while simultaneously lowering the quality of its loan portfolio. As a curious fact, Landsbanki held an astonishingly large equity portfolio that experienced large positive capital gains up till 2007. For more, see Aliber & Zoega (2011).

[162] Some mutual funds redeem in cash or in kind. Whenever a mutual fund is faced with a reflux (for instance, through redemption requests), it can sell part of its assets and redeem shares for cash. Alternatively, it can simply exchange shares directly for the underlying assets (redemptions in kind).

Federal Reserve at $t = 0$

Assets		**Liabilities**	
US Treasuries	100	Deposits	115
Gold	20	**Net worth**	**5**

US Treasury at $t = 0$

Assets		**Liabilities**	
Deposit	115	US Treasuries	100
		Net worth	**15**

Federal Reserve at $t = 1$

Assets		**Liabilities**	
US Treasuries	0	Deposits	15
Gold	20	**Net worth**	**5**

US Treasury at $t = 1$

Assets		**Liabilities**	
Deposit	15	US Treasuries	0
		Net worth	**15**

Figure 55: Example balance sheet of Federal Reserve and US Treasury

If a central bank and its government are consolidated, "taxes receivable" are both backing and a channel/source of reflux. If both are not consolidated, and if a central bank holds domestic government bonds, then taxes back government debt, and government debt backs the central bank's liabilities (hence, indirectly, taxes back central bank liabilities).

Moreover, selling assets is a way of directly backing the value of issued liabilities and can especially be observed in the case of central banks: central banks can decide to sell assets and retire liabilities in order to increase the market price of its liabilities (and thus “fight” inflation or “defend” an exchange rate[163] against foreign moneys). We will see below that a loss in the market price of an issuer’s liabilities is also a channel/source of reflux: if the issuer does not agree with the market and pretends that the value of the underlying assets are adequate to back the corresponding (nominal) amount of issued liabilities, then it can decide to sell assets and retire liabilities, thus pushing up the market price of its liabilities (if, however, the underlying assets do not suffice, the issuer will only *reduce* its effective backing).

(3) Liabilities have a (fixed) maturity

This is not the case for modern-day demand deposits, but is the case of bond financing. When a liability (debt) reaches maturity, it has to be paid back in full to its creditor. *If*, for instance, central bank deposits would have a fixed term, repayment would either lead to a loss of assets or to an increase in *callable liabilities* (such as demand deposits, in which

[163] This is, of course, what happens when an exchange rate is “pegged” to another exchange rate.

case channels one and two would function).

(4) The market price of the issuer's liabilities falls

One way for a reflux to occur is through the (market) value of its liabilities. This is generally the case when direct conversion or redemption is restricted (for example in the case of central bank currency). Take the following example (with 100 deposits equaling a nominal $1):

Bank X at $t = 0$

Assets		**Liabilities**	
Loans	100	Deposits	100 x 1
Cash	5	**Net worth**	**5**

Bank X at $t = 1$

Assets		**Liabilities**	
Loans	80	Deposits	100 x 0.85
Cash	5	**Net worth**	**0**

Figure 56: Example balance sheet of bank X at t = 0 and t = 1 (1)

In this case, we assume there was a 20% loss in the value of Bank X's loan portfolio (for any reason, either default or a decline in price due to a change in the rate of interest). The reflux in this case would occur through the market price of Bank X's liabilities (deposits). Instead of being worth $1 each, they will now be discounted in the market to a mere $0.85. This amounts to a 15% loss of purchasing power.

Now what is the source or trigger of such a decline in the market price of an issuer's liability *if* no conversion in underlying assets exists?

The weak explanation is, of course, a similar (long-term) arbitrage relationship as our q (see Section IV). Others could take the same underlying assets and issue liabilities at, for instance, a 10% premium to the "overvalued" liability. This would shift demand from the original issuer to the new issuer, lowering demand for the liabilities of the original issuer. This is what, generally, happens on foreign exchange markets. However, it is only one part of the explanation and leaves out the, by far, most important factor in triggering such arbitrage between issuers.

The strong explanation is rather the rate of interest one is able to pay *given* its underlying assets. *If* Bank X was paying 3.5% on its interest-bearing liabilities, while earning 4% on its interest-bearing assets, it would have a positive interest margin and be able to cover, for instance, its operational expenses. However, if suddenly 20% of its interest-bearing assets default, its return on (interest-bearing) assets is reduced to 3.2%. The bank is now unprofitable, and either forced to run losses or pay less in interest on its liabilities. These lower yields will lead to a decreased demand for these now lower-yielding liabilities and an increased demand for higher-yielding liabilities. Since there is no direct conversion or redemption, the supply cannot adjust to the lower demand, and therefore the adjustment is made in price or, more specifically, the exchange rate. (We will further study central banks and currencies with negative equity and losses in Section IV.)

If, in contrast, the issuer has assets of sufficient quality, it can raise interest rates (and reduce its profitability and thus its dividends to government) or sell assets to reestablish the market price of its liabilities.

Indeed, units of money ought to be considered somewhat akin to shares in a closed-end fund. Shares of a closed-end fund are not redeemable or convertible into the fund's underlying assets. Hence, we could say that there exists a regular supply and demand for the fund's shares. However, thanks to arbitrage, whenever demand increases and prices are bid up, arbitrage springs up (on the margin, investors shy away from closed-end funds where an unjustified premium to net asset value or NAV – the underlying assets – exists, besides the possibility of rival issuers to issue shares backed by similar assets at lower prices. When the price of a closed-end fund falls below the market price of the underlying assets, there is an incentive to *buy* the closed-end fund, even on the part of the issuer. Of course, such arbitrage is not instantaneous and depends on the supply of savings (that is, capitalists) and entrepreneurial effort. Inconvertible bank money (that is, our present-day central bank money, commercial banks still issue liabilities that are convertible into assets) is very similar to shares in such a closed-end fund, but with an added liquidity premium: banks offer payment systems with some *X* cost that allow for rapid payments and transfers. Yet fundamentally, the mechanism is similar.

In sum, capital is a tool for entrepreneurs that allows for economic calculation, that is, capital or profit/loss accounting (Gunning, 2004). Indeed, as *net worth*, capital is the sum or stock of profits and losses expressed *in terms of money* after a given period or a given number of adjustments. Not only is capital expressed in terms of money, money as liquid asset *par excellence* is part of capital: moreover, if held as balance, money is a form of *savings*, which allows for (generally short-term) investment. There is no difference in kind between saving and hoarding, but sometimes a difference in degree. Accumulating cash balances is akin to expressing *some* degree of renouncing consumption, but generally at a very short or immediate maturity which allows for high degrees of *optionality*. Investing in fixed income with long maturities (e.g., pension funds, life insurance, long-term bonds, et cetera) or in equity is, on the contrary, akin to expressing a *strong* degree of renouncing present consumption. Yet both are forms of savings; only their characteristics and maturities (or *durations*) differ. Since a majority of such short-term savings are held in modern forms of money (e.g., demand deposits), they can be simultaneously used to invest in short-term debt. Historically, such debt were real bills of exchange. However, bank money can be converted into any fundamentally liquid investment, such as short-term certificates of deposit, short-term commercial paper, lines of credit that serve the working capital needs of businesses, and/or short-term consumer credit (i.e., overdraft coverage).

7.5.2 The Fallacy of Distinguishing Between the Demand for Money and the Act of Saving

Many authors have defended the idea that, besides spending on consumption goods (earlier we have defined consumption *not* as ownership of a good ready for consumption, but rather as the act of consuming the services of the good) or production goods, a third alternative exists, which is increasing cash balances (e.g., Bagus, 2016). According to such authors, accumulating a cash balance (e.g., by means of demand deposit) or "hoarding" ought to be distinguished from *pure* savings. For instance, Hoppe (1994) writes:

> "First off, it is plainly false to say that the holding of money, i.e., the act of not spending it, is equivalent to saving (…). In fact, saving is not-consuming, and the demand for money has nothing to do with saving or not-saving. The demand for money is the unwillingness to buy or rent non-money goods — and these include consumer goods (present goods) and capital goods (future goods). Not-spending money is to purchase neither consumer goods nor investment goods. Contrary to Selgin, then, matters are as follows: Individuals may employ their monetary assets in one of three ways. They can spend them on consumer goods; they can spend them on investment; or they can keep them in the form of cash. There are no other alternatives (…). [U]nless time preference is assumed to have changed at the same time, real consumption and real investment will remain the same as before: the additional

> money demand is satisfied by reducing nominal consumption and investment spending in accordance with the same pre-existing consumption/investment proportion, driving the money prices of both consumer as well as producer goods down and leaving real consumption and investment at precisely their old levels." (pp. 72–73)

Whereas it is clear that these authors define commercial bank liabilities as money, do they hold the same position with regard to money market mutual shares? That is, do the liabilities of a money market fund (MMF) equal cash balances or investments? Distinguishing, from an economic point of view, between these two instruments appears untenable, at least from a practitioner's point of view. The arbitrary line between demanding money and saving becomes especially problematic since, on a historical side note, in the 1970s even MMFs came to replace banks in the U.S. As Glasner (1989) explains:

> *"*At first, the MMMF was just a way for small savers to avoid Regulation Q ceilings. But in 1976 Merrill Lynch introduced a Cash Management Account that merged a traditional brokerage account with an MMMF, periodically transferring any idle funds in the non-interest-bearing brokerage account into the MMMF. Customers could make payments either with a credit card provided them or by writing checks. The checks and credit-card drafts would be automatically debited against any cash balance in the account. If there were no cash in the account, shares in the MMMF would be liquidated to cover the payment. If there were not enough shares in the MMMF, Merrill would pledge securities in the customer's portfolio as collateral for a loan to cover the payment. The success Merrill Lynch enjoyed with its Cash Management Account induced other brokerage houses to offer similar accounts. Pure MMMFs began allowing shareholders to write checks against their shares. The explosive growth of checkable MMMFs virtually forced Congress to enact legislation relaxing the constraints Regulation Q had been imposing on depository institutions." (p. 168)

A cash balance at a bank or a MMF are held for essentially the same purpose: liquidity. And since what matters in economics is the subject valuations of the individual actor, rather than the pseudo-objective valuations of the economist, arguing that this position is untenable from a practical point of view is equal to arguing that this position is theoretically undefendable.

The fundamental flaw in this line of reasoning is that these authors cannot avoid admitting that, whenever someone increases his cash balance, he is reducing monetary demand for both consumption and production goods. This would lead, after a gradual and prolonged adjustment period, to a fall in the general price level. As we have outlined above, such a fall in the general price level, in essence, is a way of transmitting to society that resources have been saved, not consumed and "freed up". In fact, businesses can now afford spending more on production goods than before, which allows them to expand their business. Moreover, and more importantly, *real interest rates drop*, which creates an incentive to increase (new) investment in more capital-intensive projects with longer durations (or cash flows pushed out further in the future).

If this is the case, any increase in cash balances represents genuine savings just as much as a *direct* investment in corporate bonds, equity or a long-term time deposit. And, as we have explained above, savings are more efficiently converted into investments through financial intermediaries, rather than through changes in the general price level. A readjustment of savings and consumption through financial intermediaries would not force to a readjustment through the general price level or the purchasing power of the unit of account. *If* and *when* an increase in savings is handled by financial intermediaries instead of propping up the demand for nonfinancial monetary assets (such as gold), price adjustments effect one industry (where the monetary demand is reduced) and another (where the monetary demand is increased), but outside these industries no market participants would be affected by a change in prices.

The problem of financial intermediation and banking in particular is, as we have stated several times, not credit creation, but credit transformation. And, in fact, any increase in the (portfolio) demand for monetary liabilities, allows issuers of such monetary liabilities to increase lending without a *reflux* or a loss of assets. We will further elaborate on the connection between capital and money (or, more broadly, financial intermediation) in the next chapter.

Chapter 8: The Portfolio Approach to the Demand for Money and Financial Assets

Various authors have already proposed a portfolio approach to the demand for money and financial assets (e.g., Hicks, 1989; Tobin, 1963; Bagus & Howden, 2014; Duesenberry, 1965). In this chapter we will outline the key elements of the portfolio approach to the demand for money (and, more broadly, financial assets) and attempt to show how the portfolio approach leads to an endogenous theory of money, in which the money supply is endogenously determined by the demand for money. Any attempt to exogenously increase the supply of money (for instance, by central banks or commercial banks) triggers the law of reflux, which leads to negative consequences for the institutions that increase the supply of money *above* the demand for money in either the short run (through a liquidity run or adverse clearings) or in the long run (a drop in the market value of the liabilities of the issuer that brings it back in line with its backing).

Interestingly, it was perhaps Mises (1949) who for the first time pointed out the merits of the portfolio approach to the demand for money. Mises (1949) wrote: "The total amount of money and money-substitutes is kept by individuals and firms in their cash holdings. (...) **Each is eager to keep a certain portion of his total wealth in cash**. He gets rid of an excess of cash by increased purchases and remedies a deficiency of cash by increased sales." (p. 445)[emphasis mine]. Moreover, Mises (1949) clearly distinguishes the portfolio demand for money: "There exists a demand for media of exchange because people want to keep a store of them. Every member of a market society wants to have a definite amount of money in his pocket or box, a cash holding or cash balance of a definite height. Sometimes he wants to keep a larger cash holding, sometimes a smaller; in exceptional cases he may even renounce any cash holding. (...) Their cash holding is not merely a residuum, an unspent margin of their wealth. It is not an unintentional remainder left over after all intentiqnal acts of buying and selling have been consummated. Its amount is determined by a deliberate demand for cash." (p. 399)

Now, it is important to recognize at this time that the demand for non-interest bearing monetary assets (both financial and nonfinancial, that is for example, both demand deposits and gold) largely depends on its opportunity cost, that is, the return on interest-bearing monetary assets or, even more broadly, the return interest-bearing non-monetary assets and the return on equity-instruments. Indeed, this is where Tobin (1963) for instance defended a portfolio approach not to exclusively money, but to savings: “A more recent development in monetary economics tends to blur the sharp traditional distinctions between money and other assets and between commercial banks and other financial intermediaries; to focus on demands for and supplies of the whole spectrum of assets rather than on (…) “money”; and to regard the structure of interest rates, asset yields, and credit availabilities rather than the quantity of money as the linkage between monetary and financial institutions and policies on the one hand and real economy on the other.” (p. 4).

Hence, any individual’s savings can be visualized as a portfolio (a snapshot in time). As such, you arrive at a “term structure” of savings, in which an individual’s savings is divided into various types of (financial) assets at different maturities (Bagus & Howden, 2014). Among the very liquid, short-term maturities, you will have a given share in monetary financial assets, that is, financial media of exchange (in modern times, largely demand deposits) and a remainder in non-monetary financial assets (that is, for instance, liquid savings accounts).

In fact, a large part of the historic increase in the M1 money supply[164], across countries and over time, can be explained by the increase of saved income. Sometimes people, however, opt for an increase in cash balances, whereas at other times, people prefer to invest a larger part of their *flow* of savings (their unconsumed income) into other types of financial assets (for instance, on public stock markets or on other financial asset markets). Generally, in times of stress, the year-on-year growth rate in M1 money supply (which largely consists of currency and demand deposits) exceeds the personal saving rate, which indicates that all non-consumed income goes into liquid monetary financial assets and that non-monetary financial assets are liquidated in favor of demand deposits as well. (In 2008, for example, M1 grew approximately 15%, while the saving rate stood at 5%, which indicates a massive liquidation of non-monetary fianncial assets.)

On the very contrary, in normal times, the saving rate exceeds the growth in M1, which makes perfect sense. Since it would be rather costly to accumulate a 100% of savings into

[164] M1 money supply consists of all currency in circulation (cash) but, more importantly, all commercial bank demand (checking) deposits. Nowadays, in the U.S., about half of M1 consists of demand deposits and other types of checkable bank accounts.

non-interest bearing currency or demand deposits, capitalists tend to invest part of their new savings into interest-bearing currency.

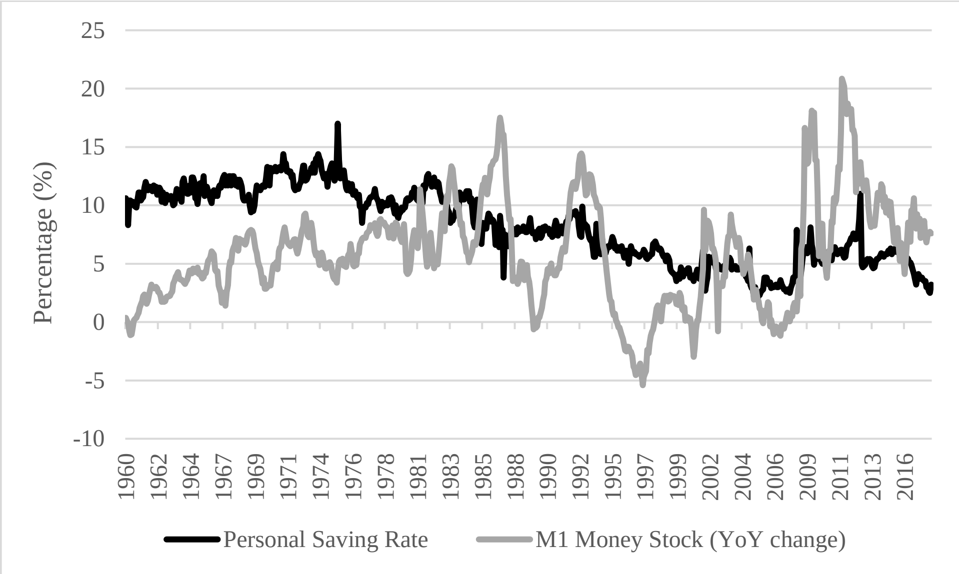


Figure 57: The personal saving rate (black curve) versus year-on-year change in M1 money stock (grey curve).

Hence, the returns on capital markets (which includes interest-bearing financial assets) determine the opportunity cost of holding M1. It is notable, as a result, that ever since 2008 (due to low or near zero interest rates on savings accounts) the opportunity cost of holding demand deposits is near zero.

The decisions of capitalists (that is, savers) hence ultimately determine the temporal composition of capital. Their preference for liquid short-term savings (in either monetary or non-monetary financial assets) determines, in absence of financial intermediaries (or maturity mismatching by financial intermediaries), determines short-term investment. Conversely, their preference for illiquid long-term savings (equities, mutual funds, hedge funds, life insurance, pension funds, etcetera) determines long-term investment. This is the essence of our portfolio approach to the demand for money and financial assets, much in line with Lachmann (1956) and Moulton (1921).

8.1 A Brief Note on Methodology and the Accounting Approach in Economics

The accounting approach, much in line with the 'Austrian' methodology of logical rigor and consistency, forces a user to be consistent in his assumptions and elaboration of a theory. In that sense, it is surprising that the accounting approach to economics, which in fact is an offspring of the earlier 'Austrian' methodological tenets (Schumpeter, 1908; Huerta de Soto, 2008), has not become more popular in recent times. Bezemer (2010) provides us with a reason 'why' the accounting approach has not become more popular:

> "If the accounting approach is theoretically and empirically well developed and predictively successful, why is it not more widely popular in academia and policy circles? Why have its methods not become part of the mainstream so that the 2007 crash was more widely anticipated (and therefore, paradoxically, perhaps avoided)? One answer, according to Pasinetti (2005), is historical and relates to the personalities of the generation of Cambridge economists working with or directly succeeding Keynes – Nicholas Kaldor, John Hicks, Joan Robinson and Richard Kahn– who apparently did not admit outsiders to their circle or sponsor their work." (p. 7).

Curiously, as we have seen earlier, Lachmann (1956) identifies his own circular flow model, which is actually, in methodological terms, an initial attempt to use the accounting approach. To Lachmann (1956), an economy can be summarized by means of three sides of an economy-wide consolidated balance sheet[165]:

1. The asset structure (consisting of the production goods – assets – on the asset side of the balance sheets of businesses)
2. The capital structure (or *control structure* as Lachmann himself calls it, consisting of the financial instruments by which the asset structure is financed).
3. The portfolio structure (consisting of the financial assets in the hands of, generally, households that finance the asset structure.

The portfolio and capital structure logically coincide. The asset and capital structure logically coincide as well. Any increase in savings, for example, would add to the portfolio structure, which would translate into additions in the capital and asset structure. There is complete consistency and coherence. Moreover, any *inconsistencies* can be easily pointed out, since their inherent to the underlying logical reasoning and assumptions.

The asset structure, following the lines of Smith (1776), consists of circulating and fixed capital. The circulating capital represents the money that changes hands in a period *t* from

[165] Unfortunately, despite ample opportunity, Lachmann (1956) did not explain his theory with the accounting rigor that the accounting approach requires.

businesses to households. Households can decide to consume in the present or postpone consumption to some given future. Any changes in the aggregate flow of savings (and, thus, the portfolio structure) affects the asset structure of producers through the decisions of financial intermediaries. Again, Lachmann's model, from an accounting perspective, allows to interpret and understand key economic relationships.

According to Bezemer (2010), this accounting approach is described by Godley and Lavoie (2007). Godley & Cripps (1983) are even more confident of their methodological approach: "The fact that money stocks and flows must satisfy accounting equalities (..) in an economy as a whole provides a fundamental law of macro-economics analogous to the principle of conservation of energy in physics." (p. 14). But, to repeat, the accounting approach is a subset of the 'logical' approach and rigor to economics which is so characteristic of the 'Austrian' school of economics (Huerta de Soto, 2010).

While the balance sheet approach (or "accounting approach in economics") resembles, in a certain way, Keynes's circular flow model, and is explicitly adopted by many post-Keynesians (e.g., Kalecki, Minsky & Tobin)[166], it has great merits. However, the accounting approach as formalized by (many) post-Keynesians lacks a key 'Austrian' ingredient: the intertemporal element of capital. **Or more specifically: balance sheet mismatches**. Although recognized in applied economics (e.g., Silvers, 1976), they are not studied from a theoretical perspective, which is a weakness of post-Keynesian thought.

In fact, Bezemer (2010) assumes, applying his proposed accounting approach, that households' net worth goes up, without noting the balance sheet intricacies of financial intermediaries that prevent them from lending, even when the value of possible collateral is up. In fact, financial intermediaries would run out of liquidity, as the earlier discussed *law of reflux* shows. We will discuss the ramifications of the law of reflux below.

8.2 Money and Financial Assets as Mediators between Savings and Capital

Indirectly, a society's savings, in terms of financial assets, are simply claims on the future productivity of a society's productive assets. Yet, *how* savings are transmitted into investment

[166] In practice, the flow of funds reports by the Bureau of Economic Analysis (BEA) are an application of such principles.

in a society's productive assets and determine a society's structure of production, is the task of a wider theory of financial intermediation (rather than a more narrow theory of banking).

The difference between money (more specifically, monetary financial assets) and other (nonmonetary) financial assets is a difference of degree, not of kind. As James Tobin (1963) wrote:

> "Neither individually nor collectively do commercial banks possess a "widow's cruse." (...) [Commercial] banks are limited in scale by the same kinds of economic processes that determine the aggregate size of other intermediaries." (p. 6).
>
> One often cited difference between commercial banks and other intermediaries must be quickly dismissed as superficial and irrelevant. This the fact that a bank can make a loan by "writing up" its deposit liabilities, while a savings and loan association, for example, cannot satisfy a mortgage borrower by crediting him with a share account. The association must transfer means of payment to the borrower; its total liabilities do not rise along with its assets. True enough, but neither do the bank's for more than a fleeting moment." (p. 7)

According to Tobin (1963), financial intermediaries – both banks and non-banks – offer assets to the saving public which, to a certain degree, constitute substitutes, with different competing yields. The saving public determines, according to its needs, how much of its portfolio (of its savings) it wishes to hold in the various assets offered by the various financial intermediaries. What limits the expansion of each intermediary, is simply the *net interest margin*: what it can possibly offer in interest on its liabilities and how much it can possibly earn on its assets (Tobin, 1963).

An efficient financial sector therefore intermediates *all* savings, so that they no longer have to be transmitted through the general price level, which implies a time lag, since adjustments are made through *real interest rates* (nominal rates do not change, but prices decline) and not *nominal interest rates* established on capital markets. In other words, financial intermediaries only monetize savings and allow them to be reinvested in other branches of the economy.

Chapter 9: The Capitalist-Entrepreneur

There already exists a rather well-developed literature on the theory of the entrepreneur and the entrepreneurial function grounded in price arbitrage (e.g., Huerta de Soto, 2010; Kirzner, 1960; Mises, 1949). I will present a brief, more extensive overview of the theory of the entrepreneur. The entrepreneur (or entrepreneurial function) is embodied by several key elements:

(1) Profit arbitrage

The central and most important tenet of the theory of entrepreneurship radiates around the fact that entrepreneurs identify (latent) profit opportunities, that is, price differences between the goods and services bought at $t = 1$ and sold at $t = 2$. Any profit opportunity represents a maladjustment of resources to consumer preferences. A pure entrepreneurial act is, hence, to arbitrage such price differentials by buying low and selling high. Time, in such a pure entrepreneurial act, is of great importance. Indeed, as far back as Cantillon (1755), it was recognized that the entrepreneur essentially has a coordinative function in aligning supply (resources) with demand (Howden, 2009). But to do so, involves *time*, specifically the time between acquiring the goods and services required in production and the selling of the final product[167]. Entrepreneurial arbitrage is impossible without time between buying and selling[168]. Therefore, the often-used distinction between *pure* arbitrage and speculation is false. Instead of a difference in *kind*, the distinction actually refers to a difference in *degree*. Spitznagel (2013) expresses this in another way: “The act of organizing production is effectively the act of lending, as inputs are paid up front in order to command product for sale much later. If the profits exceed the costs of waiting, there is an *intertemporal arbitrage* between inputs and output to be had.” (p. 114).

Mises (1949) described this entrepreneurial price arbitrage in the following way: “If all entrepreneurs were to anticipate correctly the future state of the market, there would be neither profits nor losses. The prices of all the factors of production would already today

[167] Subjective time, as Shackle (1989) and Huerta de Soto (2010) have pointed out, is different from mechanical Newtonian time, but conforms better to our concept of (Macaulay) *duration*, which we have discussed earlier. See p. 216.

[168] Even in high-frequency trading, arbitrage requires time, even if it takes as little as a split second.

be fully adjusted to tomorrow's prices of the products. In buying the factors of production the entrepreneur would have to expend (with due allowance for the difference between the prices of present goods and future goods) no less an amount than the buyers will pay him later for the product. An entrepreneur can make a profit only if he anticipates future conditions more correctly than other entrepreneurs. Then he buys the complementary factors of production at prices the sum of which is smaller than the price at which he sells the product." (p. 291). The "art" of entrepreneurship is, thus, to buy a set of production goods against "low" prices and to sell the product against "high" prices. The theory of entrepreneurship is, as a result, universal. It can be applied to spot markets, forward markets, geographical markets, financial and nonfinancial markets. The recognition of Steve Jobs that there was a price differential between the input prices of a smartphone (touchscreen, processor, memory, audio chip, battery) in 2006[169] and the price that consumers would pay for a smartphone[170] is equal to an arbitrageur identifying a (spot) price differential[171] between bananas in Europa and bananas in the U.S. (net of transport costs).

(2) *Not* a resource provider

We will see later how the entrepreneur is different from the capitalist-entrepreneur, who intermediates between capitalists (savers or resource providers) and entrepreneurs. In Huerta de Soto's (2008) words: "The exercise of entrepreneurship does not require any means. That is to say, entrepreneurship does not entail any costs and is therefore fundamentally creative. This creative aspect of entrepreneurship is embodied in its production of a type of profit which, in a sense, arises out of nothing, and which we shall therefore refer to as pure entrepreneurial profit. To derive entrepreneurial profit one needs

[169] Many of the necessary components did not exist or were not sold at viable prices earlier.

[170] In effect, Jobs speculated that by allowing any developer to develop programs ("apps") for the iPhone, consumers would find more use in their cell phones over time and would therefore be willing to pay more than for past mobile phones. Indeed, many cell phones sold at much lower prices. In fact, the iPhone specs compared unfavorably to other cell phones (such as the Nokia N95), yet its operating system (iOS) and user-friendliness was many times superior. Due to iPhone's superior iOS, Apple finally began to outsell its biggest rival Nokia in 2011.

[171] Fisher (1930) was one of the first to elaborate on the arbitrage model, including local and international spot and forward markets. See p. 141.

no prior means, but only to exercise entrepreneurship well." (p. 21)

(3) Creates new information disseminated by the price mechanism

In this sense, the Schumpeterian theory of the entrepreneur (e.g., Schumpeter, 1942) fits in with the Kirznerian theory, even though many authors consider them opposing theories (e.g., Huerta de Soto, 2008). The use of "untried technological possibilities" (as Schumpeter calls them), creates new information that lead to subsequent new maladjustments and profit opportunities. The use of such new technologies, however, perfectly fits the Kirznerian theory of the entrepreneur. Novel combinations of inputs (at their respective market prices) are sold at some expected future market price for a profit. For instance, in the development of the light bulb, many materials were tried for the filament: carbon, platinum, chromium, rhodium and finally tungsten. Even though materials such as platinum were a "technological possibility," they were unfit from an economic point of view since their prices were too prohibitive. Hence, the Schumpeterian alleged "disequilibrium" entrepreneur is just as much a profit-arbitraging entrepreneur as the Kirznerian theory of entrepreneurship presupposes.

(4) *Ex ante* expectations of profit/loss and *ex post* calculations of profit/loss "guide" entrepreneurial activity

Entrepreneurs determine their odds of success before they embark upon any venture: they want to have a reasonable chance of success. Entrepreneurs either accumulate more resources (when they realize *ex post* profits) to the degree that they are successful: a rather large winner-takes-all effect exists in increasingly global markets (Taleb, 2012).

(5) Asymmetry in risk aversion

There is no need for the entrepreneur to be rational or less risk averse than non-entrepreneurs. The fact that human beings, entrepreneurs included, are fallible (which behavioral economics often tries to underline, especially by referring to "biases") is not an argument against leaving economic coordination to entrepreneurs. In fact, the irrationality of many (Taleb, 2012) is the rationality of the market. There is no need to have a complete set of perfectly rational or perfectly risk-appraising entrepreneurs: the extreme risk aversion of one entrepreneur will, on the whole, balance against the extreme risk-seeking behavior of another entrepreneur.

Entrepreneurship is an element that only exists when profit opportunities exist. In equilibrium, no such profit opportunities exist and therefore entrepreneurs essentially cease to exist.

In our definition of the entrepreneur we deviate from the Knightian entrepreneur (e.g., Knight, 1921). In Knight's view, the entrepreneur is essentially an uncertainty bearer. As Howden (2009) writes: "[T]he Knightian view of uncertainty excluded the entrepreneur from having any sort of coordinating effect. The "fog" of the future is so thick that it is purely unmanageable. The implication is that entrepreneurial profits are as likely to develop as are losses."

Other have objected to our theory of entrepreneurship by pointing out that opportunities are often *created*, not *identified* or *discovered* (Klein & Foss, 2012). Klein & Foss (2012) claim that the Kirznerian theory wrongly pretends that opportunities have some type of objective existence. They refer to this notion as the "opportunity-discovery" bias. Nevertheless, this critique should be rejected. It is true that opportunities do not have some kind of objective existence outside of an entrepreneur's mind (here, instead, comes the entrepreneur's creativity into play). However, prices *do* have such an objective existence. In fact, historical prices are objective expressions of opposing valuations between buyers and sellers (buyers and sellers value the good they acquire more than the good they renounce). The price mechanism, which exists thanks to two immensely important social institutions (that is, money and interpersonal exchange), serves as a bridge between the subjective world (whatever occurs inside an individual's mind) and the objective world. As Huerta de Soto (2010) writes:

> "Mises's theory integrated the subjective, internal realm of individual valuations (ordinal) and the objective, external realm of estimated market prices set in monetary units (cardinal). The two realms can be bridged whenever an act of interpersonal exchange springs from the difference in parties' subjective valuations, a difference expressed in a monetary market price or historical term of trade in monetary units. **This price has a certain real, quantitative existence, and it provides the entrepreneur with valuable information for estimating the future course of events and making decisions (economic calculation)**." (p. 126) [emphasis mine]

Hence, any opportunity *has* an objective existence in the form of prices. Of course, it might turn out that there is a large difference between *ex ante* expected future prices and *ex post* realized prices, to the extent that an apparent (future) profit turns into a loss. Yet what the profit and loss mechanism effectively establishes is whether an *ex ante* anticipated opportunity was in fact an opportunity. In the context of the price mechanism, thus, profit opportunities are latent within the price system and therefore *discovered*, not *created*. Klein & Foss (2012), in a sense, deviate to the Knightian uncertainty-bearer, since entrepreneurs can only hope for *ex post* profits, rather than anticipate *ex ante* profits.

Hence, since – in contrary to Klein's and Foss's (2012) entrepreneur, our entrepreneur is essentially *resource-less*, he depends on someone providing those resources for him. This is where Lachmann's (1956) 'control structure' comes into play. What we lack, then, is a theory of how such resources *are* provided.

That theory is precisely an important part of the theory of capital. The ultimate owner of the saved-up resources (the "resource provider") is the saver (or *capitalist*). Yet, savers do not lend to or invest directly in entrepreneurs, at least in a majority of cases. What the modern-day (Kirznerian or 'Austrian') theory of the entrepreneur misses, is a theory of the *capitalist-entrepreneur* or *financial entrepreneur*. It consists of applying the same profit-seeking, opportunity/price-arbitraging entrepreneurs as many before us have already developed (e.g., Kirzner, 1996; Huerta de Soto, 2010) to the realm of financial intermediation. It includes, but is not strictly limited to banking institutions.

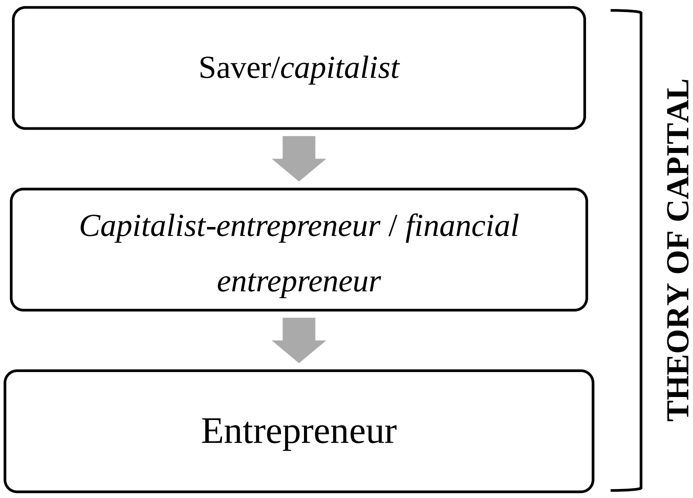


Figure 58: A diagram representing a broader theory of the entrepreneur that includes financial intermediaries as financial *entrepreneurs.*

Now, both the latter and the former are well-developed in the literature. Mises (1949) and Fisher (1930), for instance, explained the saver/capitalist side of the equation in a detailed and thorough manner. As we have seen above, the theory of the entrepreneur is also extremely well-developed (Mises, 1949; Huerta de Soto, 2010; Kirzner, 1960). What is lacking, nevertheless, is a theory of the center, a coherent theory of financial intermediation (a *financial* entrepreneur) that connects capitalists and entrepreneurs.

Now, the fact that the entrepreneur is resource-less, although in practice entrepreneurs invest part of their own capital in their business (which, nevertheless, almost never takes away the need for outside funding[172]), brings another problem when equilibrium does not exist, as is the case in the real world.

The entrepreneur might have an *ex ante* expectation of profit, for instance. Yet, if he is unable to convince any *financial* entrepreneur to back him and finance his operation, he will be unable to execute his plans. Moreover, a *financial* entrepreneur normally holds (organizational) leverage over the entrepreneur. A *financial* entrepreneur can, for instance, add board members, impose strategic business decisions or even replace the entrepreneur. As financial intermediaries, *financial* entrepreneurs, determine the liability side of an entrepreneur's balance sheet, they effectively control the business, although in a very indirect way. Moreover, as the prices of an entrepreneur's means of financing are often public (or can be accessed by, for instance, banks when a need for additional funding arises), Yet, how an entrepreneur finances himself never fully depends neither on the *financial* entrepreneur nor on the entrepreneur: an entrepreneur can find various means of funding, generally, and each has its pro's and con's.

We thus propose the term *financial entrepreneur* or *capitalist-entrepreneur* to signify the link between savers (capitalists) and entrepreneurs. As we have concluded, a large body of (economic) literature is focused exclusively on the latter: nonfinancial entrepreneurs. Yet as a financial entrepreneur allocates capital to nonfinancial entrepreneurs, and thus (partly)

[172] Elon Musk might be an exception. Musk made his first fortune as co-founder of PayPal, which was later bought by eBay. This fortune allowed Musk to fund a greater part of his later businesses (Tesla, SpaceX, SolarCity), even though Musk normally relies on outside financing as well.

determine the capital structure of a society's businesses and entrepreneurs, they are perhaps even more important than nonfinancial entrepreneurs[173].

As part of a broader theory of entrepreneurship, it is important to include the *financial* entrepreneur as a key component in the process of capital allocation. Financial entrepreneurs move capital markets (which comprise all types of debt and equity instruments). Yet, just as their nonfinancial counterparts, they are *resource-less*. The capitalist is the resource provider. The capitalist-entrepreneur is the intermediary between capitalists and entrepreneurs. They arbitrage and find opportunities, but within financial markets.

Wicksell (1898) defines the capitalist as a "dealer in commodities" (p. 137), that is, as a resource owner. Entrepreneurs, or the entrepreneurial role, are working as such with "other people's resources." Banks and other intermediaries connect resource owners to entrepreneurs, who Wicksell (1898) implicitly defines as a Kirznerian entrepreneur that only makes money by arbitraging the prices of "inputs" (wages, rents, interest) and the price of its "output" (the final product), despite making references to rather confusing notions such as the "productivity of business" (p. 104). In an imaginary state of equilibrium, the entrepreneur "meets with neither a profit nor a loss" (p. 140), but merely receives the equivalent of a wage for his provided labor services ("the same return for the trouble of conducting his business as he would have obtained for conducting a similar business on behalf of others", p. 140).

The fact that the arbitrage-relationship has a second component (from capital providers to capital allocators to nonfinancial entrepreneurs) has largely been overlooked in the literature.

As Schleifer and Vishny (1997) argue:

> "When the arbitrageur manages other people's money, however, and these people do not know or understand exactly what he is doing, they will only observe him losing money when futures prices in London and Frankfurt diverge. They may therefore infer from this loss that the arbitrageur is not as competent as they previously thought,

[173] Even though, for instance, Mises's (1949) theory of the entrepreneur could be limited to *financial* entrepreneurs, yet this would make a large part of Mises's own theory ambiguous. *Financial* entrepreneurs are capital *allocators*, *nonfinancial* entrepreneurs are capital *users*.

> refuse to provide him with more capital, and even withdraw some of the capital -even though the expected return from the trade has increased." (p. 37)

In the long run, we might assume that the market tends to take resources away from capital providers with bad judgment and increase the resources under control of capital providers with better judgment. They select capital allocators and their subsequent allocation strategies. Indeed, the distribution of fund returns (earned by intermediaries / capital allocators) is best described by a Paretian "power law." Of the few successful investment funds that survive, many more disappear.

Additionally, since we refer to an economic role (a Misesian "ideal type"), rather than a real-life person, in practice any one individual might have different roles at the same time: for instance, consumer, capitalist and financial entrepreneur. Indeed, most (hedge)fund managers are a bit of both: they invest their savings alongside their fund investors[174], and some (after accumulating enough capital) have decided to close their funds to outside investors thus combining the role of *capitalist* and *financial entrepreneur* to a greater extent (in practice, none of the roles every overlap for a full one-hundred percent: a fund might hold liquidity at a bank, for example). This and other things become clear when we distinguish the *financial* entrepreneur from the *nonfinancial* entrepreneur.

[174] This is what Taleb (2010) would call "skin the game," sharing in both the gains and losses that their actions might produce (there should exist symmetry and "agency" in risk sharing).

Chapter 10: The Return on Capital

The goal of this chapter, although a small part of the whole, is to present out theory of profits, that is, our theory of the *return on capital*. The return on capital is relative; it always depends on *what* capital we are viewing. Hence, on the stock market we could buy title to some combination of productive assets (capital); the return on *that* capital depends on how much we pay for said title: *ceteris paribus*, if we pay more than its replacement value (that is, the underlying combination of productive assets), we are bound to have low or negative expected returns.

Generally, it should be observed that returns on financial assets adjust quicker to a change in the rate of interest than returns on nonfinancial ("productive") assets. Since financial assets are traded in capital markets, their prices can be bid up or bid down every single day. Productive assets, however, only experience price changes when they are actually bought and sold (which is not often, especially in the case of long-term assets). This idea could be applied to various markets: stock prices (secondary market) adjust more rapidly than productive asset prices, housing prices (secondary market) adjust more rapidly than construction costs, foreign exchange rates (secondary market) adjust more rapidly than the underlying banking assets (especially when they are illiquid, such as mortgage loans), prices of nearby futures contracts (secondary market) adjust more rapidly than the underlying commodity prices, spot gold prices (secondary market) adjust more rapidly than the underlying cost of mining and owning a mine, etcetera, etcetera. This is important since it will help to explain why *q* ratios rise (and deviate from their mean) over the course of the business cycle.

Furthermore, it is important to highlight our conceptualization of "the" return on capital. What matters, is not the *average* return on capital, but rather the *variability* of individual returns on capital. Just as there does not exist one single interest rate, there does not exist one single profit rate. No entrepreneur is concerned with economy-wide average returns on capital and is completely in his right to ignore any economist jabbering about average returns: an entrepreneur is concerned with the odds of earning a profit in his specific venture. He is thus concerned with the existence of positive (even high) returns in certain subsets – niches – of the market. This is completely rational, since accounting profits even out (counterbalance) accounting losses. There can be no such thing as aggregate profits or aggregate losses, except for special circumstances (a consumption of capital can cause aggregate profits).

We should not forget that any calculation of profits is a simple, yet uncertain approximation of the prices one pays for its inputs and the prices one receives for its outputs. Any profitable venture moves an economic structure closer to a hypothetical "equilibrium." Of course, capitalists (or their accomplices, i.e., *financial* entrepreneurs) tend to move savings to certain subsets of the market where returns are highest. Yet they do not need sophisticated DCF analysis for the theory to be true.

They simply make a judgment call whether the industries where returns are high are likely to remain high for an acceptable period of time. Moreover, they need to exclude other factors that would impair the return: selection of entrepreneurs, contractual mechanisms to protect against any downside and/or agency costs, sudden changes in natural conditions and/or consumer preferences. Nonfinancial entrepreneurs try to identify opportunities where they (expect) returns to be high. Both financial and nonfinancial entrepreneurs can either opt to use more sophisticated tools and models to estimate the attractiveness (or potential return) on a project, or they may not. In any case, from an economic point of view, it does not matter. The one thing that sound economic reasoning shows, is that an increase in savings and investment tends to go toward the marginal entrepreneur. That is, *ceteris paribus*, the entrepreneur most willing to borrow or sell part of its equity and thus offer higher marginal returns than the other entrepreneurs demanding investment on capital markets.

Now, perhaps the most important subject we have to touch upon in this chapter, is *the law of diminishing marginal returns on capital*. This was one of the most-debated points in any of the third rounds on capital controversies. We will try to offer a solution. Multiple things have been pointed out over the years:

(1) For the law of diminishing returns to apply, it is necessary for the model to have one fixed factors (this was the Kaldor-critique to Knight);
(2) Apparently, as can be observed empirically, profits do not seem to be diminishing. In fact, some of the economists viewed the rate of interest as equal to the rate of

profit (which we have debunked). At any rate, the rate of interest, and to a lesser extent the rate of profit[175], have behaved quite stable over time;

(3) The law of diminishing returns on capital has often been interpreted as a "physical" law: say you own a field of crops and have available a given amount of labor, then every additional piece of equipment will lead to "diminishing output." In other words, the law of diminishing returns has been interpreted, generally, as a physical rather than an economic law;

(4) If the law of diminishing returns on capital applies, you need some type of periodic shock in the fixed factors (sometimes population growth and at other times, for instance as Knight argued, knowledge "shocks"), otherwise profits will disappear and capitalism will bury its own grave (this is, in fact, the Marxian theory).

Given these three points, we will see that various points are often overlooked:

(1) Production is not physical;

(2) The fixed factor is not another "input" such as labor or knowledge;

(3) The shock that "counters" the tendency of returns to disappear does not have to be exogenous (population growth or knowledge shocks);

(4) Returns are not physical but economic.

So, is there space for a theory of diminishing returns on capital? Knight struggled with this question and almost ditched the concept of diminishing returns, until years later he wrote a follow-up that explained how "knowledge shocks" periodically increase returns. If such shocks do not occur, however, then marginal returns on capital eventually lead to a zero return and an end to the system. In other words, the easy way out is to simply deny the notion of diminishing marginal returns on capital: this, however, would mean that the concentration of owners of capital would be to such a degree that, as Piketty (2014) showed, *if* the return on capital (r) is higher than economic growth (g), capital owners will eventually end up owning everything and exploit the large classes of humble middle-class wage-earners.

[175] The rate of profit has behaved very erratic over the past century, showing large deviations over the course of historical business cycles.

What is the solution to this apparent paradox? I will briefly summarize our theory on the (diminishing) return on capital:

(1) Production is a value-concept (hence related to finance and profit and losses calculations);
(2) The fixed factor that leads to diminishing returns on capital is, if we avoid falling into the trap of thinking of capital as a physical input, the price mechanism;
(3) There is no need for an exogenous shock to avoid zero returns.

With regard to the first point, we should recognize that production is not physical, but about economic value creation. Prices and, subsequently, profit-and-loss calculations, tell us whether we have created or destroyed value. Capital, as we have defined earlier in this section, is simply any given *net worth*, which represents future profits (or losses). It is not a material input that will produce, when combined with other ingredients, some output. The notion diminishing returns can, therefore, never be physical, but can only refer to such value creation. When imbalances in the price structure are great[176], it is relatively easy to create value (things could hardly be worse from the point of view of a consumer). When imbalances are rare, little profit opportunities exist, and it becomes harder to produce economic value.

With regard to our second point, let us consider this: if returns on capital can simply be explained by the existence of price differences (that is, maladjustments) which represent potential profit opportunities that can be arbitraged away by entrepreneurs, then the presence of returns must have an objective existence in the price structure (the complex of all prices that exist at any point) at any given time. Entrepreneurial opportunity is thus reflected in the present and future price structure.

Hence, if we assume prices are not in equilibrium, that is, there are currently profit opportunities, and hold the price structure equal for our present purposes, every act of buying an undervalued asset and selling an overvalued asset, gives diminishing returns, if the interim price structure does not change (we abstract away from change).

[176] A point in case is the emergence of the Russian oligarchs when the Soviet Union fell apart, who arbitraged away the huge price differentials between domestic and foreign prices, especially of commodities such as gas and oil. Since essentially no businesses existed, *any* business would have a large potential to earn profits.

Simply, *all other things equal*, more (private) investment leads to better entrepreneurial arbitrage and will thus cause a marginal diminishing return on capital. At every extra dollar of investment (and hence price arbitrage by the entrepreneur's purchasing and selling), the return on that dollar diminishes. Hence, the more investment, the lower the returns.

Now, we will delve into our third point. What is left to explain, is the existence of the countertendency of profits to rise. This is where our previously explained Fisher's "pendulum of returns" comes into play[177]. Essentially it says that when returns are low, entrepreneurs will demand less investment (profit arbitraging gets harder) and rates of interest go up. This makes price arbitrage costlier and resources to do so are less readily available.

Hence, as a consequence, profits will tend to rise again, until the very opposite occurs: with higher profits, marginal entrepreneurs will begin to step in, demand present goods, and arbitrage profits away at a positive rate of return. This dynamic explains why, although diminishing returns on capital theoretically exist *if* the price structure would not change and is given, the rate of return on capital (profit) goes up and down but never to zero. The tendency of profits is endogenous, since there exists a negative feedback mechanism *in* the market process that will push up profits once they will become too low. Thus, there is no need for an exogenous factor, such as Knight's "knowledge" shock, to counteract diminishing returns on capital. A capitalist economy will rebalance itself and will counteract diminishing returns by reducing the resources at the disposal of entrepreneurs. By restricting the supply of investment, maladjustments and price differences will grow larger until, once again, profits are arbitraged downward by entrepreneurs, essentially explaining (a) the law of diminishing returns on capital and (b) the tendency of profits not to disappear entirely.

[177] For a more elaborate explanation of our proposed Fisher's "pendulum of returns" theory, see p. 296.

Chapter 11: The Valuation and Revaluation of Capital

As we have repeated before, *capital* is not an input to the production process. It is a tool of entrepreneurs that equals valuation of future profits and losses (that is, capital accounting). As such, more "capital" that causes more "production" is an oxymoron, since capital is a reflection of production, not a cause of production. The value of capital is therefore not stable. In this chapter, we will briefly summarize the factors that lead to the valuation and revaluation of capital. The value of capital (at any *scale*) is equal to the net present value (NPV) of the future yields it is able to generate. As such, the value of capital equals:

$$NPV = \sum_{t=0}^{n} \frac{CF_t}{(1 + r_t)^t}$$

where

CF_t = cash flow at time *t*
r_t = rate of interest at time *t*

Let us now take a brief look at what happens when these inputs change:

(1) ***An increase in the rate of interest***, is a signal that consumers are discounting future income at a higher rate than at a lower rate of interest. Hence, as such, the value of capital (as a sum total) will directly fall. The value of capital falls, but some investments, ventures and assets will fall more in market value than other assets: that is, any investment with a longer duration, ceteris paribus, will be discounted more strongly than an investment with a shorter duration. There is, thus, a greater emphasis on present income relative to future income. If this income is spent on consumption, rather than saved and invested, then it also means that, simultaneously, future cash flows near consumer sectors will increase relative to the future expected cash flows *further away* from consumer sectors (Hayek, 1931). Ideally, the discount rate used by all entrepreneurs is the market rate of interest.

(2) ***A decrease in the rate of interest***, is a signal that consumers are discounting the future at a less rapid rate than at a comparatively higher rate of interest *or* that there are few present "investment opportunities" or excess returns to be earned by entrepreneurs, who in response to the absence of potential profit opportunities reduce their demand on the intertemporal market. As a direct consequence, the value of capital will tend to increase, with little to no lag.

(3) ***A higher rate of return on capital***, is a signal that there are many "investment opportunities," that is, profit arbitrage opportunities abound. As such, contrary to popular belief, high profits do not reflect a healthy economy. Quite the opposite: increasing rates of profits indicate a maladjusted, imbalanced economy[178]. That is precisely where excess returns serve a purpose: to attract profit-eager arbitrageurs than can readjust the productive structure to the (intertemporal) preferences of consumers. Such higher rates of return, show up in the entrepreneurs' calculations and projections of future profits and losses. If he discovers *ex ante* that the projected revenues at a given quantity of customers far outstrips the projected expenditures on his inputs, say cement, copper, construction labor, equipment, transport services, et cetera, and he discovers, dynamically, over time that his expectations are *ex post* reasonably met, then that is how the rate of return shows up in the entrepreneur's economic calculations and capital accounting.

(4) ***A lower rate of return on capital***, is then on the very opposite a signal that entrepreneurial arbitrage has been efficient and little distortions arise. Hence, we should applaud low profits and welcome high profits as a means to induce entrepreneurial profit-seeking price-arbitraging activity. Moreover, it should be clear that in any closed-accounting system, the rate of the return on capital is on the whole zero. Hence, when we refer to "higher" and "lower" rates of return, what we are really saying is that in some parts of the economy excess returns can be earned, while in other parts of the economy corresponding losses are incurred. When arbitrage is generally "easy," it means that the

[178] This is very much in line with Taleb's (2012) observation that regulations do not tend to decrease profits, but increase profits (at the expense of others). Profits are sign of disequilibrium, not of equilibrium. They are equal to the "price surges" in the Uber-taxi app: when profits per ride rise, it is a sign that more suppliers are necessary. Profits go down when more suppliers begin arbitraging away the difference.

absolute mean deviation of rates of return (profits) are extremely high. When arbitrage is "hard" (many prices are close to equilibrium), it means that the absolute mean deviation of rates of return (profits) is low. There are some small losses in some parts of the economy and some small excess profits in other parts of the economy, which might discourage entrepreneurs from raising funds and investing and reduce the demand for present goods on the intertemporal market.

Hence, high profits indicate large maladjustments in the price structure. Capital, if aggregated, will thus be higher than normally is the case. Hence, we can also see that aggregating capital makes relatively limited sense. It is not *as if* a higher capital or capital per capital means a society is better off. Rather, it means that implied ROIC is high and that therefore entrepreneurs should step up to reduce profits. Likewise, an increase in interest rates *lowers* aggregate capital (yet long durations more than short durations). This does not mean that suddenly the number of physical means has decreased. It merely signals to entrepreneurs what ROIC they possibly might expect at which maturity.

It is interesting to note, in conclusion, that the relationship between interest rates and capital is *convex*. Ergo, an increase in interest from 1% to 2% has a much larger impact on the value of capital than an increase from 10% to 11%. The effect of interest rates on the (re)valuation of capital is therefore nonlinear, which is counterintuitive to many economists but especially relevant in the context of price signals and incentives to prospective entrepreneurs.

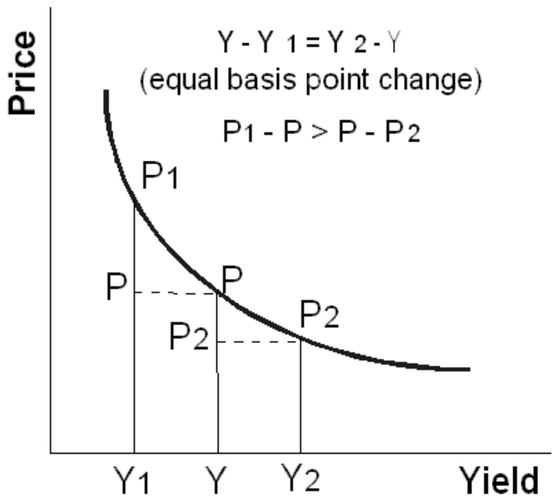


Figure 59: The relationship between interest rates and (the value of) capital is convex, which is especially important when capital gets 'revalued'.

Section IV:
Capital in Equilibrium: A Theoretical Application to Stock, Commodity, Forex and Real Estate Markets

In this section, we will define four equilibrium states in four different asset classes of *q* (where market value and replacement value are in unity) and then attempt to prove that q, in disequilibrium, is able to predict *expected* returns for each asset class. We will thus present our initial work on equity *q*, real estate *q*, forex *q*, and gold *q*, applying in all instances our (capitalized) WACC = ROIC approach, which forms the basis of our financial and subjective theory of capital.

Note that what we are doing here, is not — unlike Mises (1949) — distinguishing physical capital goods from our concept of capital. In fact, we do not need any *physical* reference to the underlying assets that are actually used by entrepreneurs (in their respective asset classes). The only thing that we pretend to do is to split the sum (of capital) into its underlying parts. But we go from capital to capital; we do not go from capital to capital goods. If we refer to a house as capital, we refer to the present value of its future benefits. Yet if we refer to cost, we merely refer to the present value of the materials (or "productive assets") and the present value of the land (which equals the present value of its future benefits). Both are capital, just at different *scales*. We therefore do not need any reference to physical objects; a mere reference to the various cash flows, composing and decomposing such "economic" flows driven by *value* exchange, suffices. We are concerned with value and price, not material objects or physical productivity.

Chapter 12: The Return on Equity and Q

As we have explained in our way up to this chapter, capital equals financial *net wo*rth which, in turn, equals the underlying productive assets net of debt. The stock market is, in fact, the *sine qua non* of capital[179]. The equity *q* ratio, then, reflects the price that is paid on the stock market for (the net present value of) a combination of productive assets. This idea originated partly from the work of Tobin (Tobin & Brainard, 1976), yet there is a key difference: we recognize that individual firms might have higher than normal returns on capital that therefore justify a higher premium over the underlying, individual productive assets. That is a circumstance of disequilibrium. Tobin (1976) discussed the Tobin *q* in the context of an individual firm relative to its enterprise value, that is, the market value of its equity *and* of its debt. As stock markets only reflect claims on equity, we will look at the broader index (in this case the S&P 500, although any stock market can be viewed from the perspective of *q*) and only to the replacement value of the equity, *not* the entire firm.

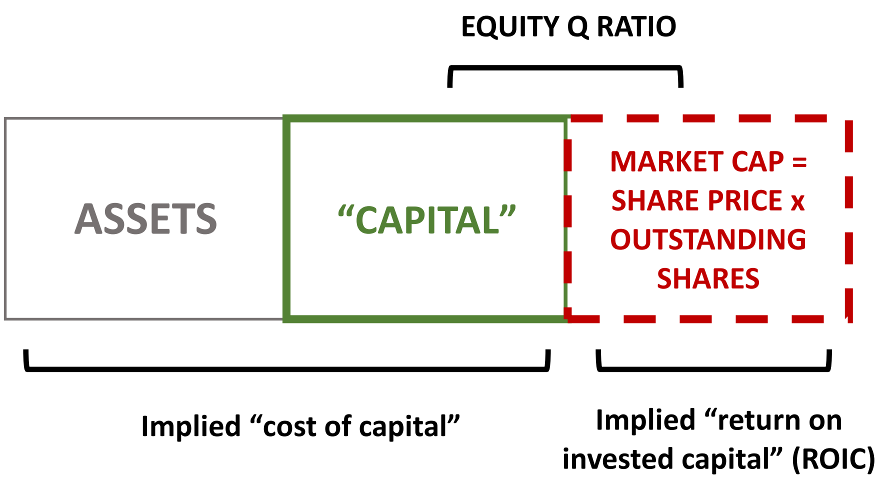


*Figure 60: The basic principle behind the equity q ratio: a financial claim (sold on the stock market) against firm "capital" (*net worth*) which is backed by a combination of productive assets.*

To arrive at the equity *q* ratio one therefore has to divide the market value of the capital with the replacement value of the capital represented by any such equity claim:

[179] As Murray Rothbard recites, when Ludwig von Mises was asked by one of his students what the difference was between a socialist and capitalist economy, he answered with a surprisingly clear answer: a stock market. Allegedly, Mises said: "A stock market is crucial to the existence of capitalism and private property. For it means that there is a functioning market in the exchange of private titles to the means of production. There can be no genuine private ownership of capital without a stock market: there can be no true socialism if such a market is allowed to exist."

$$equity\ q = \frac{Market\ value\ of\ equity}{Replacement\ value\ of\ equity}$$

Or, in slightly different terms (Smithers & Wright, 2000):

$$equity\ q = \frac{Stock\ price}{Net\ worth\ per\ share}$$

The key difference between the equity q ratio and the often-cited price-to-book value measure (which divides the market price of a stock by the book value represented by its equity) is the fact that book value does not always equal replacement value: as, in many cases, firm simply use the historical cost adjusted for some rate of depreciation in their books, the two do not coincide. As Spitznagel (2011) notes, the only correct way to estimate *q* is by taking the market prices of the underlying assets, which represent a better proxy to real replacement value. In this sense, the equity *q* ratio is superior to any price-to-book ratio. First, we want to see how this ratio has behaved historically:

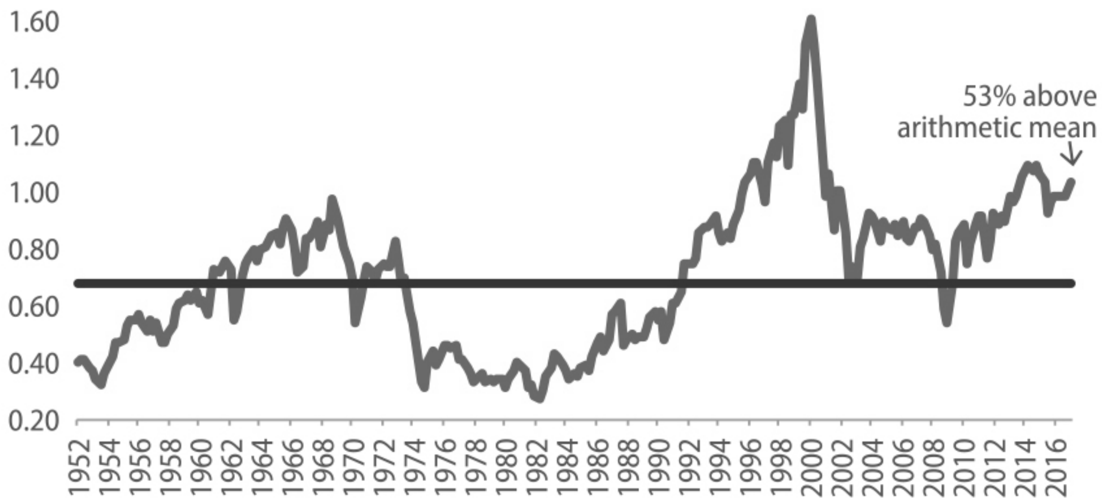


Figure 61: The historical equity q ratio in the U.S. from 1952 to 2016.

The data is from the Board of Governors of the Federal Reserve, release Z.1 Financial Accounts of the United States, Table B.103, Nonfinancial Corporate Business; Corporate Equities (representing market value) divided by Nonfinancial Corporate Business; Net Worth (representing replacement value). The theoretical equilibrium value for equity *q* is 1: the price of equity equals its replacement value and, as such, there are no arbitrage opportunities. Nevertheless, the historical mean (depending on whether we use geometrical or simple arithmetic mean) is close to 0.7. Hence, for some reason, equilibrium equity *q*, in practice, equals .7 rather than 1. As Wright (2004) offers a likely explanation: “A puzzling feature is that all resulting series for *q* have an apparently stable historic mean that is significantly less than unity. If systematic mispricing over the course of a century is ruled out, the most likely explanation (for which some circumstantial supporting evidence is presented) **would appear to be a systematic tendency to overestimate the replacement value of the physical capital stock, due to underestimation of depreciation**.” (p. 2) [emphasis mine]. In this sense, the “gap” between the observed historical mean *q* and the theoretical suggestion that equity *q* in equilibrium should equal 1 (“unity”) is more than anything else caused by limitations on data[180].

From simple observation, we can observe that equity *q* reached an absolute peak in 2000. We can also observe equity *q* was quite high (which implies that price exceeded replacement value) in late 1960s, 2006/2007 and, more recently, in 2016. Moreover, as Smithers & Wright (2000) calculated: equity *q* reached 1.75 in 1906 (before the recession of 1908 in which GDP fell 8.3%), 2.01 in 1929 (before the Great Depression in which GDP fell 27%) and 1.65 in 1937 (before the secondary recession of the Great Depression in 1938 in which GDP fell 3.6%). Here we briefly highlight equity *q* at an extreme peak compared to its lowest lows:

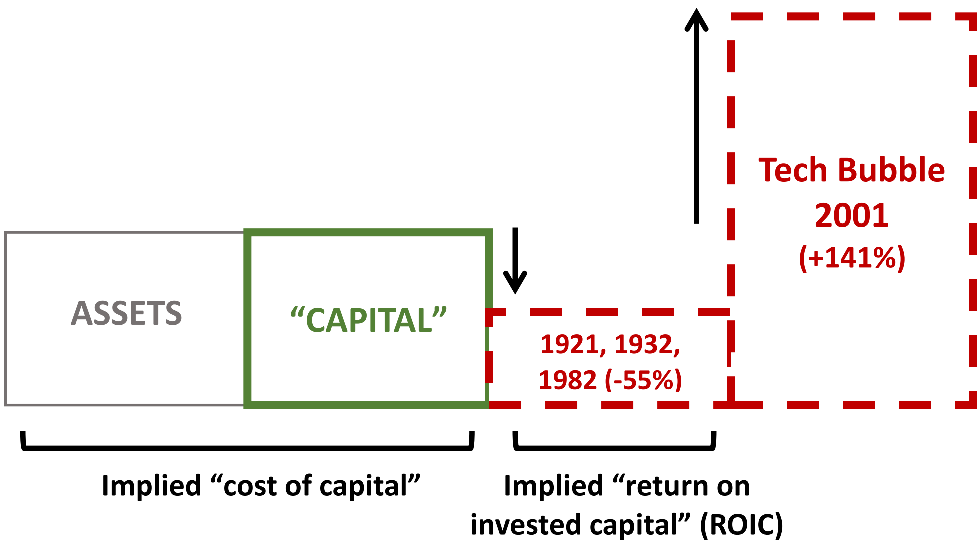


Figure 62: Historical extremes (peak and lows) of equity q.

[180] For other explanations, see Spitznagel (2011).

Our next question, thus, should be: does equity *q* predict future returns? According to our own financial theory of capital, the answer should be "yes," yet let us glimpse at the data.

To be predictive of future returns, equity q must have a "mean reversion." As Smithers & Wright (2000) argue: "The property of mean reversion means that when *q* is high it is likely to fall back, and when *q* is low it is likely to rise." (p. 121). First, they show a simply graph that charts the annual return of the S&P Industrials to changes in q. This graph shows that changes in equity *q* are mostly driven by the "stock price"-side rather than the "replacement value"-side:

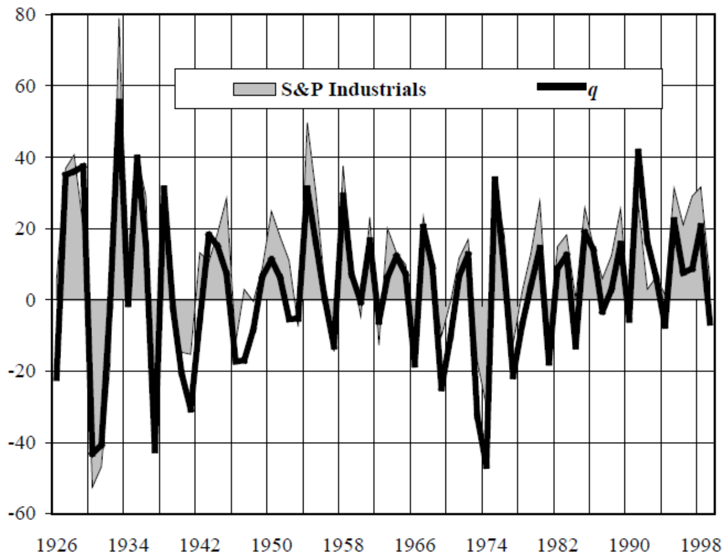


Figure 63: A chart that shows the annual change in the S&P Industrials versus annual change in equity q. *The two move closely together, showing that changes in equity q are mostly driven by the "market value" rather than the "replacement value" side of the equation (Smithers & Wright, 2000).*

Generally, it should be observed that returns on financial assets adjust quicker to a change in the rate of interest than returns on nonfinancial ("productive") assets. That is, financial assets are traded around the clock, and their prices can therefore be adjusted around the clock, whereas productive assets are rarely bought and sold and must be bid up "dollar for dollar."

So, does equity *q* predicts future returns? Spitznagel (2013) shows that equity *q* does:

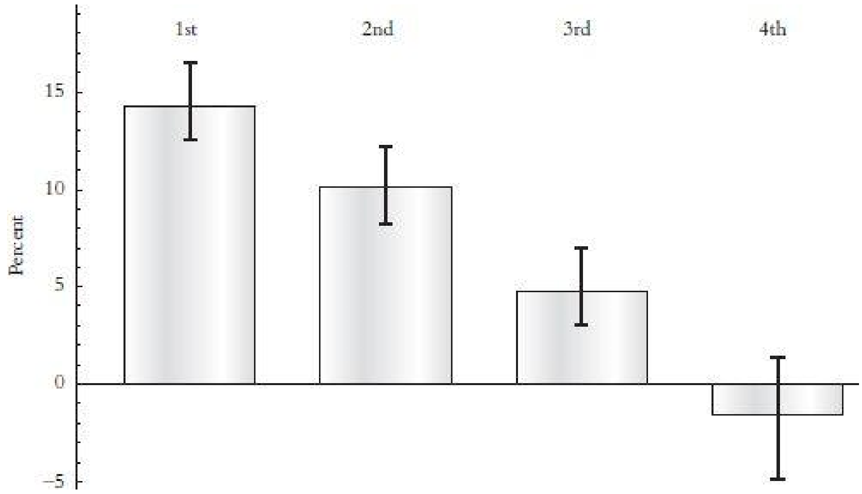


Figure 64: Equity q *is predictive of future (1-year) returns (Spitznagel, 2013).*

Spitznagel (2013) divided each year's (starting) equity *q* ratio into quartiles (from low *q* to high *q*). Then he took the subsequent S&P Composite 1-year return of each q on a monthly basis from 1901 to 2013. His conclusion is that with higher equity *q* (especially the fourth, most extreme quartile) subsequent expected returns are low or even negative. On the flipside, with lower equity *q* (the first quartile in this case), subsequent 1-year returns are higher.

Moreover, as Spitznagel (2013) also shows, with higher quartile equity *q*, the expected two-month losses are also higher. He took the two-month returns at starting equity q (again in the same four quartiles from low to high *q*) and looked at the 2 and 5 percentile of returns (in other words, how low are 2% and 5% of the two-month returns for any given *q* quartile?).

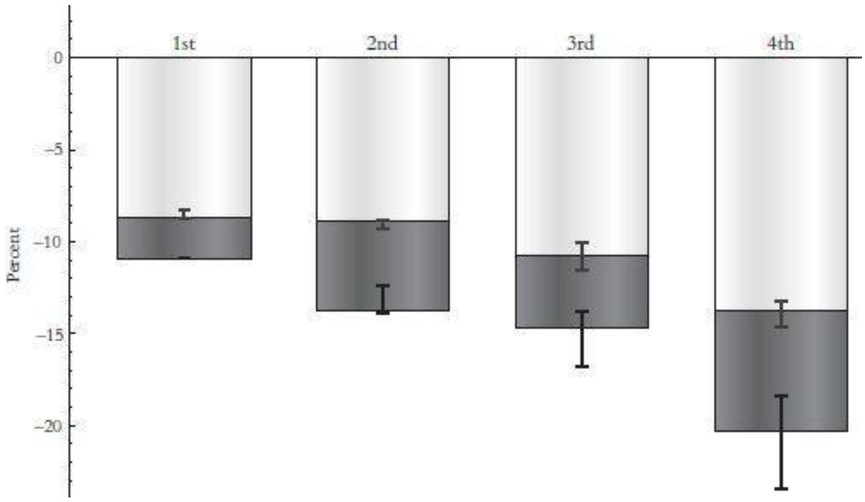


Figure 65: Equity q is also an indicator of potential risk of drawdowns. In this case, Spitznagel (2013) uses the same equity q *quartiles (low to high) to calculate the average (expected) 2-month drawdown for the 2 and 5 percentiles of each quartile.*

In this case, we can observe that if we are in the highest quartile of equity *q*, there is (indicated by historical data) a 2 percent chance (at 95% confidence) of a two-month negative return of more than 20% during any three-year timeframe. Hence, this would translate into a (two-month) decline (crash) of over 20 percent every fifty months. Put differently, when equity *q* is high, large stock market losses become *expected*.

Third, Spitznagel (2012) calculates the median number of months (at 99 percent confidence) for the S&P 500 to drop more than 20% when equity *q* is in its highest quartile:

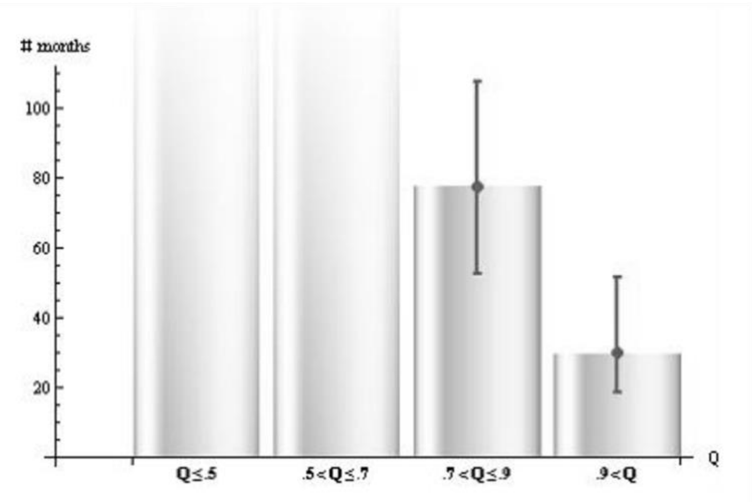


Figure 66: The median number of months before a decline of more than 20% of the S&P 500 occurs for all equity q *quartiles (again, from low to high* q*). In its highest quartile, it takes a median 30 months before a stock market crash occurs (Spitznagel, 2012).*

In other words, equity *q* is predictive of the time to a severe (20%+) decline in the stock market. Of course, these are simply historical data. Yet, the principle of *mean reversion* behind equity *q*, is a theoretical necessity. Therefore, even though we cannot predict the exact number of months before a stock market crash might occur, we can calculate the (historical) odds of a crash when equity *q* is at a certain level[181].

Now, what can be done to improve on equity *q*?

As we will discuss in *Section V* in greater detail, our theory of capital in disequilibrium revolves around the practice of *maturity mismatching*. This practice of maturity mismatching is reflected in the movements of the yield curve spread. During a phase of economic expansion, the yield curve spread expands (that is, the spread between short-term and long-term interest rates increases).

Nevertheless, when the phase of economic expansion is nearing its end, the yield curve spread contracts. Months before a recession, the yield curve often inverts: long-term interest rates exceed short-term interest rates. As has been generally recognized (e.g., Huerta de Soto, 2006; Bernanke, 1990), an inverted yield curve is one of the best predictors of a recession. Moreover, recessions and stock market crashes often coincide (Huerta de Soto, 2006).

Let us, therefore, attempt to replicate Spitznagel's (2012) and estimate the median number of months before equity *q* 'mean reverts' or, more specifically a 20%-drawdown occurs. Since we use equity *q* data from 1945 onwards (to 2016) we get a slightly different result.

We get equity *q* by dividing market value of equity by corporate net worth (from the Flow of Funds Release Z.1 Financial Accounts of the United States, Table B.103, as we did earlier). We calculate subsequent annual returns for every month with S&P Composite data from Shiller. Akin to Spitznagel (2012), we divide the observations into quartiles according to equity *q* (low, medium-low, medium-high, high). We get the following results:

---

[181] From personal research, I observed that the big exception among stock market crashes was the 1987 crash better known as Black Monday (which, also, as up to the moment of writing, is the largest daily loss in stock market history). Both equity *q* and yield curve spreads were not signaling any danger to come; it should be noted, however, that 1987 was not a bad year for equity investors at all; the S&P 500, for instance, closed in the black over the year.

	3rd and 4th quartile	4th quartile
Median	35 months	25.5 months
Average (arithmetic)	39.1 months	29.8 months

Figure 67: Median (and average) number of months before a 20%+ (annual) drawdown in the S&P Composite occurs (monthly data from 1945 to 2016).

Here we get a slightly lower median number of months (4th quartile of *q*), namely 25.5 months, than Spitznagel (2012). The difference can be attributed to data; we use a data series of *q* beginning in 1945 while Spitznagel (2012) uses a data series of *q* beginning in 1901. Apparently, the adjustment of market prices to deviations in equity *q* is faster (or, put differently, 20% drawdowns occur more frequently than before 1945).

As a next step, we want to see if we can lower the median number of months before a 20%+ drawdown by incorporating the yield curve spread. We collect annual data on interest rates from the St Louis Fed: specifically, the Fed discount rate (provided by the Federal Reserve of New York) until 1954 and the effective Federal funds rate from 1954 onwards as a proxy of short-term interest rates and the 10-Year US Treasury rate as a proxy of long-term interest rates. With both rates, we calculate the spread (long-term rate minus short-term rate). We then divide all the observations into two groups (a low spread and a high spread group) split by the *median*. In the low spread group, the probability of a recession would thus be higher than in the high spread group (a high spread is often considered healthy and, as we have seen above, common in a phase of economic expansion). Then we calculated the *median* (and average) number of months when equity *q* is high and the yield curve spread is low (low spread group):

	3rd and 4th quartile	4th quartile
Median	25 months **(-10)**	16 months **(-9.5)**
Average (arithmetic)	28.4 months **(-10.7)**	21.4 months **(-8.4)**

Figure 68: Median (and average) number of months before a 20%+ (annual) drawdown in the S&P Composite occurs (monthly data from 1945 to 2016) for equity q *quartiles and the* low *yield curve spread group.*

We see a marked improvement compared to Spitznagel's (2012) median months. That is, the odds of a 20% drawdown increase when both equity *q* is high and the yield curve spread is *low*, compared to when equity *q* is high and the yield curve spread is high (curiously, when equity q is high and the yield curve spread is high as well, the median number of months increase to 44). Hence, any tail hedging strategy on the stock market (to protect a portfolio against extreme declines, that is, drawdowns of 20% or more) is benefited by taking into account not only the equity *q* ratio, but the yield curve spread as well.

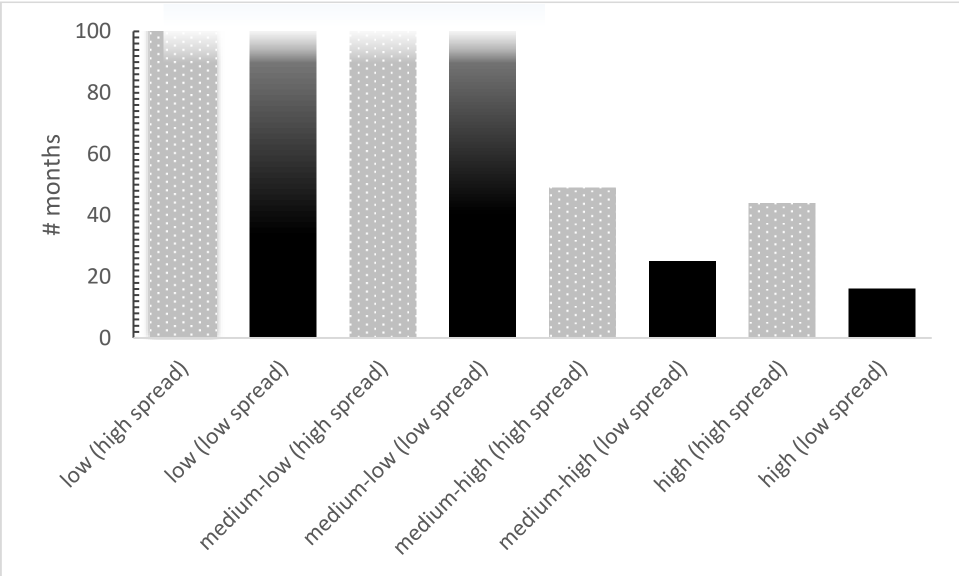


Figure 69: A comparison of the median number of months before a 20% (annual drawdown) given an equity q quartile (low to high) and yield curve spread (high spread versus low spread).

Chapter 13: The Return on Real Estate and Q

Another asset class where our principle of q (and thus our theory of capital) can be readily applied is the real estate sector. The two variables needed to apply the principle of q are clearly marked in case of real estate: on one side, the observed market value of a property or a sum of properties and, on the other side, the observed replacement cost of the same property (or sum of properties). In general, we could assume the replacement value to consist of two elements: (1) construction costs (which includes materials, labor, et cetera) and (2) residential land prices. Hence, our real estate (or housing) q would equal:

$$housing\ q = \frac{Market\ value\ of\ housing}{Replacement\ value\ of\ housing}$$

Where:

$$Replacement\ value\ of\ housing = construction\ costs + land\ costs$$

First, we will limit ourselves to residential real estate due to availability of data. Second, we gathered data from Knoll, Schularick, & Steger (2017)[182], which includes annual (residential) housing price indexes gathered from a variety of historical sources from 1950 to 2012 for twelve countries (including Australia, Belgium, Canada, Switzerland, Denmark, Finland, France, Great-Britain, The Netherlands, Norway, Sweden and the United States).

One of the limitations of the data provided by Knoll, Schularick, & Steger (2017), is the fact that land prices are "imputed." As the authors noted: "Primary historical data for the long-run evolution of residential land prices are extremely scarce." (p. 32). Using their own data on housing prices and construction costs, and assuming a stable share of land prices to construction cost at 0.50, they basically explain land prices by taking the differences between observed construction costs and housing prices. Hence, for our purposes (our attempt to identify deviations from a mean residential real estate q) this approach leads to nowhere. We are looking or desequilibrium away from replacement value, whereas in Knoll's et al. (2017) approach there is per definition an equilibrium between market value and replacement value (in this case construction and land costs). Since we have no reliable data on land prices, we will simplify and use exclusively a construction cost index as a proxy. Using the construction

[182] We are grateful to Knoll *et al*. (2017) for making their data publicly available.

cost index, we have calculated a housing q. To arrive at a provisional housing q, given the available data, we have thus used:

$$housing\ q = \frac{Housing\ price\ index\ (current)}{Construction\ cost\ index\ (current)}$$

We used the Knoll et al. (2017) data to construct a (country-specific) time series of housing prices and construction costs. We have disregarded the imputed land price data from the same series, as we have discussed above. With both housing prices and construction costs we estimated housing q. Moreover, with the housing price index we calculated the expected 1-year, 3-year, 5-year and 10-year future (average annual) returns. The results of our models were surprisingly robust over most time frames, albeit less pronounced in the 10-year average. We collected a total of 714 observations among the 12 different countries we have mentioned above.

Our first step, was to run a one-way ANOVA model with the housing *q* quartiles (1 to 4, low to high) as independent variable and the average 5-year annual return on housing as dependent variable. Our ANOVA p-value (0.00001) is lower than the 99% confidence level. We therefore conclude with 99% confidence that the mean returns on housing at the various q's are *significantly* different.

```
                      Sum of squares       df      Mean square

  Treatment                 0.290417        3        0.0968057
  Residual                   4.68096      686       0.00682356
  Total                      4.97138      689       0.00721536

  F(3, 686) = 0.0968057 / 0.00682356 = 14.187 [p-value
0.00001]

  Level         n       mean     std. dev

  1           143  0.0815098     0.098166
  2           171  0.0477377     0.079394
  3           129  0.0526648     0.072090
  4           247  0.0253143     0.080035

  Grand mean = 0.0476311
```

Figure 70: Our one-way ANOVA model output, which at a 99% confidence level shows that the average mean return among the q quartiles is *different.*

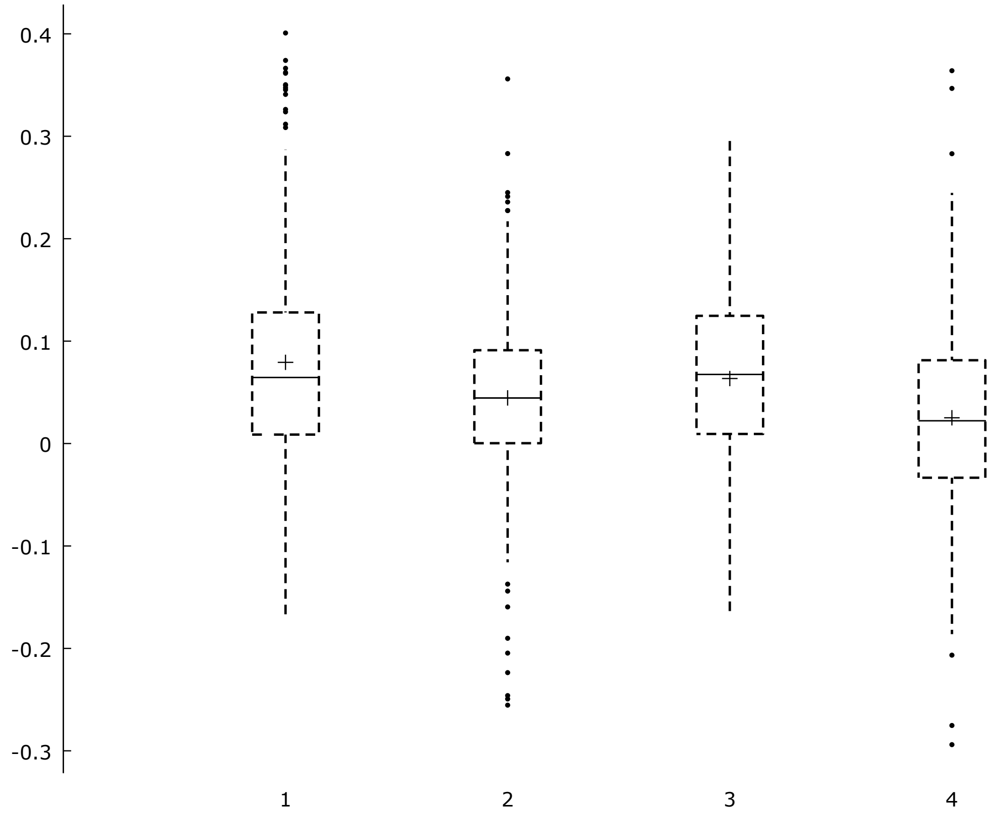


Figure 71: A factorized boxplot of different quartiles of housing q *with the future average annual 3-year return as dependent variable. Quartiles range from low* q *(1) to high* q *(4).*

q quartile	mean	min	Q1	median	Q3	max	
1	0.079267	-0.170060	.0087618	0.064424	0.12798	0.40082	(n=153)
2	0.044512	-0.255440	.00033770	0.044431	0.091099	0.35598	(n=176)
3	0.063441	-0.163600	.0093734	0.067560	0.12469	0.29767	(n=131)
4	***0.025123***	***-0.29399***	***-0.033396***	***0.022248***	***0.081330***	***0.36402***	***(n=254)***

Figure 72: A numerical summary of the median and mean expected annual 3-year return for every q *quartile.*

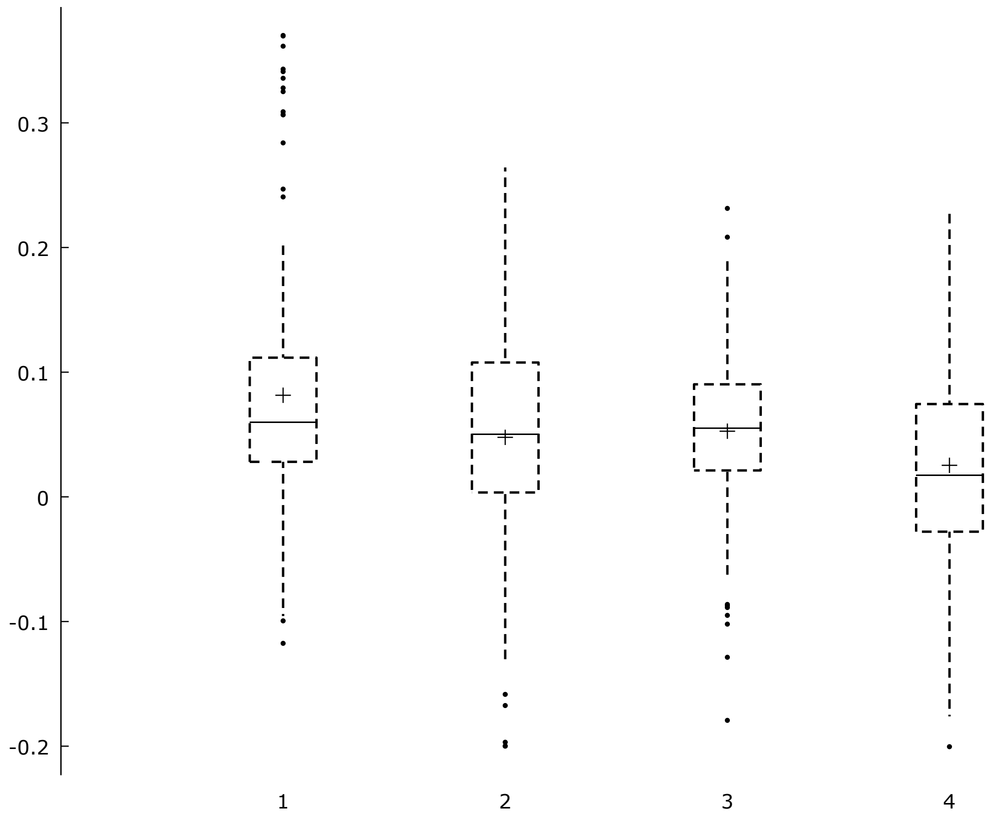


Figure 73: A factorized boxplot of different quartiles of housing q *with the future average annual 5-year return as dependent variable. Quartiles range from low* q *(1) to high* q *(4).*

q quartile	mean	min	Q1	median	Q3	max	
1	0.081510	-0.11749	0.028075	0.059765	0.11164	0.37011	(n=143)
2	0.047738	-0.19997	0.0034999	0.050101	0.10778	0.26420	(n=171)
3	0.052665	-0.17928	0.021173	0.055154	0.090300	0.23139	(n=129)
4	***0.025314***	***-0.20041***	***-0.027916***	***0.017302***	***0.074455***	***0.22801***	***(n=247)***

Figure 74: A numerical summary of the median and mean expected annual 3-year return for every q *quartile.*

From the boxplots, we see a less renounced effect of *q* quartile on future returns, although some pattern is visible. In the most extreme quartile (4, *overvaluation*) the (average 3-year and 5-year) expected mean and median return are *much lower* than for other quartiles. The median (mean) 5-year annual return drops to a mere 1.7% (2.5%), while the median (mean) 3-year annual expected return drops to 2.2% (2.5%).

We can also observe that at low valuations (quartile 1), future returns on (residential) real estate are relatively high: the median (mean) 5-year return reaches 6.0% (8.2%), while the median (mean) 3-year return reaches 6.4% (7.9%). The results of real estate *q* between 3-year and 5-year returns are robust. Hence, our theoretical housing *q* measure seems to make practical sense given the patterns in the data.

As a next step, we ran several panel data models with the 3-year (average annual) return as dependent variable and housing *q* (excluding land prices) as independent variable, with fixed and random effects, and with or without time effects. The model accounts for random country effects. The Hausman test, however, suggested that a random effects panel data model should be used. Nevertheless, the Breusch Pagan test suggested heteroscedasticity is a problem. Therefore, we choose a weighted least squares (WLS) panel data model. The results are summarized below:

Panel data model: WLS, using 714 observations
Included 12 countries
Dependent variable: 3-year average annual return
Weights based on per-unit error variances

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
constant	0.0870104	0.0107730	8.077	<0.0001	***
housing *q*	**−0.0519595**	0.0110570	−4.699	<0.0001	***

Figure 75: Output of our weighted least squares (WLS) panel data model of fourth quartile housing q *on future 3-year average annual returns on housing* Δ.

The Standard Error (S) of the regression shows high precision at 0.99 and a good fit of the model to the data. Both the constant and the housing q independent variable are highly significant (at a 99% confidence level). In effect, the results illustrate that when housing *q* increases, future average (3-year) returns are lower, in line with our theoretical exposition of housing *q*. In brief, we can conclude that the same principle of q can be applied to (residential) housing markets by taking the ratio between the market value and replacement value of real estate. Overvaluation is an important predictor of low future returns. Conversely, undervaluation is an indicator of possible higher future returns, although the effect of q on housing returns is especially strong at high levels of overvaluation.

In sum, we found evidence for our hypothesis that low housing q (high housing q) results in higher (lower) future returns on housing. Housing q is a significant and powerful predictor of future housing returns, akin to the more well-known equity q ratio.

However, efforts can be made to improve the quality of the data. As we have seen, there are fundamental problems with the available land data. Nonetheless, the Japanese housing market seems an interesting target for future empirical investigation since there is ample historical data on Japanese (both urban and rural) land prices, which can be combined with the construction cost index and the Japanese housing price index to arrive at a case study of the Japanese real estate market, especially in the context of the enormous bubble in the 1970s that eventually led to a decade-long incredibly depressed housing market.

.

Chapter 14: The Return on Foreign Exchange and Q

14.1 The Theoretical Foundation of Forex Q

It is generally accepted that interest rate differentials have an impact on exchange rates (e.g., Barker, McNelis, & Nickelsburg, 2013). The price level in any given economy (the general rate of inflation) is caused by changes in the prices of the underlying goods, often referred to as Cantillon effects (Huerta de Soto, 2006). It is similar to the mechanism behind our equity *q* ratio, in the sense that exchange rates (being prices of financial rather than productive assets) can easily adjust before prices of goods and services adjust (their prices adjust with a substantial lag since more monetary demand needs to bid up all such individual prices).

In some limited cases, it is possible that an increase in central bank liabilities increases domestic inflation before the exchange rate is affected, but this is very likely not a consistent phenomenon in economics, since exchange rates tend to reflect expectations of *future* inflation rates (that is, they should price in, to a certain degree, future losses of purchasing power). If domestic inflation precedes exchange rate depreciation, financial entrepreneurs effectively *failed to anticipate inflation*. If a currency is overvalued, exchange rate depreciation will occur. Consequently, if a currency is undervalued, exchange rate appreciation will occur.

So, how do we go about devising a *q ratio* for currencies on foreign exchange markets?

As we have seen before (e.g., Glasner, 1992) the (asset quality of the) backing of a monetary unit, despite the absence of direct convertibility, is still of vital importance. For Ize (2005), in order to maintain credibility, a central bank must have a positive comprehensive net worth (the present value of its future real profits), even when *current* profits and/or *current* equity are negative. Ize (2005), in his own words, proposes "a simple methodology to assess the need for central bank capital, **based on a forward-looking projection of profits, that is, on expected net worth.**" (p. 23) [emphasis mine].

One of the gaps in Ize's contribution is that he is unable to tie the value of the liabilities of an issuer (that is, central bank) to the net present value of its assets. Ize (2005) exclusively focuses his analysis on the latter, without establishing a clear and easy to grasp ratio against the former. Bringing central banking to its very essence, the future profits of a central bank are a function of:

(1) Its interest income (principally on foreign and domestic exchange reserves from seigniorage)
(2) Its interest expenses (for instance from interest on deposits)
(3) Its operating costs

Which provides us with the following net present value (NPV) equation:

$$NPV_{CB} = \sum_{t=0}^{n} \frac{(income_r - expenses_r - OPEX)_t}{(1+r)^t}$$

Where the NPV per unit of currency issued:

$$NPV_{currency\ unit} = \frac{NPV_{CB}}{Outstanding\ units}$$

And hence where the essential ratio of over- or undervaluation of a given currency is:

$$q_{currency} = \frac{P_{exchange\ rate}}{NPV_{currency\ unit}}$$

Since the NPV of a currency unit of a given currency (and, therefore, of an issuing central bank) is *per definition* denominated in a foreign or an exogenous unit of account, in order to estimate *q*, the exchange rate against that foreign unit of account should be used. Since we are talking about *relative* overvaluation in a dynamic context of multiple issuers and currencies trading against each other, each currency can have a *q* against another currency. For all practical purposes, it could be useful at this point to use a base currency such as the US dollar and, consequently, if the dollar and the Federal Reserve are subject to an analysis with the use of *q*, another large currency, to use a rival currency such as the euro[183].

[183] Forex traders could use our *currency q* theorem for their forex trades. However, as the inputs to our models are highly subjective, our scheme is a novel way of understanding foreign exchange rate movements, rather than a profitable trading strategy. Moreover, central banks (issuers of currencies) could use our scheme as part of their monetary policy decisions. At a first glance, however, many contingent factors seem to exist that determine exchange rates in the short term.

Any investor should, however, account for the probability of a government recapitalizing the central bank or, at a larger extreme, include the present value of future tax receivables in a consolidated view of both the central bank *and* the government backing it. These factors are, however, more difficult to capitalize and incorporate in our analysis. In such cases, it might even be useful to see a central bank from the perspective of a consolidated balance sheet of both the central bank and the government backing the central bank (Jordan, 2017).

All this is to imply that "vicious circles" of inflation and currency depreciation are overstated. This vicious circle is described, for instance, by Ahmad (1984): "A depreciation of the domestic currency initially aggravates the rate of domestic price inflation through an immediate rise in the price of traded goods expressed in domestic currency, which is quickly fed into the domestic price level. **This inflation, in turn, requires a depreciation of the exchange rate, which causes further inflation**." (p. 143) [emphasis mine]

Such a "vicious circle" can, in a majority of cases, be explained by some type of positive feedback mechanism[184] created by the very own central bank. For instance, if asset losses are the cause of declining exchange rates (or increasing domestic inflation), yet the central bank persists in continuing adding the same assets, they themselves are responsible for said "vicious circle."

Archer and Moser-Boehm (2013) argue clearly in favor of our theory of forex *q* and estimated comprehensive and book net worth of major central banks:

> "[T]heory suggests that central banks can get into financial trouble despite the clear financial advantages that come with their monopoly right to create base money, protection from bankruptcy proceedings and the backing of an owner with exceedingly deep pockets. **Such trouble is characterised by negative comprehensive net worth – that is, insufficient profitability over the entire (discounted) future to offset deficits**. Only two escape routes appear available to a central bank that might be at risk of finding itself in such a situation, and neither is attractive. The first is to alter policy course: ease up on inflation control, or eschew

[184] Soros' (2007) theory of reflexivity is another alternative to explain this "vicious circle." Nevertheless, Soros (2007) in his theory sometimes confuses the demand for a currency with the demand for financial assets (such as stocks and bonds) denominated in that same currency.

desirable though financially risky policy actions. And even this escape route is not without limits, as the revenue gains from higher inflation ultimately fall, and a poorly-functioning financial market may eventually drive intermediation offshore. The second escape route – fresh real resources transferred from the taxpayer – may conflict with the policymaking incentive structures purposefully constructed by central bank independence, since taxpayer resources are intermediated through the political process. And public finances may not be in good enough shape for governments to forgo the chance to dip into inflation taxes." (pp. 17-18) [emphasis mine]

Buiter (2008), meanwhile, argues that a great part of estimating the NPV of a central bank's net worth consists in estimating its net present value of *future seigniorage profits*, with its other assets, compared against base money (and other liabilities) and the net present value of future operational expenses including the net present value of (dividend) payments to the government. something that generally does not appear formally on a central bank's balance sheet. His approach is very similar to our theoretical understanding of central bank finances and the asset backing any given currency.

We could also take a slightly different approach. The relatively straightforward relationship between real and "financial" markets has often been misunderstood.

(1)

$$P_{assets} - P_{liabilities} = P_{net\ worth}$$

Where:
P_{assets} = the current market price of assets (factors of production)
$P_{liabilities}$ = the face value of outstanding liabilities (debt)
$P_{net\ worth}$=the replacement value of the factors of production net of debt

Financial assets are simply contractual claims on these assets, therefore:

(2)

$$P_{liabilities} + P_{net\ worth} = P_{financial\ assets}$$

Where:

(3)

$$P_{financial\ assets} = P_{debts} + P_{equities}$$

Stocks are simply titles to net worth; therefore, we get to our equity q ratio:

(4)

$$equity\ q = \frac{P_{equities}}{P_{net\ worth}}$$

Where:

(5)

$$P_{equities} = \sum_{i=1}^{n} \left(\frac{FCFE_i}{(1 + r_d)^i} \right)$$

Remembering:

(4)

$$P_{liabilities} \approx P_{debts}$$

These debt instruments (P_{debts}) are held by either financial intermediaries or outright by households.

(5)

$$P_{debts} = (Q_{institutional} + Q_{retail})\, P_{debts.avg}$$

From institutional holdings (debt instruments held by financial intermediaries) we go to bank holdings. Therefore:

(6)

$$Q_{institutional} = Q_{banks} + Q_{nonbanks}$$

Therefore, the banking system’s liabilities and net worth are equal to the quantity (Q) held by banks times the average-weighted price of corporate debts:

(7)

$$P_{debts.avg}\ x\ \left(\frac{Q_{banks}}{Q_{institutional} + Q_{retail}}\right) + \ P_{other\ bank\ assets} = \ P_{bank\ liabilities} + \ P_{bank\ equity}$$

Of which:

(8)

$$P_{bank\ liabilities} = P_{bank\ deposits} + P_{bank\ debts}$$

Of which:

(9)

$$P_{bank\ liabilities} = \ P_{central\ bank\ liabilities} + P_{commercial\ bank\ liabilities}$$

Here we arrive at our *banking q* (the primordial relationship between the replacement value of bank assets and the market value of bank liabilities), which is the (partial) premise of the backing (asset quality) theory of money:

(10)

$$banking\ q = \frac{P_{bank\ liabilities}}{P_{debts}\ x\ Q_{banks}}$$

Whenever $P_{\text{bank liabilities}}$ exceeds P_{debts} x Q_{banks}, either holders of bank liabilities dispose of them (reducing the market price, that is, the purchasing power of the liabilities), redeem them for base money (if convertibility is possible) or exchange them for rival banks' liabilities (negative clearings) *or* the bank is forced to raise interest rates on its liabilities and/or withdraw outstanding bank liabilities from circulation by selling assets and cancelling liabilities.

To sum up, there exists a close theoretical relation between the backing of a currency (reflected in the net worth and future profits and losses of a central bank) and its future fate (in terms of inflation and/or currency depreciation).

14.2 Empirical Results

We have used data from Klüh & Stella (2008), *Table A3. Raw Data for Latin America*, which is derived from different sources: Leone (1994), IMF staff reports and various local central banks, which spans a period from 1987 to 2005. This data on central bank losses was supplemented with data from other sources, such as Gross Domestic Product (GDP) figures (source: St Louis Fed), monetary base (source: International Monetary Fund), inflation as measured by consumer prices (Feenstra, Inklaar, & Timmer, 2015), and exchange rate against the U.S. dollar (St Louis Fed). Our data series spans 253 observations across 15 (Latin-American) countries in 19 periods (1987 to 2005). With the data available, we attempted to calculate a crude proxy of forex *q*:

$$forex\ q = \frac{Market\ value\ of\ monetary\ base\ (in\ dollars)}{Market\ value\ of\ monetary\ base \pm\ profit/loss}$$

First, we ran a one-way ANOVA model with the forex *q* quartiles (1 to 4, low to high) as independent variable and the (change in the) exchange rate as dependent variable. Our ANOVA p-value (0.0277) is lower than the 95% confidence level, therefore we conclude with 95% confidence that the means of various q's are *significantly* different.

	Sum of squares	**df**	**Mean square**
Treatment	0.692493	3	0.230831
Residual	18.8965	253	0.0746897
Total	19.589	256	0.0765195

F(3, 253) = 0.230831 / 0.0746897 = 3.09053 [p-value 0.0277]

Level	**n**	**mean**	**std. dev**
q = 1	89	0.120727	0.23190
q = 2	42	0.125788	0.22244
q = 3	62	0.207342	0.30838
q = 4	64	0.239476	0.31709

Grand mean = 0.172021

Figure 76: Our one-way ANOVA model output, which at a 95% confidence level shows that the average mean return among the q quartiles is *different.*

As a next step, we estimated a panel data model, with random effects and robust standard errors but no time dummies, with the (change in the) exchange rate as dependent variable and a dummy variable for our fourth q quartile (which indicates currency *overvaluation*) as independent variable. A Hausman test gives a highly significant value, which indicates that we should take a panel data model with fixed effects. The Breusch-Pagan test is not significant, which indicates that heteroscedasticity is not an issue, as expected since we exclude the other quartiles. We thus repeated the panel data model with fixed effects, robust standard errors and without time effects. The result is a significant effect of fourth quartile *q* on future exchange rate (**when forex *q* reaches the fourth quartile, foreign exchange losses are 8.3% higher at a 90% confidence level**) with R^2 at *0.32*. Moreover, there is a significant one-year lag in the dependent variable (exchange rate). Below we summarize briefly the output:

Panel data model: fixed-effects, using 244 observations
Included 15 countries
Time-series length: minimum 5, maximum 18
Dependent variable: exchange rate Δ
Robust (HAC) standard errors

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
constant	0.0718189	0.0238517	3.011	0.0093	***
q = 4 dummy	0.0830284	0.0409598	2.027	0.0621	*
exchange_rate_1	0.415791	0.136931	3.036	0.0089	***

Figure 77: Output of our fixed effects panel data model of fourth quartile q *on exchange rate* Δ.

We have used factorized boxplots to show the expected returns (increase/decrease of exchange rate against the US dollar and increase/decrease of domestic purchasing power) for every quartile of our forex *q* derivative / proxy. Beginning with plotting the relationship between *q* and the exchange rate, we have capped extreme outliers and changed their values to 1 (which amounts to a 50% depreciation, understood as, for example, a change in a hypothetical exchange rate of 2-to-1 to 4-to-1). Then we divided all historical observations into four *q* buckets: from 1 (low, signaling undervaluation) to 4 (high, signaling overvaluation). Below you can observe the result:

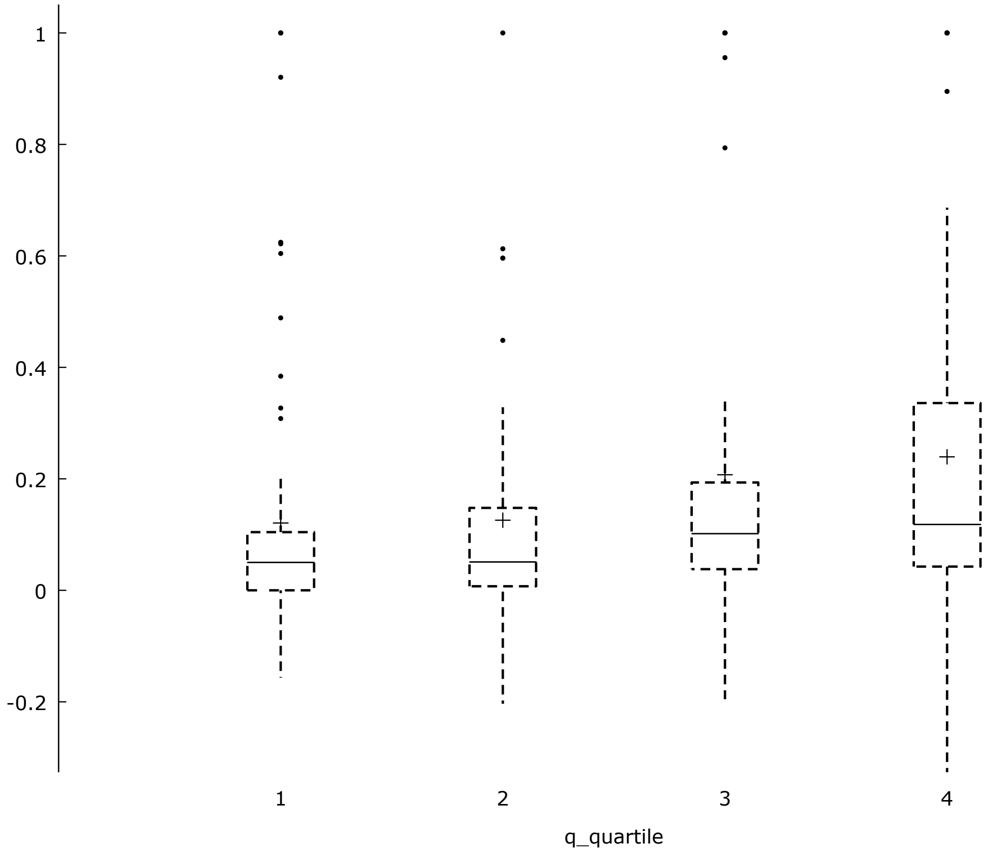


Figure 78: A factorized boxplot of different quartiles of q *with the (annual rate of change in the) exchange rate as dependent variable.*

q_quartile	**mean**	**min**	**Q1**	**median**	**Q3**	**max**	
1	0.12073	-0.15664	0.00000	0.049951	0.10443	1.0000	(n=89)
2	0.12579	-0.20351	0.0073111	0.050663	0.14786	1.0000	(n=42)
3	0.20734	-0.19533	0.037963	0.10164	0.19349	1.0000	(n=62)
4	0.23948	-0.39767	0.042537	0.11796	0.33601	1.0000	(n=64)

Figure 79: A numerical summary of the median and mean expected return (as measured by the 1-year change in the dollar exchange rate) for every q *quartile.*

With the above data, we can appreciate that, effectively, the median (mean) expected 1-year return when forex *q* is in the fourth quartile equals negative 11.8% (negative 24%). On the other side of the spectrum, we can observe that the median (mean) expected 1-year return when forex *q* is in the first quartile (undervaluation) equals a slightly negative 5% (negative 12%). Hence, the higher *q*, the larger the expected foreign exchange rate losses.

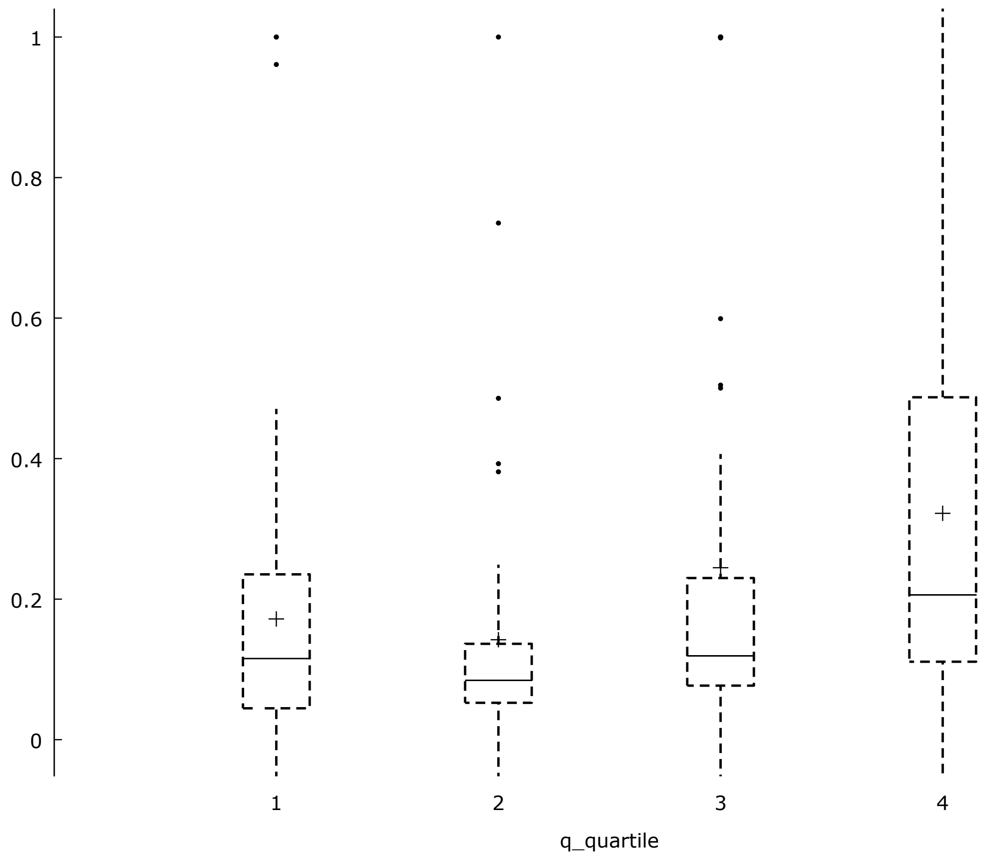


Figure 80: A factorized boxplot of different quartiles of q *with (the annual rate of) inflation as dependent variable. Higher* q *(higher overvaluation) points to higher expected future inflation.*

q_quartile	mean	min	Q1	median	Q3	max	
1	0.12073	-0.15664	0.00000	0.049951	0.10443	1.0000	(n=89)
2	0.12579	-0.20351	0.0073111	0.050663	0.14786	1.0000	(n=42)
3	0.20734	-0.19533	0.037963	0.10164	0.19349	1.0000	(n=62)
4	0.23948	-0.39767	0.042537	0.11796	0.33601	1.0000	(n=64)

Figure 81: A numerical summary of the median and mean expected return (as measured by the 1-year rate of inflation) for every q *quartile.*

With the above data, we can recognize a similar tendency. The higher q, the higher the future median rate of inflation. For instance, in the fourth quartile, the expected future inflation equals a median rate of 11.8% and a mean rate of 24%. At its lowest quartile, the expected future (domestic) inflation drops to a median rate of 5% and a mean rate of 12%. The mean, as we can observe, is very unstable due to the non-normal (power law) distribution of annual changes in the foreign exchange and inflation rate: the mean has therefore limited meaning.

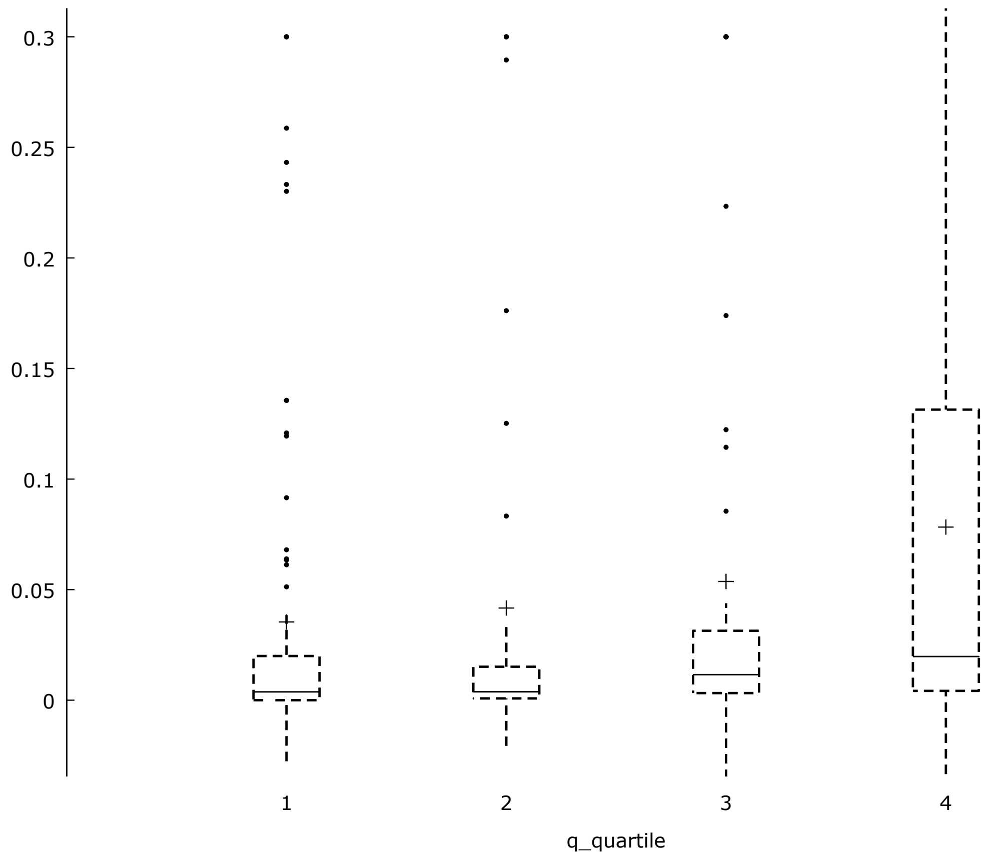


Figure 82: A factorized boxplot of different quartiles of q *with the* product *of inflation and exchange rate appreciation/depreciation (inflation multiplied by change in exchange rate). Now the extreme, the so-called tail, becomes an even more expected event.*

Another interesting observation is that, regardless of the example, deviations from the median (volatility) increase exponentially when forex *q* is high. The risk of large drawdowns apparently increases when *q* is high.

There are many factors, however, that are left out of our model that, as a result, are not captured in our estimates. I will briefly outline the factors that we have missed by simplifying the theoretical foreign exchange *q* to an easily obtainable *q* based exclusively on central bank losses:

– A central bank profit or loss might be temporary or permanent; if the loss is caused by exchange rate appreciation (and, hence, a downward revaluation of foreign exchange reserves), the loss will not endure and, moreover, easily turn into a gain *if* the exchange rate begins to fall against other currencies.

- A central bank might incur an enormous financial loss in its balance sheet without recognizing it, by not marking its asset to market. This is particularly in case of worthless public debt. In volatile countries, the central bank (as a public institution) might be unwilling to consider the loss of principal of the public debt it holds. In some scenarios, governments turn the loans into a perpetuity bond with zero interest. In this case, no direct loss is recognized, but there is a great real loss that, without a doubt, will not be left unnoticed by market participants. In this case, losses might now show up, while its exchange rate and/or inflation will explode. A case in point, with regard to our own data, includes the case of the Bank of Jamaica, which merely includes cash losses, but excludes the Bank of Jamaica's 'Special Issue Bond'.

- A central bank might have assets that are not marked-to-market, essentially serving as a "buffer" against the liabilities it has issued: a good example is the Federal Reserve, which has gold on its balance sheet against a value of $35/troy ounce. Nevertheless, its *real* market value is substantially higher than reported. Therefore, any recognized loss might be compensated by an unrecognized gain, which could not be captured in the data. Here, again, we see the difficulty in estimating forex q as many central banks around the world are not bound by the same accounting rules and principles as ordinary businesses or even governments.

- Market participants might price in the probability of the government directly backing and recapitalizing the central bank. In this case, the losses are shifted to government and, thus, the balance sheets (and net *worth*) of the government and the central bank ought to be consolidated. If a government is still able to finance itself on credit markets with relative ease, and if a government has a sound tax base, then the backing of government assets (and the governments' ability to tax its citizens) backs the value of the currency since, at any time, the assets of the central bank can be replaced or supplemented with others.

- Official exchange rates might diverge from real market exchange rates; even though we have used a data series that attempts to partly account for such divergences, it is clear that large differences between official and observed exchange rates might imply less severe foreign exchange rate losses than truly is the case. In these cases, central banks will have a high forex *q*, yet not high (officially recognized) exchange rate depreciation.

- We did not capitalize future losses or gains to get a picture closest to our theoretical forex *q* principle. We simply adjusted the value of the monetary base by the extent of a given year's loss or gain and, subsequently, try to determine whether market value of a currency unit diverges positively or negatively from the market value of the assets backing the currency unit.

- We were unable to include the present value of all future seigniorage of the central. As one of its main sources of income, it is important to estimate the value of a central bank's seigniorage. This depends, however, on both its operational costs and the risk-adjusted yields that can be earned in the assets to which it is bounded[185].

Future research will be necessary to further explore methods to estimate the (non-accounting) net worth of central banks, based on the present value of all future gains and losses of the central bank, including recapitalization attempts and a parent government's ability to recapitalize the central bank. Moreover, we need more data to estimate the net present values of future seigniorage, future operational expenses and future interest payments to government (Buiter, 2008). However, clear explanations of how the asset composition of central banks backs their respective currencies have lacked so far, especially in 'Austrian' literature. The recent attempts to use central bank net worth to explain and predict inflation and/or currency turmoil should therefore be embraced and extended further. This preliminary study has been one of such attempts to further the theory of foreign exchange (or central bank) *q*.

[185] Such a limit can be largely self-imposed, of course, as is the case with the present-day European Central Bank (ECB).

Chapter 15: The Return on Monetary Assets (Gold) and Q

Gold is a strange animal. It has a high stock-to-flow since all gold ever mined in the world forms continues to form part of its (potential) supply. Its supply and demand dynamics are, therefore, completely different from other financial assets. Gold, however, is not a liability to anyone. It is a physical asset. It most likely has been the source of most confusion in money and banking, too. Since regular laws of supply and demand apply to gold, just like coffee beans, the quantity theory of money appears to be a direct inheritance of the historical emergence of (physical) gold as money. Indeed, when physical gold moved around as medium of exchange, the quantity theory of money applied without exception. Moreover, when physical gold stopped circulating, but promissory notes payable in physical gold, and banks were supposed to a hold a 100% gold reserve against notes, again the law of supply and demand simply hold. It is when banking becomes more complex, and financial sectors more elaborate, that the law of supply and demand breaks down. Fisher (1912) recognized this very fact, with his famous Fisher equation of exchange:

$$Fisher\ equation = MV = PT$$

Where:
M = Money (cash) in circulation
V = Velocity of circulation of M
P = Price level
T = Volume of trade (number of units of goods which were exchanged for M)

Since the quantity theory no longer fitted due to the emergence of complex banking relationships, Fisher changed his equation to an “augmented” version to include checking accounts (checkable bank deposits):

$$Fisher\ equation = MV + M'V' = PT$$

Where:
M’ = Bank deposits
V’ = Velocity of circulation (rate of turnover) of deposits M’

What remains is, however, exactly the same. It is the quantity theory of money applied to both gold and bank demand deposits. And the quantity theory of money can be readily applied to simple monetary assets such as gold.

What determines supply, however, is in a flexible banking system, demand. Demand for means of payment depend on the liquidity preferences of an individual. In other words, it is an exchange in time. Given a certain amount of savings (stock not flow), individuals (households) must determine how much of their savings they wish to hold in means of payment or liquid assets (e.g., bank deposits) and how much they wish to invest in less liquid credit which cannot be used as means of payment or exchange[186]. They determine what specific qualities each financial asset must possess in their portfolio (e.g., means of payment, convertibility and/or redemption conditions, frequency of yield payments, et cetera) and at what maturity (that is, weighted maturity or duration) they wish to invest their savings. Hence, money supply depends (1) on the willingness to save income and (2) on the relative preference to invest part of that savings in monetary financial assets. This is the fundamental law of modern banking.

Now, to apply the same principles which we have touched upon many times in this work (i.e., ROIC = WACC, or *Capital value* = *Replacement value of assets*), we assume the market price of a troy ounce of gold (implied ROIC) and therefore begin with estimating the replacement value of a troy ounce of gold (implied WACC). The fundamental law of capital makes sense in this case: there is arbitrage between the price of gold and the replacement value of gold.

The replacement value of gold can be determined by estimating the cost of buying a mine and the cost of extracting and processing the gold. There are many measures of gold mining cost. In fact, "cash cost" is often reported by the gold mining industry. But it does not include (a) depreciation, (b) general and administrative expenses, and (c) other overhead costs. "Cash cost" is therefore not a sound long-term indicator of replacement value. "All-in sustaining cost," which is an alternative devised by the World Gold Council and various gold companies, was meant as another standardized measure. It includes *sustaining* CapEx directly related to the gold mining operations and general and administrative expenses. However, mining companies are in practice at times surprisingly creative in shifting CapEx around, failing to include certain capital expenditures in this cost measure. Another way of measuring

[186] For more on our "portfolio approach to the demand of money", see p. 349.

the production cost of gold is by deriving the cost of sales of gold of gold mining companies, which can be a viable alternative, at least at aggregate or industry level.

The troubling issue with such measures is the fact that they do not take into account the cost of *purchasing ownership in the mining company* (which is a rather blunt omission if we want to compare gold prices with their respective replacement value). Let us see how this works by taking one of the largest (and most economical) gold mining companies in the world, Barrick Gold Corp.

For the full year 2016, Barrick reports an all-in sustaining mining cost of $730 per troy ounce. In addition, Barrick reports 5.517 million ounces of gold production over the same period and 85.9 million ounces in proven reserves, which multiplied by the all-in sustaining cost gives us $4.027 billion for the year and $62.707 billion for proven reserves (which would equal the total cost of mining those 85.9 million ounces), assuming a stable all-in mining cost over time. Barrick has 1.166 billion shares outstanding at $14.50 a share. Hence, we would pay per share $14.50 for 0.074 ounce of gold in proven reserves **or 13.57 shares (times $14.50 a share gives $196.77) for one ounce of gold in proven reserves**. If we look at the rate of mining, 2016 annual production is 6.5% of total proven reserves, which gives the company a lifespan of 15 years with no increase in proven reserves. To get the future value of $196.77 we calculate:

(11)

$$\$196.77\ x\ (1+0.065)^{15} = \$506.06$$

Next, we add the all-in mining cost of $730 to get:

(12)

$$Replacement\ value\ (adjusted\ mining\ cost) = \$506.06 + \$730 = \$1236.06$$

Or, alternatively, the cost of sales of $790:

(13)

$$Replacement\ value\ (adjusted\ mining\ cost) = \$506.06 + \$790 = \$1296.06$$

This provides us with a realistic proxy of replacement value for an outside investor: he or she can either buy a share in a gold mining company that extracts its gold reserves at a certain cost or buy a troy ounce outright in the market. Therefore, our measure of *q*, or *gold q ratio*, equals:

(14)

$$gold\ q = \frac{P_{gold}}{Replacement\ value}$$

In our example, with all the inputs from Barrick for the full year of 2016 and the gold price in December 2016, this would amount to:

(15)

$$gold\ q = \frac{1{,}150.00}{1{,}236.06} = 0.93$$

Which would signal *relative undervaluation* (1 < q). Indeed, gold prices rose from $1,150/oz to over $1,300/oz. Since the data (beginning of 2016), the average annual return (which includes the two years 2016 and 2017) equaled 11.35%, which is higher than its average annual return since 1970 (7.9%) and higher than its average annual return since 1980 (1.9%).

For 2018, we can take the most recent share price of Barrick Gold Corp (14.80) and the current gold price, which would give us:

(16)

$$\$200.84\ x\ (1 + 0.065)^{15} = \$516.53$$

With Barrick's latest guidance on all-in sustaining costs per ounce (of $720 per troy ounce):

(17)

$$Replacement\ value\ (adjusted\ mining\ cost) = \$516.53 + \$720 = \$1236.53$$

Inserting the latest gold price (January 25, 2018) gives us:

(18)

$$gold\ q = \frac{1{,}350.00}{1{,}236.53} = 1.09$$

Which would indicate *relative overvaluation* (q > 1). In this case, expected returns on gold would be either low or negative (or positive but even more positive for mining shares).

Alternatively, we should take into account either the interest rate or the gold basis (the spread between the gold price on the spot market and the future gold price) to increase the discount rate to account for the time value of money (not just the extraction rate of proven reserves). This would then give us a higher replacement value and thus a lower *gold q ratio*.

Moreover, our example suffers from many other fundamental deficiencies. We have used only one gold mining company (Barrick Gold Corp), only one data point (end-of-year 2016) with many choices that could have affected our final *q*. We could use the average stock price for the year instead of end-of-year stock price[187]. We could use cost of goods sold (of gold) instead of all-in sustaining cost. We could use all large mining companies instead of only one. We could account for exploration in some way or another. All these factors would, however, perhaps disappear *if* we had a long time-series and a geometric average. Mean Q could for instance turn out to be nearer to $q = 0.9$ instead of $q = 1$, with the difference between both being all the factors that could possibly alter or impair our estimates.

The issue with gold, however (and the same applies to other monetary assets with relatively fixed and stable supply over time), is that corrections to *q* can come from both sides. Stock prices of mining companies could rise to raise our gold replacement value estimate (to reflect the new, higher market gold prices). This would actually be expected if the recent rally in gold prices would be viewed by stock market investors as “sustained” instead of “temporary.” Conversely, gold prices could fall to its replacement value. Moreover, mining companies could begin shutting down their marginally least economical mines, which could lead to a fall in the all-in sustaining mining cost per troy ounce. This, however, would make no sense for the moment since the gold price of a troy ounce exceeds the replacement value of a troy ounce.

For future research, we recommend the construction of a time series including multiple publicly-listed gold mining companies, deriving a geometrical mean, and calculating 3 to 5-

[187] This actually makes sense since Barrick (and other companies) only publish their data at year-end or even later. The share price would adjust to the new, published information (all-in mining cost, total production, proven reserves, etc.) which serve as the other inputs to our preliminary model.

year expected returns at different quartiles for *q*. In addition, the series could help us discover whether adjustments tend to come from the “gold price” side or the “replacement value” side and what they key drivers behind such moves are.

Our other model is based on Barsky’s and Summers’s work. They draw an inverse relationship between gold returns and real returns on capital markets. When capital markets returns are low (interest and stock market returns are low), then gold returns are high, and *vice versa*. This relation is quite stable. However, Barsky & Summers (1988) do not apply the term structure to real returns. Hence, every year gold’s “appreciation” would be the inverse of a negative real return on capital markets (interest rates as a convenient proxy), or “depreciation” if the real return were to be positive. However, this approach appears fruitless. It leads to very small changes and forecasts that do no justice to real price volatility.

A clear avenue for future research is to extend the same principle behind monetary *q* to cryptocurrencies. Cryptocurrencies behave much in the same way and have a cost of mining per unit. Nevertheless, data is difficult to get and mining rewards are highly unstable if new miners enter. One of the main differences between gold and cryptocurrencies is that if gold mining would cease, gold would continue to exist, whereas if cryptocurrency mining (e.g., Bitcoin mining) would cease, the underlying asset would also cease to exist.

Section V: Capital in Disequilibrium: A Reformulation of the Austrian Business Cycle Theory

Chapter 16: A Reformulation of the Austrian Business Cycle Theory

The 'Austrian business cycle theory' is in fact a theory of how capital gets distorted and no longer reflects final consumer preferences. The business cycle theory is, in essence, a theory of capital distortion.

One of the most grandiose syntheses of the 'Austrian' theory is Huerta de Soto's magnum opus, *Money, Bank Credit and Economic Cycles* (2006), which opened an entirely new field of investigation to a once forgotten and mistreated subject.

In this line, we extend the work of Huerta de Soto (2006) on business cycle theory, expanding it by integrating it with the term structure of interest rates (the *yield curve*), the practice of maturity mismatching, and the theory of (equity) q. The theory of the business cycle presently proposed is therefore an intent to improve and build upon the work that already exists.

The version of the 'Austrian' business cycle theory presented in this chapter also adds to the robustness of the theory, as (nominal) increases in money or credit supply (e.g., Huerta de Soto, 2006) or central bank reserves (e.g., Selgin, 1988) are no longer required for a business cycle to occur or to explain past recessionary episodes. Moreover, it will be easier for empirical studies to back the 'Austrian' explanation of recessions and crises.

As a side effect, we also offer a coherent theoretical justification of the commonly observed inverted yield curve (inversion of the term structure, in which short-term interest rates briefly exceed long-term interest rates). Indeed, a large body of empirical literature shows that the yield curve is the most informative predictor of a crisis (e.g., Bernanke, 1990; Estrella & Mishkin, 1996). The fact that the term structure and, specifically, an inverted yield curve signals an economic downturn has been well-documented.

The trigger that sets in motion a business cycle is closely related to one of the most important prices in the economy: the rate (or, more aptly put, the rate*s*) of interest. The business cycle is an *intertemporal* distortion in the capital structure: it consists of a misalignment between the productive structure of society and the time preferences of savers, caused by the network of financial intermediaries that exists in between. As Huerta de Soto (2006) explains, if the rate(s) of interest decline(s):

> "(…) then the market price of capital goods and durable consumer goods will tend to increase. Moreover it will tend to increase in proportion to the duration of a good (…). Capital goods already in use will undergo a significant rise in price as a result of the drop in the interest rate and will be produced in greater quantities." (p. 325)

What Huerta de Soto (2006) does not mention in this segment, is the fact that the prices of financial assets rise much faster than the prices of productive assets (capital goods): financial claims are traded every day on public markets, yet the underlying assets are only bought and sold every once and a while. In these instances, entrepreneurs will *eventually* bid up the prices of the productive assets, yet the adjustment period is rather long. This is what the equity *q* ratio, for instance, indicates. In that sense, it is noteworthy that Huerta de Soto (2006) shows that:

> "(…) uninterrupted stock market growth never indicates favorable economic conditions. Quite the contrary: all such growth provides the most unmistakable sign of credit expansion unbacked by real savings, expansion which feeds an artificial boom that will invariably culminate in a severe stock market crisis." (p. 462)

This is very much in line with our idea of *q*. Financial asset prices move ahead of entrepreneurs bidding up the productive assets' prices. Moreover, a fall in the rate of interest, suddenly turns previously unprofitable projects profitable. Hence, marginal investment projects will be undertaken if the rate of interest and cost of capital decreases. Again, in Huerta de Soto's (2006), words:

> "[T]he fall in the interest rate will reveal that many production processes or capital goods which until then were not considered profitable begin to be so, and consequently entrepreneurs will start to introduce them. In fact in the past entrepreneurs refrained from adopting many technological innovations and new projects because they expected the cost involved to be higher than the resulting

market value (which tends to equal the value of the estimated future rent of each capital good, discounted by the interest rate)." (p. 326)

To sum up, a decline of the rate of interest, as Huerta de Soto (2006) outlined beautifully, has three important consequences for markets:

(1) Asset prices go up (that is, the value of capital goes up), since future cash flows are now discounted at a lower rate of interest. In Huerta de Soto's (2006) words: "The *market value* of [a] capital good tends to equal the value of its expected future flow of rents, discounted by the interest rate. An inverse relationship exists between the present (discounted) value and the interest rate." (p. 325).
(2) Financial asset prices go up first. As indicated by *q*, financial asset prices adjust (almost) instantaneously, while productive asset prices (capital goods) only adjust to the extent that they are bid up by (nonfinancial) entrepreneurs.
(3) Asset prices of assets and investments with *longer durations* go up proportionally more than assets with *shorter durations*. This is due to the fact that at a lower rate of interest previously unprofitable ventures suddenly become profitable (and feasible) to begin.

Nevertheless, interest rates can fall either because of real changes in supply and demand of ultimate resource providers (saver) and resource users (entrepreneurs) or because of distortion in the capital structure induced by financial intermediaries. In light of the latter, recessions are a violent realignment of the time composition of capital to the time preferences of consumers.

While we are aware that there are different monetary theories that address the how behind the distortion in the capital structure, we will simply leave the debate for another moment and outline our own theory of how the cycle is spawned. There is, generally, agreement among defenders of the 'Austrian' theory that a business cycle involves an intertemporal mismatch between consumers' future consumption intentions and plans and producers' production plans. Intertemporal, as both consumption and production are planned for with a great regard of time.

We could say, however, that capitalists (savers) express their time preferences by holding the maturity of their liking. That is, if a capitalist would be willing to part for ten years with (part of) his income, he would simply buy a ten-year bond or pick any other investment with a 10-year maturity. Hence, if capitalists decide to hold more *liquid* investments, for instance demand deposits, they are essentially expressing that they wish to hold optionality over what they have or even specifically wish to consume in the near future, in short, that they are focused on the short term rather than the long term. They also might see few "investment opportunities," which could induce capitalists to keep liquid assets in a bank account or any other liquid investment (money market fund shares, for example). The point here is that when capitalists save part of their income as zero maturity bank liabilities, they are communicating their time preference.

The trouble is thus that banks take a large part of these liquid savings and invest them in the long term. This is problematic, because:

(1) They increase investments at longer *durations* relative to shorter durations;
(2) As loans are spent, accounting profits (returns) will rise for longer *durations* where the flow of investment ends up, which in turn shifts additional resources to longer *durations* (both by attracting investors and by reinvesting retained earnings), while average returns will remain unchanged for shorter durations.

Moreover, long-term demand for present goods is quite elastic: lower the rate of interest a little bit, or consider marginal borrowers and lower lending standards, and you can invest funds at ten years without much of a hassle. This partly explains why *average* 10-year yields are relatively stable, even when commercial banks begin arbitraging the yield curve: the real story is partly hidden by the fact that 10-year rates for non-marginal borrowers decline while new, less creditworthy marginal borrowers are charged more (but now have access to such funding, something that under earlier conditions was not the case).

The latter explains why corporate risk spreads narrow heavily in the boom phase. Not every 10-year yield is made equal: there are US Treasuries, AAA/AA-rated, A-rated, BB-rated and junk bonds available at 10-year maturities. When corporate risk spreads narrow on long-term debt, it is clear that financial intermediaries are arbitraging the yield curve[188]. More than a radical lowering of the 10-year rate on, for instance, US Treasuries, it lowers substantially more 10-year rates of less creditworthy borrowers.

However, this phase of exuberant economic expansion inevitably comes to a halt, if the reduction in long-term rates came off the back of yield curve arbitraging, liquidity arbitraging,

[188] What market commentators call a "hunt for yield" is one of the many characteristics of the business cycle.

or maturity arbitraging[189]. Three things occur that lead to upward reversal in WACC, after banks arbitraged the yield curve and brought down WACC:

(1) One of the first effects that occurs, is that savers will – on the margin – save less. Indeed, the personal savings rate *tends* to decrease over the expansion phase as they are being confronted with lower rates of interest[190].

(2) A second effect is that long-term investment will go up (in *productive assets*, as *q* indicates), which inevitably raises the prices of (a) wages and consumer goods, (b) commodities, (c) producer goods, (d) real estate or any combination of the prior. This leads to price pressure that reduce *ex post* profits of *ex ante* expected profits, even leading to outright losses.

(3) A third effect is that depositors begin withdrawing and liquidating savings to either consume or otherwise dispose of their bank holdings. This leads, marginally, to a loss of deposits for the bank, which will on the whole forces the bank to either raise interest rates to attract depositors or dispose of assets (thus increasing supply of loans and securities elsewhere). Short-term interest rates, therefore, rise.

Hence, the reversal in the business cycle occurs when WACC experiences an upward reversal and that upward reversal can come from either higher borrowing costs or higher input prices. Eventually, this increase will lead to an inverted yield curve (or, at the very least, to an extreme narrowing of the yield curve spread).

At this point, pressures will gradually build (either by higher input prices[191] or by higher short-term interest rates) until the cycle turns and a recession sets in. A liquidity crunch ensues. At this point, there is a deep correction on stock markets (*q* ratios fall) and in other financial asset prices, and a widespread liquidation sets in which leads to the usual symptoms: a cluster of defaults, an increase in unemployment, large write-offs on capital and distress on capital markets (especially among commercial banks).

---

[189] We consider all three different sides of the very same coin.

[190] The decrease in savings is a general tendency, not an economic law; it is one of the possible factors that begins the process of readjustment.

[191] Of course, central banks, generally when faced with an increase in prices, begin contracting monetary policy as well. However, even if a central bank would not do so, price inflation would trigger the crisis (even if price inflation would *not* lead to higher rates).

We summarize the different stages of the business cycle as follows:

Expansion

S1. Savers/capitalists decide to hold liquid, short-term debt (instead of holding a greater part in illiquid, long-term investments).

S2. Banks begin arbitraging the yield curve by expanding the maturities of their loans and investments, financed by liquid, short-term debt (mostly demand deposits).

Boom

S3. Long-term interest rates fall and the yield curve spread begins to narrow.

S4. Financial asset prices rise; q ratios go up.

S5. At lower long-term interest rates, businesses begin (on the margin) to invest in projects with longer *durations* that are more illiquid.

S6. Long-term (capital or fixed) investment increases and pushes up (marginally) the prices of productive assets (including commodities), the *average* or aggregate rate of profit (across the economy) peaks.

Pre-crisis

S7. Savers/capitalists begin to liquidate their holdings and (on the margin) consume savings (the savings rate drops). Either the prices of present goods rise or short-term interest rates begin to increase as banks are faced with decreasing demand for their deposits.

S8. On the margin, businesses begin to default due to *ex post* losses due to price increases or higher rates of interest (in short, higher WACC).

S9. The yield curve spread becomes negative (the yield curve inverts).

Crisis

S10. Financial asset prices collapse and q ratios go down (possibly below their equilibrium point due to liquidation of capital).

S11. Massive liquidation and liquidity crunch: defaults mount, weak lenders and marginal overleveraged borrowers go bankrupt, workers are fired and capital is destroyed.[192]

Recovery

S12. The time preferences of consumers are realigned with the financial maturity structure and the productive duration structure.

S13. Once the readjustment has run its course, if financial intermediaries begin arbitraging the yield curve again, the cycle repeats.

[192] If and when governments try to assume the systemic illiquidity of the financial sector by, for example, bail-outs and nationalizations, a public debt crisis might be the result, as the 2013 European debt crisis has convincingly proved.

Yet, surprisingly, the liquidity crunch or squeeze is a mere process of realignment, of readjustment: a reestablishment of consumer sovereignty. The maturity structure of the economy is violently aligned again with the time preferences of consumers.

Hence, the 'Austrian' business cycle theory is a theory of illiquidity, of a mismatch or a misalignment of consumer time preferences and the *duration* of production. In Huerta de Soto's (2006) words: "[T]he recession always originates from an absence of the voluntary saving necessary to sustain a productive structure which thus proves too capital-intensive. The recession is caused by the credit expansion the banking system undertakes without the corresponding support of economic agents, who in general do not wish to augment their voluntary saving." (p. 399).

Despite Huerta de Soto's (2006) tremendous contributions to the field, we want to close our proposed business cycle theory with a brief comment on Hayek's (1932) and Huerta de Soto's (2006) use of the 'Ricardo-effect' over the course of the business cycle. This notion is very much related to capital theory: as Hayek (1941) distinguishes "capital" as "produced means of production" from "original" factors of production, specifically, labor and permanent resources such as land, the Ricardo effect treats workers and labor as homogenous goods. The "Ricardo effect" describes how in the expansion phase of the business cycle labor is substituted with capital goods, which reverses in or after the recession when wages fall.

Nevertheless, human capital is very diverse, in the sense that the economic behavior over the course of the business cycle of a highly skilled and trained professional (who, generally, made a large investment in his education or training) resembles much more the behavior of a complex piece of equipment than it resembles the dynamics of unskilled labor. Hence, isolating "labor" from "capital" is meaningless. It is the cash flows related to investment in human capital (both the outlays and future wages to be earned) discounted at a rate of interest that are of importance, just as any other economic good.

Take the training of astronaut. It is suggested that training one astronaut costs anywhere in the neighborhood of $15 million dollars and approximately six years of training. Hence, the cash flows of an astronaut (or contribution to cash flows) only starts in year seven after a $15 million-dollar investment. It is obvious that interest rates play a very big role in whether such an investment is deemed profitable. In this sense, the astronaut has more in common with an oil well than a cashier.

This means, moreover, that the 'Ricardo effect' is not generalizable as a phenomenon over the business cycle. CEO pay, for instance, even excluding stock option compensation, follows a much more cyclical pattern. In a certain sense, Hayek (1932) unfortunately seems swayed by the classical triad of capital, labor and land and the 'Ricardo effect' is a by-product of the material line of thought. As a result, many nuances are lost.

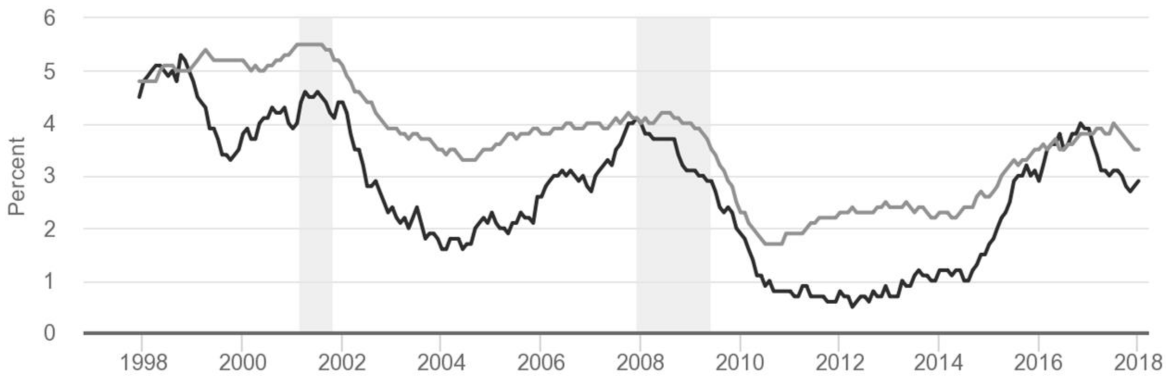


Figure 83: Wage rate growth in leisure and hospitality industry (dark curve) versus trade and transportation industry (source: Atlanta Fed)

Chapter 17: A Review of Historical Crises and the Role of Maturity Mismatching

17.1 The Great Depression of the 1930s

The Great Depression is perhaps the most interesting outcome of any historic business cycle so far, given its devastating consequences and its global scale. While there is much to the Great Depression that lies far beyond the scope of this study[193], we will focus on an overlooked part of the story of the run-up to the Great Depression: the degree of maturity mismatching.

There are a few important peculiarities to the Great Depression. First, liquid secondary markets were becoming increasingly sophisticated. Second, the *shiftability* theory gained popularity and justified a rotation from short-term commercial credit to long-term corporate, government and consumer debt[194]. Third, many (new) durable consumer goods were introduced in the 1920s, such as cars, vacuum cleaners and washing machines, which were financed with (long-term) consumer credit[195].

The gradual shift toward the *shiftability* theory (and the subsequent retirement of the liquidity theorem) turned out to have an important historical impact on monetary history as one of the key causes of the Great Depression. As Glock (2017) explains: "These new ideas on the importance of the long-term rate did not remain long in academic ivory towers, and began to influence the Federal Reserve beginning in the late 1920s. (…) [An] analysis of Federal Reserve discussions and actions in the late 1920s and early 1930s, showed that much of the system focused on managing and lowering the long-term interest rate." (pp. 10-11).

---

[193] For instance, the 1930 Smooth-Hawley Tariff, which imposed import tariffs on over 20,000 goods in the U.S., had devastating consequences for global trade and led to an rapid impoverishment at a scale that has rarely been seen before. However, an analysis of such reactions to a crisis is beyond the scope of this present work.

[194] We gave a brief overview of the shiftability debate on p. 132.

[195] In this case, we refer to "long-term" as anything that exceeds twelve months to consider it "current." Normally, consumer financing was at maturities ranging from one year and a half up to five years.

In addition, there was a rather large "shadow banking" system that is often not accounted for in the 1920s in the U.S. Two-thirds of all of all commercial banks was *not* a member of the Federal Reserve and as such did not form part of the Federal Reserve System on the eve of the Great Depression in 1929 (Mitchener & Richardson, 2012). These banks, all together, made up approximately half of all deposits. Nevertheless, they relied on interbank loans of Fed member banks to comply with state-level reserve requirements. A majority of non-member banks who depended directly on member banks were located outside the big cities (Mitchener & Richardson, 2012). Non-member banks would also include savings and loan associations (S&Ls). The early S&L's of the 1920s, however, were very antifragile since member shares were pretty much equal to equity (Pyle, 1995).

These non-member banks behaved very similar to the so-called "banking trusts" of the 1907 crisis, which eventually led to the creation of the Federal Reserve. These "banking trusts" were *de facto* banks that, however, depended on a (single) commercial bank for access to the national clearinghouse (which was the New York Clearinghouse which is still in existence as of today). The only difference with the pre-Great Depression period is that instead of the private New York Clearinghouse, now the Federal Reserve was directly in charge of reserves. This matter complicates gathering data on banks' maturity mismatches, since the degree of maturity mismatching could just as well be much more extreme in the case of the "shadow banks" compared to the more broadly referenced member banks.

So, is there any data that could suggest that commercial banks were, in fact, engaging in maturity mismatching?

Baum & Thies (1989) demonstrated that in the period preceding the Great Depression of the 1930s (1927 to 1929) the yield curve was flatter than during the Great Depression (1929 to 1932). In the pre-depression period, short-term rates were higher and long-term rates were lower, which makes perfect sense if one realizes that banks were engaging in maturity mismatching. Again, commercial banks arbitrage the yield curve by borrowing short and lending long, which leads, *ceteris paribus,* to a narrowing of the yield curve spread.

During the depression, Baum & Thies (1989) observed that: "The changes from 1931 to 1933 (…) were unprecedented. (…) Private-sector loans, mortgages and bonds were liquidated in favor of cash and secondary reserves. This shift to cash and secondary reserves (…) [lowered] rates on liquid securities and [raised] rates on illiquid securities" (p. 492).

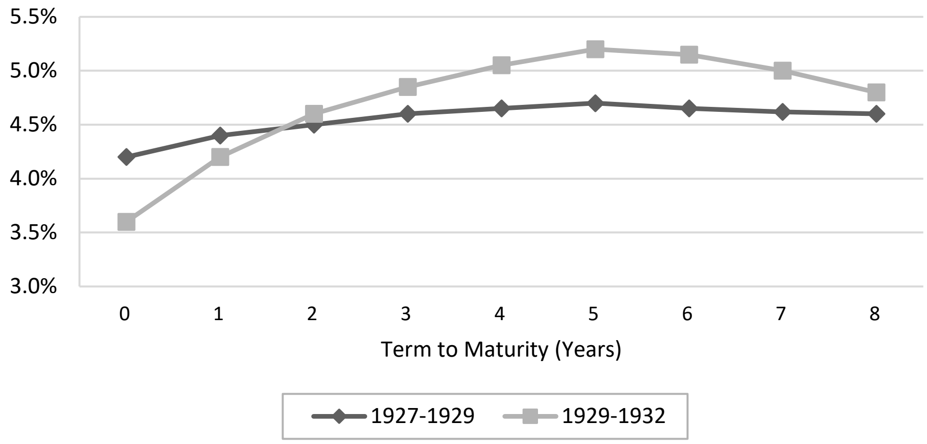


Figure 84: Average term structure of interest rates of 1927-29 and 1929-32, data from Baum & Thies (1989)

More than anything else, there is substantial anecdotal evidence with regard to maturity mismatching in the run-up to the Great Depression. Lieber (1931), for instance, stated:

> "The inconsistency of considering demand money, money that has been invested in mortgages repayable by amortization over a period of 15 to 20 years is apparent. Whatever the desires or preferences, however altruistic the attempt to have money invested in building and loan associations payable on demand, the theory of long-term mortgage and demand funds will no more work than will oil and water mix." (p. 274).

With regard to consumer durables bought on credit, Eichengreen & Mitchener (2003) remark: "consumer debt as a percentage of personal income doubled from 4½ per cent in 1918-20 to more than 9 per cent in 1929." Consumers were thus leveraging up, yet the banks that would ultimately finance such leverage would themselves be financed with short-term (zero maturity) demand deposits. According to Eichengreen & Mitchener (2003), almost 9% of consumer spending in the pre-Great Depression period was spending on consumer durables. Of that 9%, roughly 70% was financed with consumer credit. This boom in consumer durables, financed with callable bank deposits, was partly responsible for the (artificial) economic boom of the 1920s.

To sum up, we have proposed a different take on the Great Depression, based on both bank and non-bank maturity mismatching of *zero maturity* savings invested in credit related to consumer durables (which would include housing). Both banks and nonbanks arbitraged the yield curve by borrowing short and lending long. This led to a narrowing of the yield curve spread in the years prior to the Great Depression.

Even though our historical case study of the Great Depression is preliminary, we do have reason to suspect that maturity mismatching played an important role in the most severe recession in history. A more extensive study on the Great Depression in the context of the term structure of interest rates and maturity mismatching by financial intermediaries would, however, be necessary.

17.2 The Savings & Loan (S&L) Crisis of the 1980s

The S&L crisis of the 1980s was severe: the direct cost of the crisis equaled an astonishing $124 billion dollars. Close to a third of all S&L Associations were wiped out. The Dow Jones Industrial Average fell 20% from 1980 to 1982, with most of its decline clustered in 1981. The unemployment rate rose to 11%. The downturn lasted for 16 months. The crisis began when Fed Chair Paul Volcker raised interest rates aggressively to counter the high and escalating rate of inflation. By pushing up short-term interest rates, Volcker inverted the yield curve. It is generally forgotten that the 1981-82 recession was prior to the Great Recession of 2008 the worst crisis since the Great Recession.

Historically, S&L Associations were very stable. As early as in 1831, these mutual associations arose to fund housing projects. Funds were pooled by members (which received shares) and members made regular contributions. The proceeds were invested to finance mortgages. Initially, the shares in S&Ls were equity, with no fixed interest payments or right to withdraw. Since S&Ls were very local, members knew each other and allowed for an efficient way to select and monitor borrowers. Moreover, since the members' participations were shares, no possible liquidity problems can arise and losses are fully assumed by the S&L's members. Nonetheless, little by little, shares were turned into semi-bonds, with fixed dividends (effectively interest payments) and maturities, which could also be withdrawn if necessary (albeit at the expense of penalty). Mortgages typically had a maturity of approximately 11 years and a duration of about 7 years (Morton, 1956).

Gradually, Savings & Loan Associations were engaging in maturity mismatching by expanding the share of short-term deposit liabilities to other liabilities and equity. After the New Deal, however, average maturities shot up to roughly 15 years (Pyle, 1995). Many safeguards were put into place by the Roosevelt administration to protect mortgage lenders.

Yet, with interest rates on long-term mortgages locked in for years, S&Ls were extremely vulnerable to changes in interest rates, since their liabilities essentially roll over daily. Any increase in interest rates would increase its interest expense without increasing its interest income. Curiously, as short-term interest rates already began to rise in the 1960s, the interest rate S&Ls could pay on deposits was capped. This, more than anything, sustained the unsustainable degree of S&L maturity mismatching.

Goodhart & Perotti (2015) would draw a parallel with the 2008 crisis[196]: "The mortgage business ensured banks a steady cash flow, funded largely at the short-term rate. But it also represented a dramatic increase in maturity mismatch. **This construction was at the heart of the Savings and Loan crisis in the 1980s, caused by a sharp rise in interest rates**. Its format was also replicated in the massive expansion in shadow bank operations during the credit boom. In many countries banks managed to set such lending outside their balance sheet, on the pretence that these entities were bankruptcy remote. Investment banks also pursued a related form of shadow banking, with massive holdings of securities based on long-term mortgages funded mostly by short-term repos, at the extreme overnight." (p. 2).

Ely (1991) commented on the S&L crisis: "*Borrowing short to lend long* was the financial structure that federal policy effectively forced S&Ls to follow in the aftermath of the Great Depression. S&Ls used short-term passbook savings to fund long-term, fixed-rate home mortgages. Although the long-term, fixed-rate mortgage may have been an admirable public-policy objective, the federal government picked the wrong horse—the S&L industry—to do this type of lending because S&Ls funded themselves primarily with short-term deposits. The dangers inherent in this "maturity mismatching" became evident every time short-term interest rates rose. S&Ls, stuck with long-term loans at fixed rates, often had to pay more to their depositors than they were making on their mortgages. In 1981 and 1982 the interest rate spreads for S&Ls (the difference between the average interest rate on their mortgage portfolios and their average cost of funds) were −1.0 percent and −0.7 percent, respectively." (p. 2).

Ely (1991) furthermore argues that government made a big mistake by allowing losses to accumulate in insolvent S&Ls backed by taxpayers. In effect, U.S. Congress kicked the can further down the road at a huge cost to the taxpayer. Insolvent S&L Associations were not

[196] We will discuss the 2008 crisis further below, see p. 201.

closed after the 1981-1982 crisis, which led to an enormous accumulated cost. Half of these eventual losses are due to delayed closures (Ely, 1991).

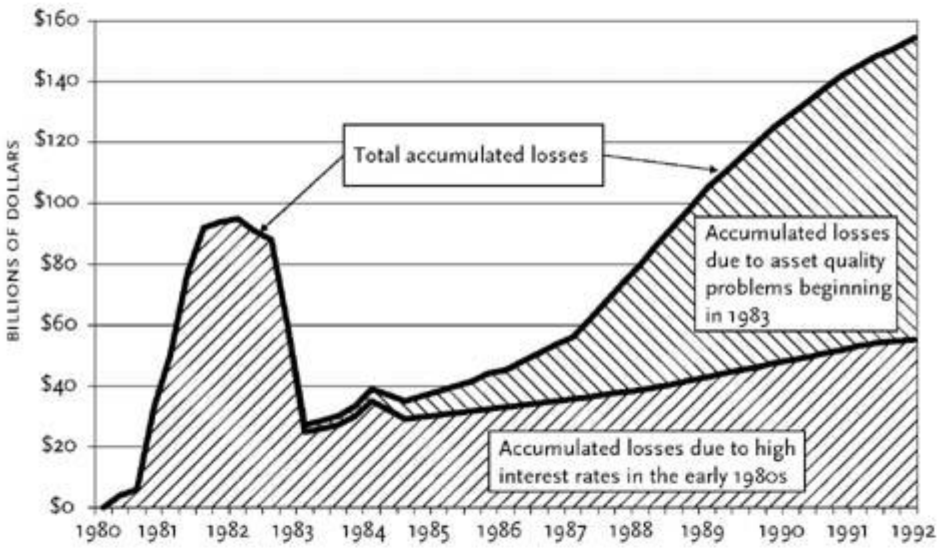


Figure 85: The losses of S&Ls in the 1981-82 recession and later losses due to a lack of decisiveness on the part of the Carter administration (Ely, 1991)

As Pyle (1995) adds:

> "By 1981, the strategy of housing construction support and S&L cost containment through the use of deposit interest rate ceilings was in complete disarray. (…) [T]he S&L's average interest cost had risen to 10.71% (…). In market value terms, it has been estimated that aggregate S&L net worth which in 1966 was around -$2.8 billion fell to about -$100 billion by 1981. By contrast, Federal Savings and Loan Insurance Corporation reserves were $1.8 billion in 1966 and $6.2 billion in 1981. The misguided and ultimately failed attempt to shelter the structurally defective S&L industry from the effects of interest rate risk had convert an unfunded federal liability from about $1.0 billion to one of over $90 billion. (…) Variable interest rate mortgages (VRMs) provided an alternative method of reducing the maturity imbalance faced by S&Ls (…). Despite periodic support for VRMs by the Federal Home Loan Bank Board, Congress throughout the 1970s steadfastly opposed VRMs for federally-chartered institutions, belatedly authorizing their use in 1979." (pp. 15-16).

So, what does the data tell us?

Let us first take a look at the yield curve. We will observe a similar pattern as in the other historical episodes that we review in this chapter. We gathered data from the St Louis Fed and, specifically, selected the 10-year US Treasuries minus the 2-year US Treasury rate to calculate our yield curve spread:

Figure 86: The yield curve spread in the run-up to the 1980-82 recession.

In the above chart, we are able to observe a familiar pattern: the yield curve is arbitraged since the beginning of 1976 with the yield curve spread narrowing. In 1975, the U.S. economy was barely coming out of another recession. Yet, the conditions in the high-inflation environment of the 1970s enabled another brief cycle.

While the S&P 500 suffered a 20% drawdown in 1981, our equity *q* ratio failed. Most of it can explained by, again, the data: the high inflation of the late 1970s led to an understating of the measured replacement value of capital. As a result, equity *q* gave the impression of undervaluation but it was likely that *q* would have been much higher if replacement value was well-adjusted for inflation.

17.3: The Asian Financial Crisis of 1997-98

The 1997 Asian financial crisis affected many countries and had global repercussions. The crisis took a large toll on the economies of Thailand, Indonesia and South-Korea and to a lesser extent of Hong Kong, Taiwan, Vietnam, China, Singapore, Malaysia, Laos and the Philippines. Local currencies (sometimes pegged to the dollar) were hammered in international foreign exchange markets, stock market and asset prices collapsed and many governments were on the brink of default (or even ended up in default). The Asian financial crisis of 1997 eventually led to a marked decline in the price of oil, which triggered the 1998 Russian financial crisis, which in turn led the world-famous hedge fund Long-Term Capital Management to bankruptcy. While the U.S. largely avoided a severe economic downturn (the bursting of the Dotcom bubble would come later), major economies such as Brazil and Argentina were left in shambles.

The Asia contagion crisis was mostly attributed to dollar appreciation and currency mismatching (with Asian commercial and central banks that issue dollar-denominated debt to invest in local-currency-denominated assets (e.g., loans denominated in Thai baht). Moreover, many central banks were either explicitly or implicitly pegging their currencies to the U.S. dollar, which magnified the negative impact of an appreciating dollar. Many central banks were losing incredible amounts of foreign exchange reserves in attempt to stabilize or prop up the value of their currencies in international foreign exchange markets. An often overlooked and arguably more important issue, however, were the maturity mismatches in the underlying Asian banking systems, rather than the more “obvious” currency mismatches. Our thesis in this chapter is, therefore, that the fundamental cause of the Asian financial crisis of 1997-98 was maturity mismatching, rather than currency mismatching.

In fact, from a theoretical point of view, currency mismatches have no reason to be a life-threatening and abrupt problem *if* maturities are matched. Let us assume, for instance, a bank that has a present currency mismatch, but not a maturity mismatch:

Figure 87: An example bank balance sheet with USD-denominated liabilities and THB-denominated assets (THB is the acronym of the Thai currency, the baht)

Assets (in THB)	**+10%**	**Liabilities + equity (in THB)**	**(-5%)**
THB Loans (10-years)	30,000	USD Deposits	5,000
THB Reserves	5,000	USD Bonds (10-year)	25,000
		Capital (equity)	5,000

In the above example, there are no maturity mismatches. The bank issued 30,000 in THB-denominated loans at a 10-year maturity (let us assume, for simplicity sake, that maturity and duration are equal), financed with 5,000 in equity and 25,000 in USD-denominated bonds issued by the bank on foreign capital markets. Moreover, the bank has 5,000 in USD-denominated demand deposits, backed by 5,000 in liquid THB-denominated cash reserves. Let us, moreover, assume that the bank earns an average 10% on its asset-side and spends on average 5% in interest on the USD-liabilities it issued.

Now, let us assume that, as a shock, the THB depreciates 20% against the dollar (which simply means that the USD-denominated liabilities appreciate 20% in THB-terms). This results in:

Figure 88: An example bank balance sheet with USD-denominated liabilities and THB-denominated assets after *a 20% devaluation.*

Assets (in THB)	***+10%***	**Liabilities + equity (in THB)**	***(-5%)***
THB Loans (10-years)	30,000	USD Deposits	6,000
THB Reserves	5,000	USD Bonds (10-year)	30,000
		Capital (equity)	-1,000

In this example, the USD-denominated debt increases in terms of THB: the debt burden increases by 20% overnight. Moreover, the previous cost of credit of 5% jumps as the principal (in THB-terms) increases. In effect, the interest expense goes up from 1,500 to 1,800 and the cost of credit goes up from 5% to 6%.

While it is clear that this change would lead to balance sheet *insolvency*, it does not follow that it would lead to *illiquidity*. Even if deposit holders would withdraw every single deposit, it can make whole on such liquid promises by liquidating its reserves and part of its loan portfolio. Moreover, the net interest margin (NIM) remains positive, which would allow the bank to make up for the foreign exchange loss in the remainder of the maturity. While it may be stating the obvious, no real-world bank would finance a 100% of its balance sheet in a foreign currency. Goldstein and Turner (2004) suggest that at its peak (in 1997), the effective currency mismatch equaled 12% (net exposure to foreign currencies), while it rose from slightly negative in 1994 to 6% in 1996. Overall, currency mismatching as such could not explain the Asian financial crisis and much less its incredible rippling effect across countries.

In conclusion, maturity mismatches are more important than and primary to currency mismatches. Maturity mismatches lead to uncontrollable liquidity runs, currency mismatches by themselves would only cause insolvency, which – while still problematic – does not explain the dynamics of the 1997-98 Asian financial crisis. As we will see and apply our business cycle theory, we could conclude that maturity mismatching *caused* the crisis while simultaneous currency mismatching *intensified* the crisis.

Indeed, this is how we could characterize the 1997-98 crisis. Slightly adapting the stylized balance sheet of our example to the conditions during the 1997-87 crisis, we can get a quick glimpse into the fragile nature of the East-Asian banking system[197] (Yoshitomo & Ohno, 1999):

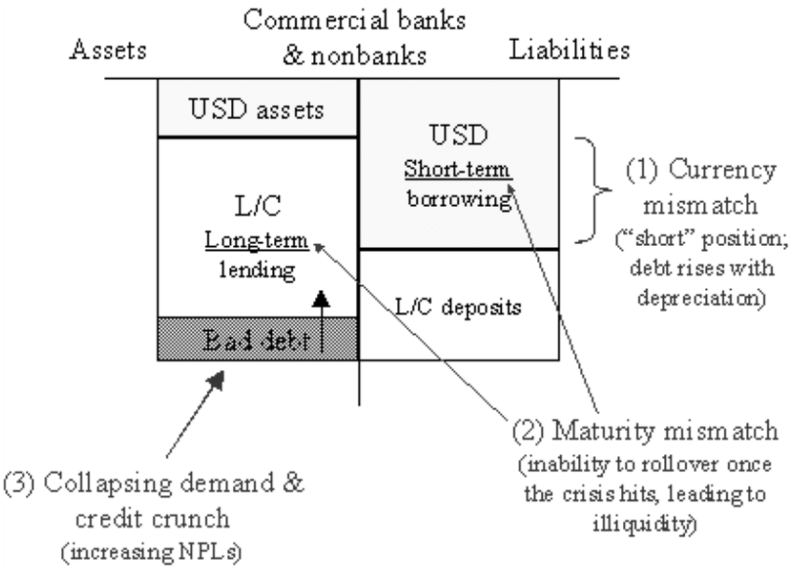


Figure 89: The 1997-98 Asian crisis was characterized by three key factors: maturity mismatches, currency mismatches and increasing non-performing loans (Source: Yoshitomo & Ohno, 1999).

Contrary to what many think, governments such as the Korean government, discouraged long-term borrowing and created perverse incentives for short-term borrowing. As Kihwan (2006) explains: "[The Korean] government in effect discouraged long-term foreign borrowing by business firms as it required detailed disclosure on the uses of the funds as a condition for its permission. On the other hand, short-term borrowing was mainly regarded as trade-related financing requiring no strict regulation. These *de facto* incentives for short-term borrowing led banks and business firms to finance long-term investments with short-term foreign borrowings." (p. 5). Before 1994, capital controls inhibited foreign borrowing: after

[197] What Yoshitomo and Ohno (1999), like many others, fail to realize is, however, the fact that maturity mismatching triggers a cycle that inevitably leads to an increase in non-performing loans. They cannot be separated from each other.

1994, however, only restrictions on short-term overseas borrowing were lifted, while restrictions on medium and long-term borrowing from abroad were maintained (Smith, 2000). This led Korean banks to invest 80% of short-term foreign credit into 70% of long-term loans (Kihwan, 2006). Furthermore, policies by the Bank of Korea biased “borrowing toward the short-end of the term spectrum” (Noland, 2007, p. 501).

Another distortion was caused by the Basel I accord: short-term loans (less than a year) by OECD-banks to non-OECD banks received only a 20% risk-weighing, whereas long-term loans (longer than ten years) received a 100% risk-weighing[198].

Additionally, as Smith (2000) writes: “Merchant banks and insurance companies (…) were a major source of short-term liquidity to corporations. By mid-1997, there were 30 merchant banks, many established between 1994-6. These institutions borrowed large amounts of short-term loans in international capital markets and then re-lent onto firms on a long-term basis, taking advantage of differences in short and long-term rates.” (p. 122). The latter is evidenced in the data on gross fixed capital formation.

Chang and Velasco (2001) try to understand why the 1997-98 crisis was mostly limited to Asian countries, which experienced huge outflows of capital. Chang and Velasco (2001) compare Asian with Latin-American countries, since they were affected similarly by the appreciation of the dollar and was not fundamentally sounder than the Asian economies. They found the degree of maturity mismatching to be one of the main differences: “[T]he Asian countries were in a situation of international illiquidity evidenced by sharply rising ratios of hard currency short-term liabilities to liquid assets.” (p. 3). According to Chang and Velasco (2001), much of the Asian borrowing was, “(…) especially in the period right before

---

[198] Interestingly, the South-Korean financial crisis began in November 1997, followed by an IMF-organized bailout (in fact, the largest bailout package ever up till that day) of $55 billion (the IMF itself contributed $21 billion. South-Korea would join the OECD in December 1997, effectively eliminating the regulatory bias that Basel I caused toward short-term (international) borrowing. The unintended consequences of top-down regulatory intervention can have devastating effects. However, the people responsible for creating such rules and regulations, do not bear the cost of their bad decisions. Moreover, crises are then perceived as “market failures,” when in fact they are often induced by bad regulations and government policies.

the crisis, short term" (ibid). The need to roll over large amounts of short-term debt meant that the Asian economies were extremely fragile.

The data, however, is hard to come by in the case of the affected Asian economies. Many were in the process of opening up their (capital) markets; the Asian banking industries were in their infancy. The difficulty in getting reliable data on long-term and short-term (market) interest rates, for instance, is high. Of the countries that were most affected (Thailand, Korea and Indonesia), the best data is available for Korea. Therefore, we will focus for the remainder of this chapter on South-Korea, with the assumption that Korea is representative for the other Asian countries.

To get more information about how the term structure of interest rates (the yield curve) has behaved over the course of the Asian business cycle, we tried to approximate the yield curve by using various inputs. Since most short-term borrowing was in U.S. dollars, I decided to take the 12-month (U.S. dollar) LIBOR as a proxy of short-term borrowing costs. Local short-term interest data was contradictory, in the worst of cases, since they did not reflect true market conditions or, at the very least, irrelevant since the driver of this business cycle was mostly the (international) short-term borrowing in U.S. dollars at much lower yields. Moreover, much data on long-term bond yields is unreliable as well. I selected the long-term Korean housing bonds as proxy for long-term rates[199]. Nonetheless, to account for the exchange rate difference, I adjusted the proxy of the long-term bond yield for (core) inflation, as reported by the OECD. As a last step, to derive the yield curve or term structure *spread*, I simply subtracted the 12-month LIBOR from the Korean housing bond yield adjusted for inflation. The result is the following chart:

[199] Korean "Housing Bonds" are issued by the government, but may only be used according to the National Housing Fund Act.

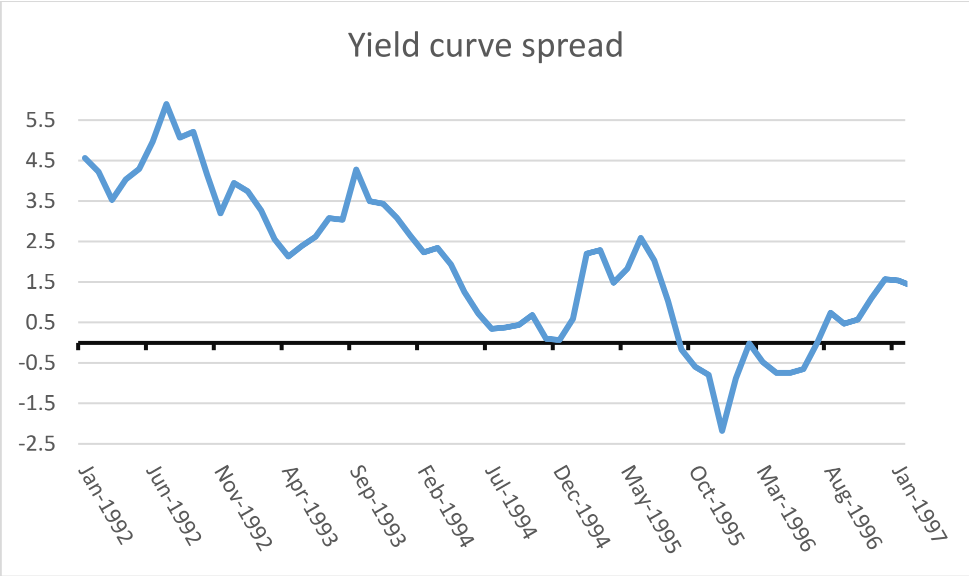


Figure 90: A proxy of the yield curve spread in Korea from 1992 to 1998 prior to the Asian contagion crisis.

What we can appreciate from the above yield curve spread, is that the observed behavior of the term structure coincides with our theory of maturity mismatching. In 1992, financial intermediaries appear to begin arbitraging the yield curve by increasingly relying on short-term funding to invest in the long term. As such, the yield curve spread gradually narrows, until in the beginning of 1995 the yield curve almost inverts[200]. Nonetheless, the yield curve truly inverts from late 1995 to late 1996, which indicates a crisis is near.

In the first months of 1997, the first cracks begin to appear. One of the largest Korean business conglomerates, Hanbo, filed for bankruptcy in January 1997. Seven other large businesses, with Kia Motors being the most notable, were facing insolvency. Non-performing loans of Korean commercial banks also rose steeply. By September 1997, the narrowing of the yield curve and tightening of conditions led to a doubling of the amount of non-performing loans in merely half a year (Lee, 1999). A devastating banking crisis was the result.

It is important to note that the sudden and devastating foreign exchange losses only arose *after* the crisis, which severely worsened the crisis. In the case of Korea, exchange rates began their real slide in November 1997. Currency depreciation was, thus, a *symptom* (albeit important) rather than a *cause* of the 1997-98 Asian financial crisis.

[200] This was the year of the Mexican "Tequila" crisis, which led to a bail-out in 1995.

The real crisis erupted in November 1997, when the Korean won depreciated by twenty percent against the dollar and the Korean stock market fell by roughly 30 percent, reaching a ten-year low. As the Korean central bank maintained a peg to the dollar, they were forced to sell assets to defend the won. Yet, by the end of November, they had burned through practically all of their foreign exchange reserves. Korea asked the IMF for help by late November and as soon as in December the IMF signed a deal with the Korean government for an emergency rescue package of $21 billion dollars.

Another ratio, which is more often used, relates the amount of short-term debt to the amount of international reserves:

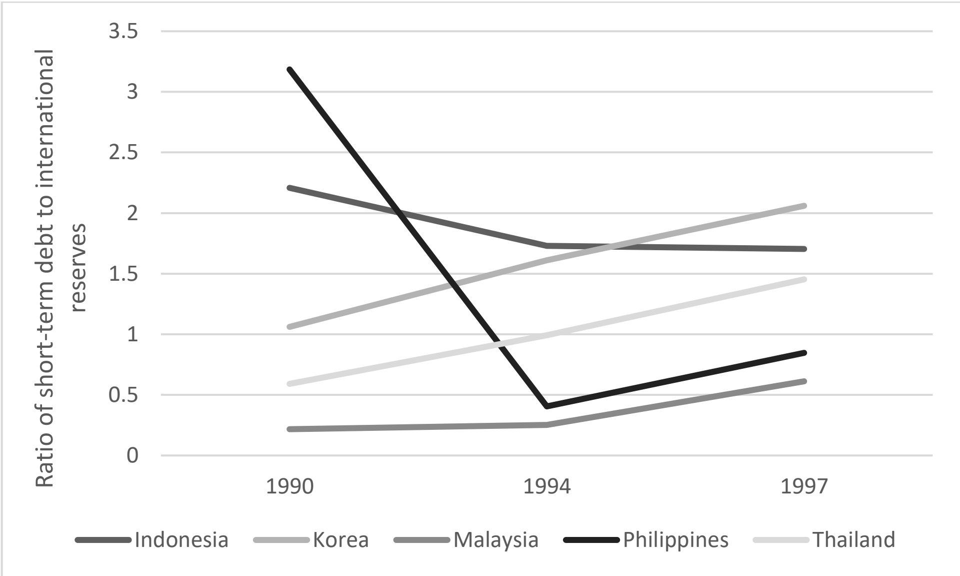


Figure 91: The ratio of short-term debt to international reserves of the affected Asian countries in the run-up to the 1997-98 Asian financial crisis.

In this comparison, we can appreciate the fact that the countries that had the highest ratios of short-term debt to international reserves, were the countries that were most affected by the 1997-98 crisis. Malaysia and the Philippines were economically less affected than Thailand, Indonesia and Korea. We could conclude that the degree of maturity mismatches was greater in these three economies.

At this point, we can briefly sum up our main conclusions with regard to this historical episode: (1) currency mismatching alone is not sufficient to cause a liquidity crisis and explain the 1997-98 Asian financial crisis, (2) there existed yield curve arbitrage in the run-up to the crisis, beginning in 1992, which narrowed the yield curve spread until it inverted in

1996, (3) the anecdotal evidence clearly points at (Korean) banks and nonbanks to borrow short in dollars and invest on the long term, (4) the currency devaluation was a consequence, not a cause of the crisis, (5) there were clear government incentive

17.4: The Great Recession of 2008

The Great Recession of 2008 was the most severe economic downturn since the Great Depression of the 1930s: the total cost of the 2008 crisis is estimated to range from 40 to 90 percent of one year's output (or from $6 trillion to $14 trillion or $50,000 to $120,000 per U.S. household), which is based on "lost output" as opportunity cost. As Atkinson et al. (2013) summarize the 2008 crisis:

> "The Second Great Contraction, the worst economic downturn since the 1930s, was unusual because it stemmed from an easing of credit standards and an abundance of financing that had fueled the prior expansion. This fuel also helped create imbalances - an overextension of mortgage - financing and capital market financial intermediation. A housing collapse and credit shocks, culminating in a financial crisis, hit the economy as these financial practices generated new losses. Home construction plunged, the stock market crashed, commodity prices tumbled, job losses mounted, credit standards tightened, and short-term funding markets seized up." (p. 2).

There are multiple theories that attempt to explain the financial crisis of 2008: among the most prominent, complete and coherent ones are, however, the 'Austrian' business cycle theory[201], extremely well-documented by Huerta de Soto (2006). In Huerta de Soto's (2011) words:

> "The expansionary cycle that has now come to a close was set in motion when the US economy emerged from its last recession in 2001 and the Federal Reserve embarked, again, on a major artificial expansion of credit and investment, an expansion that was not backed by a parallel increase in voluntary household saving. (…) The media of exchange originating from this severe fiduciary inflation had been placed on the market by the banking system as newly created loans granted at extremely low (and even negative in real terms) interest rates. **This fueled a speculative bubble in the**

[201] It is well worth noting that many 'Austrian' economists and practitioners were among the few *prepared* for a crisis. As Taleb (2012) argues, talk is cheap, actions speak louder.

shape of substantial rise in the prices of capital goods, real estate assets, and the securities that represent them and are exchanged on the stock market (…)" (p. 34) [emphasis mine]

As we have seen earlier, our equity *q shows* that that financial asset prices (the numerator in equity *q*) move ahead of the prices of the underlying productive assets (the denominator in equity *q*). As Smithers & Wright (2002) pointed out, as financial asset prices move ahead of the underlying productive asset prices, capital investment is induced and saving is actually discouraged.

Huerta de Soto's view on a too low rate of interest was confirmed by Fed official Fisher, who argued that as a result of looking at the wrong data monetary policy led to excessive "speculation" in the housing market (Atkinson, Luttrel, & Rosenblum, 2013). Many of the other "factors" that caused the 2008 crisis are mere symptoms of Huerta de Soto's interest rate theory of the business cycle: for instance, "bad loans made by banks", "failing rating agencies", "government incentives that encouraged banks to be reckless in their lending" and many other factors can be reduced to one common denominator: the 'Austrian' business cycle theory which shows the inevitable results of a decrease in interest rates *not* caused by an increase in savings.

Nevertheless, Huerta de Soto's (2011) focus is on the *creation* of credit, which, as we have expressed earlier, deemphasizes the role of financial intermediaries as intermediaries (Jordan, 2017) and deemphasizes the role of the composition of credit (and its *duration*) to the quantity of credit. Maturity mismatching is, in Huerta de Soto's views (2006), a symptom rather than a cause of the business cycle. The contribution we make, in this article, can therefore be seen either as a supplement to Huerta de Soto's (2011) explanation of the 2008 crisis, or as a slightly altered version of Huerta de Soto's (2006) otherwise excellent analysis of the phenomenon of the business cycle.

One of the problems in highlighting the role of maturity mismatching in the 2008 crisis, is a lack of data.

Since there exists no (publicly available) balance sheet data on commercial bank asset and liability maturities, it is difficult to estimate to what degree asset-liability maturity mismatching increased or decreased over the course of the business cycle that preceded the Great Recession of 2008. However, since rather recently, U.S. publicly listed banks are obliged to report *part* of their assets' maturities and the maturities of their financial obligations. Therefore, in an attempt to gather the data necessary to illustrate our theory, we gathered the 10-K Forms from the SEC of the five largest U.S. banks (measured by assets) and began collecting the maturities of loans and wholesale credit (which are categorized as maturity shorter than twelve months, between one year and five years and longer than five years), securities (divided by maturities' shorter than twelve months, between one year and five years, between five and ten years, and longer than ten years). The same was done for the banks' so-called "cash obligations." The banks included in the sample are J.P. Morgan Chase, Wells Fargo, Bank of America, Citi, and U.S. Bancorp for the years 2004 to 2012, which represent over 50% of (U.S.) commercial bank assets.

We divided both assets and liabilities in two groups: short-term (maturity of less than one year) and long-term (maturity greater than one year). In some cases, banks would not include their demand deposits as part of their "cash obligations," in which cases we supplemented the liability maturities by including the total amount of demand deposits. We then calculated the long-term asset/short-term asset and long-term liability/short-term liability ratios. Then we calculated, what we call, the *liquidity gap* by:

$$Liquidity\ gap = \frac{Long\ term\ assets/Short\ term\ assets}{Long\ term\ liabilities/Short\ term\ liabilities}$$

If the ratio long-term assets to short-term assets is greater than 1, it means that there are relatively more long-term assets relative short-term assets. Conversely, if the ratio is less than 1, it means that short-term assets have a greater weight than long-term bank assets. Equally, if the ratio long-term liabilities to short term liabilities is greater than 1, it means a bank (or the banking system) has more long-term liabilities than short-term liabilities. Conversely, if the ratio is less than 1, it means that short-term liabilities exceed long-term liabilities.

In general, the tendency is to have longer-term assets and shorter-term liabilities. In 2016, for example, the asset maturity ratio was *1.28*, whereas the liability maturity ratio was *0.37*. This proves that banks, as is generally known, finance themselves at a (very) short-term, as a majority of its liabilities consists of demand deposits, yet invest for the long-term.

We will chart the evolution of the *liquidity gap* in the U.S. (five major banks) below:

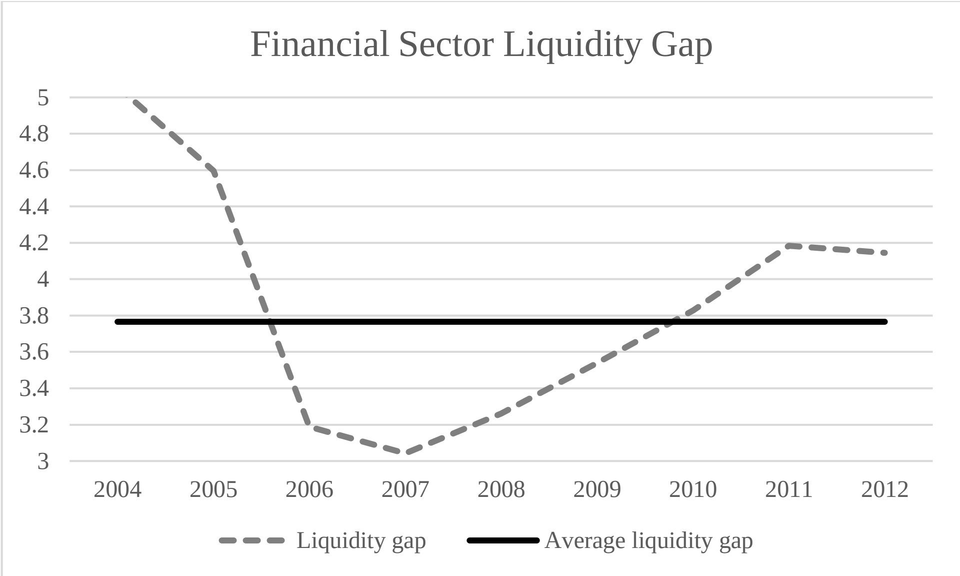


Figure 92: The liquidity gap *of the five major U.S. banks from 2004 to 2016. The (arithmetic) mean* liquidity gap *equals* 3.77. *As can be observed, banks engaged in extreme maturity mismatching in 2004 and 2005, but began reducing the liquidity gap in 2006 and 2007 (right before the 2008 financial crisis.*

As we can appreciate in the above graph, banks markedly increased the degree of maturity mismatches on their respective balance sheets. The *liquidity gap* reached a cyclical high in 2004 as banks were arbitraging the yield curve by increasing the degree of maturity mismatching. What we observe here, is essentially how financial intermediaries (*financial* entrepreneurs) transform a society's given set of time preferences into a misaligned financial maturity structure (Bagus & Howden, 2014). Whereas consumers express a desire to merely forego consumption in the short run, intermediaries transform this expressed preference into a demand for investment in the long run as they arbitrage the yield curve by "borrowing short and lending long." Lachmann's (1956) capital structure is thus mismatched or misaligned: consumers preferences do not match with the maturity structure of financial intermediaries.

Moreover, (nonfinancial) corporate businesses also increased *their* maturity mismatches. The typical theory of the business cycle, as presented by Huerta de Soto (2006), is very much focused on distortions induced inside or by the banking system. In that sense, a broader theory of maturity mismatching as cause of the business cycle would show that the very same intertemporal distortions can *be induced even in absence of banks[202]*. In this specific case, in the run-up to the 2008 crisis, we can observe that nonfinancial firms also increased their *liquidity gap* (which equals the ratio of long-term versus short-term assets divided by the ratio of long-term liabilities versus short-term liabilities) and thus the degree of maturity mismatches:

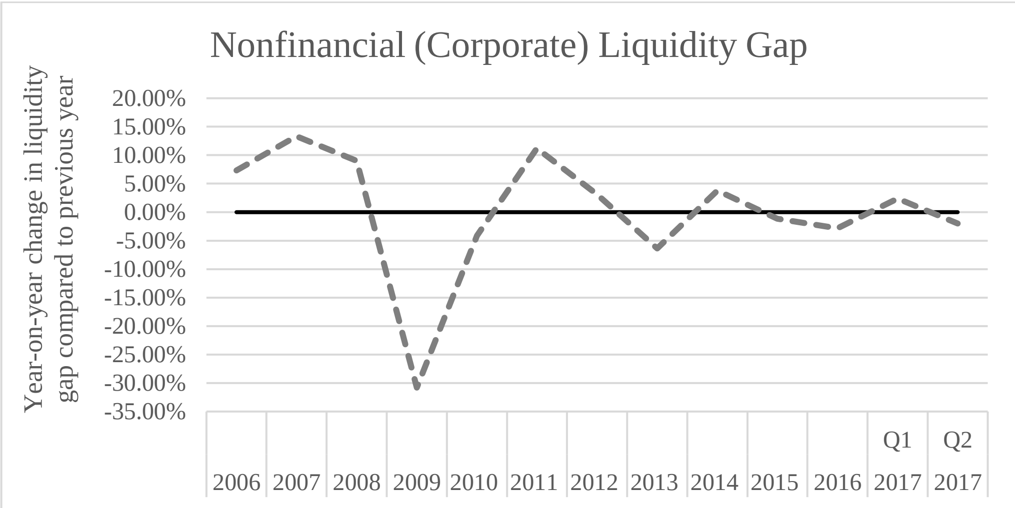


Figure 93: Nonfinancial (corporate) businesses also increased their degree of maturity mismatching. Data is from the Federal Reserve, Release Z.1, specifically: liquid assets (broad measure), total short-term liabilities, liquid assets as a percentage of short-term liabilities, short-term debt as a percentage of total debt).

Hence, not only did commercial banks began lending at longer maturities, nonfinancial corporate businesses also financed long-term assets with increasingly greater amounts of short-term debt. How this maturity mismatching affect the yield curve spread?

[202] As a quick thought experiment, it is interesting to see whether a theory of the business cycle is still valid after we replace all commercial banks with mere money market mutual funds (MMMFs). Recessions would necessarily disappear in Huerta de Soto's (2006) theory, but in a theory of maturity mismatching business cycles could continue to exist.

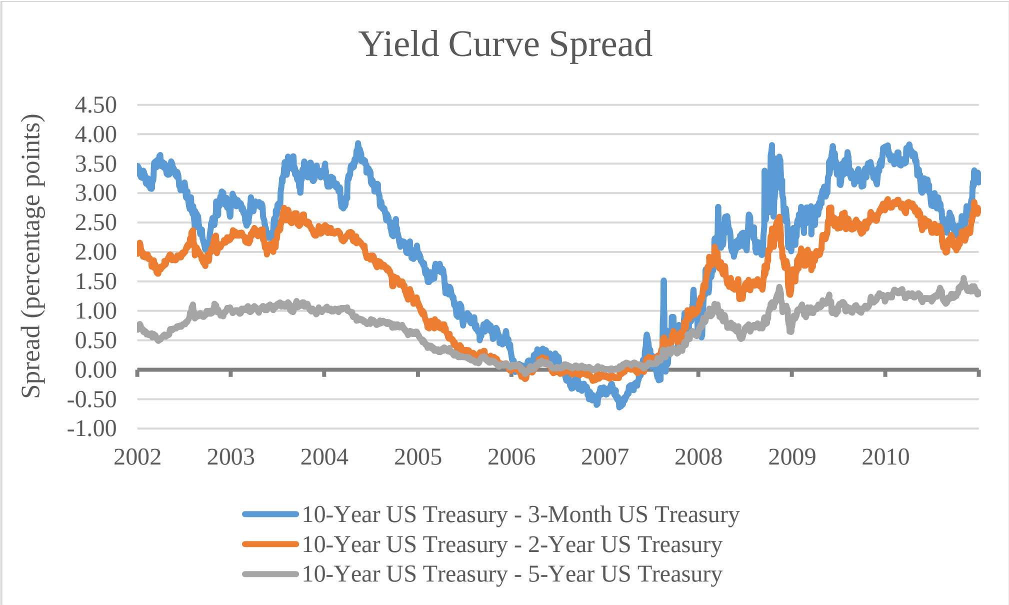


Figure 94: Different yield curve spreads from 2002 to 2011. Data from St Louis Fed.

In the above chart, we can appreciate the fact that from 2003/2004 (depending on the specific spread) onwards the spread began contracting, which is a sign that financial intermediaries are engaging in maturity mismatching, as they arbitrage the yield curve by borrowing short and lending long. Generally, as the 'habitat theory of the term structure' would predict, the first spreads to narrow are the "closest" spreads, in this case for example the 10-year / 5-year spread (which reaches a peak of approximately *1.10* in 2003 and from 2003 onwards contracts/narrows).

The same tendency can be observed in the net interest margin of banks (that is, the spread banks earn between their assets and liabilities):

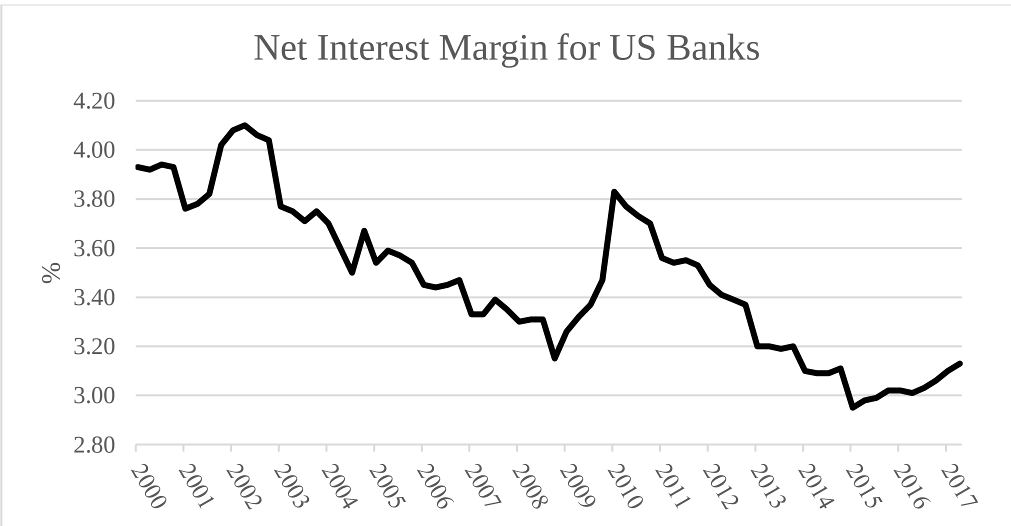


Figure 95: Ever since 2002/2003, when the yield curve spread began narrowing, banks' net interest margin began falling as well. Data from St Louis Fed.

Moreover, the (average) maturity of commercial and industrial loans also went up markedly since 2002 (at which average maturity topped at 400 days). In 2005, 2006 and 2007 average loan maturity went up to between 500 and 600 days (reaching even 700 days in 2007):

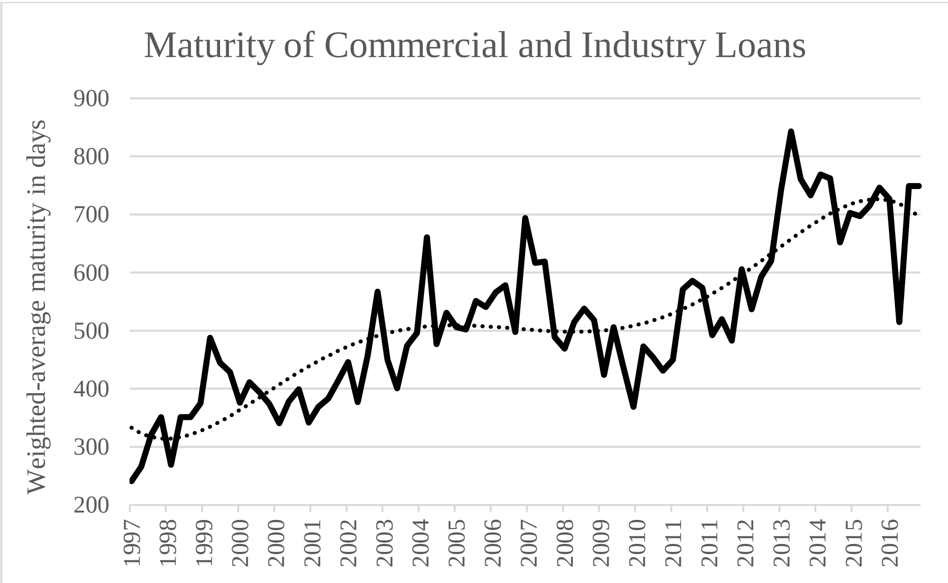


Figure 96: The (average) maturity of commercial and industrial loans held by commercial banks. Data from St Louis Fed.

Therefore, we can conclude that the conventional 'Austrian' theory of the business cycle, as presented by one of its main proponents Huerta de Soto (2006), has another component that has received little attention thus far: the degree of maturity mismatching.

In an alternative view, it is the degree of maturity mismatching that produces an artificial boom during which maturities and durations are lengthened above and beyond the maturities at which consumers/households wish to consume or wish to have disposal over their savings. A recession is then an inevitable outcome of an illiquid capital structure, in the sense that the preferred maturities of consumers are misaligned with the maturities in the financial system and the maturities at which corporates finance themselves. The result is an artificial boom, also evidenced in our equity *q* measure, in which financial asset prices move ahead of the underlying productive asset prices which, moreover, induces capital investment (the

‘Austrian’ equivalent of “malinvestment”) at unsustainable durations and discourages household savings (Smithers & Wright, 2000).

The contribution of this brief reflection on the crisis of 2008 is thus that it shows how maturity mismatching has behaved over the course and in the run-up to the 2008 financial crisis, with data that was previously not available.

Section VI: Concluding Remarks

Chapter 22: Theoretical Implications

Our conclusions (and contributions to the field) related to our theoretical work are numerous:

1. Böhm-Bawerk (1888) criticized the theories of interest based on "productivity" in a convincing way: if these theories were to be correct, entrepreneurs would simple arbitrage prices until excess returns disappear. Therefore, the productivity theory of interest is unable to explain the interest phenomenon. In essence, the proponents of this "naïve" productivity theory of interest (e.g., Clark, 1899; Ricardo, 1817) do not explain why entrepreneurs do not eliminate profits, taking advantage of the returns that a capital good offers. Moreover, Clark (1899) was mistaken when he distinguished between individual capital goods and "capital" as a fund: both are essentially capital, the same theory of capital applies to them, yet only the "scale" differs.

   Some of the defenders of the productivity theory of interest, such as Frank Knight (1934) adopted a pessimistic view on the future of capitalism. They thought that, due to diminishing marginal returns on capital, profits would eventually disappear completely. A zero rate of return would then, one way or another, lead to an existential crisis of capitalism. On top of everything, many modern economists, such as Solow (1953), base their theory of diminishing returns on capital not on the notion of entrepreneurs successfully arbitraging away profits and thereby eliminating profit opportunities, but on the notion that every additional physical unit of capital (that is, another machine) has a diminishing *physical* return. There exists a wide divide between the economists that attempt to study economics and economists that confuse studying physical production with studying economics.

2. Böhm-Bawerk, despite his impeccable critiques, based his own theory of capital on the classical triad of production (e.g., Smith, 1776): labor (wages), land (rent) and capital (interest). According to this material theory of capital, capital consists of "produced means of production." Capital, therefore, is a historical combination of land and labor. Moreover, capital is heterogenous, since *no factor of production is the*

physical equivalent of another. This notion of capital as something physical, was later adopted by Cobb and Douglas (1928) in their article about the production function: the two economists attempted to estimate the empirical tendency between material inputs (specifically labor measured in weeks worked and capital measured in physical terms expressed in dollars according to a capital good index) and some physical output (produced products). As such, Cobb and Douglas estimated the coefficients that labor and physical capital contributed to physical output. Unfortunately, the sad legacy of the classical economists still haunts modern-day applied economics. The same production function, or any of its intellectual spin-offs, continues to be one of the favorite tools among modern economists to estimate economic future, to project public finances and many other types of macroeconomic predictions, following a method that did not change much since the original Cobb-Douglas paper (1928), popularized by the (neoclassical) participants of the third round of capital controversies: the Cambridge Controversies (e.g., Solow, 1953, Solow, 1963; Swan, 1953).

3. The Knight-Kaldor-Hayek controversy, as all historical debates on capital, never reached its intellectual pinnacle. Hayek's critiques of Knight were clear: according to Hayek, they did not take into account the element of time. This omission brings the Knightian capital scheme down to its knees. Essentially Knight, as his intellectual predecessor J.B. Clark, argues that the theory of time preferences and (average) period of production could not be correct, because capital is "perpetual," in the sense that ever since the moment of its creation (its "genesis") it automatically renews and sustains itself. As it takes into account depreciation in its profit calculus, capital is essentially an inexhaustible source of income. It no longer requires abstinence; abstinence was only necessary at the start, when the first human beings needed to abstain from consumption to "establish" the initial capital and its subsequent perpetual flow of income. Ever since, the theory of abstinence plays no part. Knight, as Hayek recognized, was wrong. The fact that one accounts for depreciation in economic calculation does not prevent him from consuming his entire capital. Every cash flow implies a *new* intertemporal decision and, therefore, an infinity of repeated acts of "abstinence." Far from being "perpetual" and "automatic," to maintain capital is a daily decision that requires deliberate and continuous abstinence. Incapable of recognizing the role of time, Knight never succeeded in defining a coherent theory of

interest: interest is simply the equivalent of profit. Nonetheless, Knight criticized Hayek and Kaldor for their *material* theories of capital: a production good is from an economic point of view nothing different from a laborer or a piece of land. To distinguish between labor (wages), land (rent) and capital (interest) is pointless. Hayek's error, according to Knight, consisted in his backward-looking approach to capital (which involves arbitrarily determining whether a given good is "produced" or "original"), whereas capital is only concerned with the future (capital, hence, is forward-looking). We conclude that the *subjective* theory of capital of Knight, in this sense, is superior to the *material* theory of capital of Hayek, supported with several examples regarding, for instance, the notion of human capital.

4. We conclude that Fisher (1930) made a defining and extremely valuable contribution to economics: contrary to the "naïve" theories of interest, based on the physical productivity inherent in capital (such as apple trees naturally yield apples), Fisher supplemented the theory of interest based on abstinence or, more precisely, subjective time preferences (e.g., Mises, 1949). Fisher (1930) agrees with the theory of time preferences, but only as an explanation of the supply side of the *intertemporal* market. Nevertheless, the theory does not satisfactorily explain the other side of the equation, necessary to explain the phenomenon of the *market* rate of interest, which is the demand side of the *intertemporal* market.

   Demand, as Fisher (1930) explains, depends on the available "investment opportunities." When entrepreneurs do a good job in arbitraging the price differences that exist in disequilibrium, there will be few investment opportunities and, therefore, reduced demand by the different entrepreneurs on the *intertemporal* market. Conversely, whenever there exist many maladjustments in the price structure, many profit opportunities exist (that is, high returns) and, therefore, a high demand by entrepreneurs on the *intertemporal* market. Both factors, subjective time preferences and investment opportunities, explain the interest phenomenon. In this way, Fisher (1930) contributes to the theories of interest based exclusively on subjective time preference (the "pure" time preference theory of interest).

5. The neoclassicals of the Cambridge Controversy contradict themselves in their capital theory in various ways: when capital is an input that is measured in terms of money,

and when you need a rate of interest to discount future income in terms of money to obtain an estimate in terms of money, and when the rate of interest – as is custom in neoclassical theory – simply equals the rate of profit, then you have a capital theory that depends on circular logic. Therefore, the neoclassical capital theory cannot be possibly correct. Essentially, the neoclassicals (e.g., Solow, 1963) have refuted themselves. Another valid critique on the neoclassical theory of capital by the economists of Cambridge, U.K., was the notion of "aggregating capital" to later use the result as a sum input in an "input-output" model. In other words, as we can appreciate in Menger (1888) and Mises (1949), capital is an entrepreneurial method of economic calculation that allows the entrepreneur to estimate future profits and losses. That is, capital is an *outcome* of the production process, given the fact that it consists in the present value of income streams that the many different production goods may be able to yield in the future. Capital represents, in a certain way, much better a future "output" than an input into the production process that yields a present "output."

6. The alleged problem of *reswitching* of capital is not a true problem. Economists of Cambridge, U.K. (e.g., Robinson, 1953; Sraffa, 1960; Pasinetti, 1966) thought that the phenomenon of *reswitching* of capital completely refuted Solow's, Swan's and Samuelson's theory of capital and, therefore, all neoclassicals pretensions to use the production function as a foundation of applied economics. The problem of *reswitching* consists in the fact that, according to the theory of Böhm-Bawerk and Mises and to the application of the Solow-Swan production function (1956), in which the "capital stock" is measured in terms of money that later depends on a rate of discount, that is, a rate of interest, some production techniques or net present values (NPVs) appear to be more profitable at a high rate of interest, less profitable at a medium rate of interest, but suddenly become more profitable again at low rates of interest. This paradox or deficiency refutes the idea that a production technique/NPV is favored at a high interest rate, while another production technique/NPV is favored at a low interest rate. The notion that the rate of interest determines the "roundaboutness" of a capital structure or the "period of production" is therefore, according to the Cambridge, U.K. economists, false. This study provides two solutions: (1) to use the term structure of interest rates (instead of a single rate of interest) in case the phenomenon appears and (2) to apply the methodology of "multiple interest rates" advocated by Osborne (2014), which consists of using both

the orthodox and unorthodox rates in calculating the NPV, since the equations used to discount cash flows are polynomials. We have presented various counterexamples in which the *reswitching* phenomenon completely disappears. Both approaches have a lot in common, since the second is a way to “weigh” the discount rate with respect to when cash flows occur, which even leads us to a better approximation of *duration* (Macaulay, 1938). The phenomenon of *reswitching* was first noticed by Fisher (1930), yet he never followed up on his discovery (more than anything, it was a simple curiosity of seemingly little practical importance), but became extremely important in the Cambridge Controversies.

7. We conclude, after reviewing the often-overlooked debate on shiftability of the beginning of the 20th century, that there is a large difference between “systemic” and “individual” liquidity. What applies to one bank (the ability to “shift” illiquid assets to stronger banks in times of trouble) does not apply to the system as a whole. According to the defenders of the shiftability theory, commercial banks can invest in illiquid assets (for instance, long-term mortgages, medium-term consumer credit, long-term corporate debt), since the development of secondary markets and more sophisticated capital markets meant that a bank could always “become liquid” by selling its “illiquid assets” to other, stronger banks. Nonetheless, this point of view assumes that the “illiquidity” events occur more or less randomly distributed over time and banks, but illiquidity in a majority of cases do not affect banks in an isolated way, but all at once for the entire banking system in periods in which a majority of banks are illiquid. This occurred in the Great Depression of the 1930s. Unfortunately, the defenders of shiftability gradually came out on top despite the Great Depression and partly due to the start of the Second World War, which favored investment in long-term public debt through the Federal Reserve to finance the war effort.

8. The Hayekian triangle is based on an erroneous concept of capital: it is not approximity to consumption that matters, but *optionality to consume* (liquidity). Duration applies to any income stream, not just to income streams close to consumption. Moreover, even at stages near final consumption, some investments require a lot of capital and only produce cash inflows after several years. We then backed our critique of the Hayekian triangle with historical data on the sensitivity of

the maturities (average age) of different types of assets to changes in interest rates.

9. We have derived and coined the Fisher's "pendulum of returns" theory. Fisher (1930), possibly as a result of his non-academic work, recognized an important dynamic between demanders of present goods and suppliers of present goods on intertemporal markets, which explains why the long-term interest rate (contrary to the short) behaves stable over long stretches of time. The more savings available to invest, the more resources entrepreneurs have to arbitrage prices, earn profits and arbitrage away price differentials. All other things equal, if investment become less attractive, the demand for present goods goes down, which lowers the market rate of interest and discourages suppliers of savings, in the margin, to continue saving at the same pace or save more (empirical studies that generally show an inverse relationship between interest rates and consumption backs this feedback mechanism). Now, when savers start to substitute, in the margin, savings for consumption, they reduce the supply of present goods in the *intertemporal* market. This, in turn, leads to an increase in the rate of interest and to a decline in the resources available for entrepreneurs to arbitrage away profits, which therefore results in greater maladjustments in the price system due to the absence of arbitrage and, as a consequence, higher profits. This provokes, again, a return of entrepreneurs and higher *demand* in the intertemporal market. This dynamic process in the intertemporal market can be visualized as a pendulum, which oscillates from side to side but, due to the presence of natural feedback mechanisms, always gravitates to some steady center. This theory is applied to other realms: principally, to the recent controversy on passive and active investment strategies. When the amount of passive investment (and passive investors) passes some optimal point, price arbitrage will become less efficient and greater profit opportunities will arise. This will, in turn, lure active investors back into the market (and favor active investment strategies). This until returns are largely arbitraged away and more difficult to find, after which a countertendency will happen in which passive strategies will be favored over active investment strategies. Hence, we applied our formulation of Fisher's theory of pendulum returns to the passive-active investment controversy.

10. The theory of the entrepreneur is well-developed (Mises, 1949; Kirzner, 1970; Huerta de Soto, 2010): the entrepreneur is an arbitrageur of profit opportunities that are

implicit in the present and future price structure. As such, entrepreneurs have a coordinative role until, in equilibrium, no price differentials remain and no profits are left. In his very essence, the role of the entrepreneur is “resource-less.” Resource providers are capitalists, or savers, that provide the resources of the entrepreneur. Our conclusion is that in this theory one important layer has been overlooked: the *financial* or capitalist-entrepreneur. These intermediaries are in charge with selecting entrepreneurs and allocating capital. Financial entrepreneurs determine which of its nonfinancial peers receive resources and under what conditions.

11. We have extended the theory of *q* (e.g., Wright, 2004; Spitznagel, 2012) into the context of banking (principally to the realm of foreign exchange rates), real estate and nonfinancial monetary assets. The theory of *q* predicts future returns across asset classes and, when it indicates large *overvaluation* (fourth quartile), shows that expected future returns are greatly diminished or even negative.

12. Capital is *financial net worth* in line with contributions by Menger (1888) and Mises (1949). The conclusion of this work is that this is the only coherent definition from a subjectivist point of view. In this theory of capital, the entrepreneur, both as a profit-arbitraging and capital-valuing entrepreneur, becomes the key feature and driver of economic activity, instead of the expert economist that arbitrarily classifies production goods according to some physical or objective characteristics from the convenience of his ivory tower. The theory of capital is thus based on (sometimes asymmetric) economic calculation by entrepreneurs. In essence, it is a forward-looking rather than a backward-looking theory of capital.

13. The theory of capital is “scalable”: our theory of capital does not revolve around classifying goods: it only establishes the different scales of capital and the entrepreneurial arbitrage between each scale. It argues that capital markets (including banks) are intimately connected to the capital structure, since the decisions of savers (capitalists) will determine the maturity or duration of the investment of the not-consumed resources in a capitalist society. As such, the capital structure can be more or less liquid, given the time preferences of savers and the investment opportunities of entrepreneurs at each maturity/duration. The “scale” of capital refers to the fact that the same concept can be applied at many levels: the level of the individual asset (for

example, a building), at the firm level (a combination of assets), or at the stock market level (a combination of firms). Between every level, every scale, there exists entrepreneurial arbitrage. This is how we, theoretically, bridge and connect the theory of the *q* ratio by Tobin and Brainard (1976) to our capital theory. Tobin and Brainard (1976) established a theory of arbitrage between financial claims (debt and equity) and the underlying assets by which they are backed. The prices of the individual productive assets represent (together)) the "replacement value," whereas the prices on publicly-listed stock markets represent the "market value." The ratio between the two equals *q*. This theory is a crude and narrow formulation of the broader theory of capital (and the principle of *q*) that we proposed in this work. In effect, our definition of capital, based on the entrepreneurial function, economic calculation, and subjectivism, naturally flows to the theory of *q*.

We conclude that a specific good, such as a building, is "capital" equivalent to the net present value of the future flows or contribution to the flows it is able to generate. A combination of goods, such as a firm, is "capital" equivalent to the net present value of the future income streams it is able to generate. Both are "capital" but at different "scales." The *q* ratio is a result of such "scale" differences. Nonetheless, there exists a tendency for the price of the sum to equal its parts: in the opposite case, arbitrage opportunities would exist for Kirznerian entrepreneurs. Therefore, the principle of q is characterized by a regression to the mean: it can never deviate permanently from its "equilibrium." The theory of capital and the theory of *q* are unity.

14. We show that financial entrepreneurs are in charge with assuring that the time preferences of capitalists align with the temporal structure of capital. In equilibrium, the maturities of capitalists align with the maturities of financial entrepreneurs, which in turn align with the *duration* of the productive structure. Income streams of businesses align with income streams of financial intermediaries which align with the (expected) income streams of consumers. This intertemporal dimension in capital theory is of utmost importance.

15. The 'Austrian' theory of the business cycle, so well-espoused, defended and developed by Huerta de Soto (2006), is refined in the following ways: by incorporating an explanation of the cyclical nature of the *q* ratio, by including

maturity mismatches in the capital structure, by emphasizing the relationship between maturity mismatching and yield curve arbitraging, and a focus on the financial Macaulay *duration* instead of the Hayekian theory based on nearness to final consumption (or to sectors near final consumption). As such, a reformulation of the 'Austrian' business cycle theory is proposed. To sum up, the economic cycle is characterized by the following stages: (a) capitalists opt for liquid, short-term savings, (b) banks arbitrage the yield curve, (c) long-term interest rates begin to decline and the yield curve spread narrows, (d) financial asset prices and *q* ratios rise, (e) with lower long-term rates, businesses expand the *duration* of their investment projects, (f) long-term (fixed) investment rises, profits peak and prices of "productive assets" (such as commodities) begin to rise; (g) capitalists begin to liquidate their holdings and begin to consume savings (the savings rate tends to go down), which leads to the start of a tendency toward higher prices or higher short-term rates or both; (h) businesses begin to default because WACC is rising (higher prices or higher rates); (i) the yield curve spread goes negative and the yield curve inverts; (j) financial asset prices collapse and *q* ratios drop (possibly below their "equilibrium" levels due to forced liquidations; (k) massive liquidation and liquidity crisis: rise in defaults, weak lenders and borrowers go bankrupt, unemployment rises and (the value of) capital is reduced violently and collapses; (l) recovery: time preferences align again with the temporal financial structure and the *duration* of the productive structure; (m) once the recovery is completed, any attempt by financial intermediaries to arbitrage the yield curve again leads to a new cycle.

16. Applying our reformulated business cycle theory sheds a new light on prior recessions. Specifically, we take a look at the Great Depression, the S&L crisis of the 1980s, the Asian crisis of the late 1990s, and the Great Recession of 2008. With the data we gather, we show that maturity mismatching and yield curve arbitrage were important phenomena in these instances.

Chapter 23: Practical Implications

The implications of our work for practitioners can be summarized as follows:

1. By using our proposed measures of *q*, practitioners can take into the growing probability of mean reversals (toward mean *q*) by either protecting their downside[203] or increasing their exposure to possible upside. As we have shown in our studies, high levels of *q* are related to lower expected returns and higher probabilities of a drawdown. Likewise, low levels of *q* offer opportunities, since lower levels of *q* are related to higher expected returns. Moreover, the odds of a (large) drawdown increases considerably, in addition to the median period before a drawdown occurs, when we include the yield curve spread (we divided the yield curve spread in a high/low group divided by the median spread). The theoretical explanation for doing so is that maturity mismatching is one of the key drivers of recurrent economic crises. The yield curve is one of the more robust predictors of volatility and economic crises.

2. In general, when *q* ratios are high, future expected returns are low. Hence, practitioners are advised to take into account this measure of over- and undervaluation when, for instance, they invest on the housing market. It would, without a doubt, be profitable to reduce exposure to the market when *q* is high, and increase exposure to the market when *q* is low. By “optimizing” for *q*, higher returns can be earned.

3. With regard to economic forecasting, it is clear that there are deep fundamental problems with the neoclassical production function and all its derivations. As a result, governments, institutions such as the IMF, BIS and World Bank, (central) banks, should proceed with caution and should avoid using exact GDP growth estimates Moreover, central banks should abandon the use of “output gap” (potential GDP) measures, which is largely based on the flawed production function and is an input to monetary policy decisions. The “output gap” can be positive, which indicates that the economy is “overheating” (and the Fed should contract monetary policy), while when

[203] Our reader should be reminded that even with low *q* (and a high yield curve spread) true “black swans” are possible. *Some* downside protection seems therefore reasonable at all times.

the "output gap" is negative the Fed should keep policy loose[204]. However, there exists a (theoretically refuted) tendency for the production function toward diminishing returns, as well as a built-in tendency for the production function to overestimate the contribution of labor and underestimate future "output growth."

4. Governments should stop pursuing wealth equality policies that destroy wealth, since they are based on false premises: the neo-Ricardian theory of capital misunderstands the nature of capital and therefore leads to flawed conclusions about wealth distribution. There is no justification for the tendency of r to exceed g indefinitely.

5. Gold investors should consider opting for investing in gold mining stocks whenever the gold q, if calculated in a reliable and consistent fashion, indicates relative overvaluation.

6. Governments should analyze the ways in which policy can – as an unintended consequence – remove one of the feedback mechanisms in the various applications of Fisher's pendulum of returns. Investors and savers should be wary for "overoptimization" and "overexposure" to passive investment strategies, taking into account our application of our theory of Fisher's pendulum of returns.

7. Central banks should carefully determine and monitor their own net present values (NPVs) and the quality of their assets to prevent future bursts of inflation or foreign exchange rate depreciation.

8. Practitioners that use DCF-models for capital budgeting decisions, should be aware of the effect of the term structure and the polynomial nature of NPV calculations (and the possible occurrence of *reswitching* in practice). We, therefore, insist on using the term structure of interest rates to discount future cash flows rather than using one single rate, as well as possibly applying Osborne's (2014) methodology to include the product of both orthodox and unorthodox rates when discounting cash flows.

[204] It is currently estimated that the (estimated) "output gap" will remain negative until 2019/2020.

Chapter 24: Future Research

The principle of *q* can be applied to a myriad of different asset classes and circumstances; the concept serves as a principle; its application is up to its user. Future research of our principle *q* involving other assets, as well as other markets (for instance, the Chinese real estate market), would add tremendously to our understanding of *q*. As such, another quite obvious application of our *q* principle, is deriving and estimating an equity *q* for European, Chinese, Russian and/or Japanese stock markets (in short, non-U.S. markets) and do empirical testing on whether equity *q* is predictive of future returns on their corresponding stock markets.

As we have stated, money and capital are intricately related: nothing becomes so clear when it comes to the use of collateral on capital markets, reuse of collateral, collateralization, hypothecation and rehypothecation. The widely used (reverse) repurchase agreements (or repo's) are, practically, equal to collateralized loans. These subjects, so important in modern times[205], have nevertheless received precious little attention in economics. It is necessary to see how the theory of capital links collateral to capital markets and what the broader implications are. Perhaps the most interesting avenue for future research is, therefore, collateralization in the context of modern financial markets. There appear many gaps in economic literature on these complex subjects, which opens a myriad of academic opportunities. Especially the use and reuse of collateral in the context of banking and the business cycle deserves close attention. In this sense, collateralization (and securitization) might play an important role in the *composition* of capital. These instruments are relatively new (at least in the manner how they are used by modern banking institutions) and must be analyzed from a broader, economic point of view.

One of the more obvious applications of our capital theory and q, is to non-Latin American countries and central banks. A case study for the ECB since inception would be interesting, although such an application requires a more elaborate model and balance sheet (*net worth*) analysis. Nevertheless, the theory of negative central bank net worth and asset quality of central banks' as key determinants of both domestic inflation and exchange rates is a relatively new field with little research effort thus far. It will also become much more

[205] Recently, the term "collateral runs" became more popular. Collateral runs are the 21st century equivalent of 20st century bank panics and require our utmost attention.

important in the near future, as modern central banks in developed countries now pay interest on reserves. As a consequence, it is likely that at some point modern central banks will incur operational losses. A more developed and integrated theory of how central bank profitability (and the quality of a central bank's underlying assets) is therefore absolutely necessary, considering how and to what extent it affects the *composition* of the capital structure. Moreover, the fiscal consequences of such policy are important to study. There are some obvious links with interest rates (attractiveness of currencies) and domestic inflation that could be, possibly, explained by one common denominator: a banking system's underlying assets. There is much to be done with regard to the 'backing theory' and our banking *q*.

To close our round of obvious avenues for future research, it would also be interesting to gather data and apply our reformulated theory of the business cycle to other historical episodes, such as the recessions in the 1970s, the *Dotcom* bubble of 2000, and other episodes of recession. Given our theoretical foundation, we expect either central banks, commercial banks or shadow banks to engage in maturity mismatching and lowering the long-term rate of interest compared to the short-term rate of interest.

Another interesting yet less obvious avenue for future research are maturity mismatches between risks and rewards. Taleb (2012), for instance, argues that bankers generally received annual bonuses, whereas statistical blow-ups would happen every few decades. Hence, the *duration of risk* could be ten years, whereas the *maturity to reward* is a single calendar year. As a result, individuals push as much risk into the future as possible to optimize for the short-term reward. *If*, however, the maturity of risks and rewards could be aligned, such incentives disappear. Contract theory could provide important insights in this regard, as bonus packages can be contingent upon certain longer-term parameters to reestablish maturities between risk and reward.

Moreover, there is precious little economic (and 'Austrian') literature on the role of risk, especially in relation to the concepts we discussed in our work, "roundaboutness" and capital-intensiveness. Moreover, in the context of the ('Austrian') business cycle theory there is no research, so far as I know, on the convexity of interest rates and their role over the business cycle.

Last but not least, with regard to modern-day applications of the neoclassical production function, is remans important to examine how much the practical applications vary to identify how the defective neoclassical theory of capital affects their outcomes. For instance, some forecasters use the production function on an industry level before aggregating them, opposed to other applications of the production function directly on the aggregate economy. It is a mystery how present-day uses of the production function in practice differ from the theoretical textbook examples. In this sense, an extremely interesting avenue for research is to investigate the tendency, identified by this present work, of the production function to overestimate the contribution of “labor” to growth. How large is the bias and how large will it become in the future? In addition, the modern dynamic stochastic general equilibrium (DSGE) models use, in principal, the same methodology: a simple production function is often used to “represent” the productive structure. The consequences of these models in the context of capital theory remain unclear and require further research.

Bibliography

Ahmad, J. (1984). The Vicious Circle of Inflation and Depreciation. In J. Ahmad, *Floating Exchange Rates and World Inflation* (pp. 143-162). Palgrave Macmillan.

Aliber, R., & Zoega, G. (2011). *Preludes to the Icelandic Financial Crisis.* New York: Palgrave Macmillan.

Archer, D., & Moser-Boehm, P. (2013). Central bank finances. *BIS Papers No 71*.

Artis, M., & Marcellino, M. (2001). Fiscal forecasting: The track record of the IMF, OECD and EC. *Econometrics Journal*, 20-36.

Atkinson, T., Luttrel, D., & Rosenblum, H. (2013). How Bad Was It? The Costs and Consequences of the 2007-09 Financial Crisis. *Federal Reserve Bank of Dallas Staff Papers*(20), 1-22.

Bach, L., Calvet, L., & Sodini, P. (2017). Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy. *CEPR Discussion Paper No. 11734*.

Bagus, P. (2016). A Comparison of Direct Investment of Savings and Cash Building of Savings: A Response to Alexandru Patruti. *The Quarterly Journal of Austrian Economics, 19*(4), 359-375.

Bagus, P., & Howden, D. (2014). The Term Structure of Savings, the Yield Curve, and Maturity Mismatching. *Quarterly Journal of Austrian Economics, 13*(3), 64-85.

Bahegot, W. (1873). *Lombard Street.*

Balassa, B. (1989). The Effects of Interest Rastes on Savings in Developing Countries. *PPR Working Paper Series*, 1-16.

Barbon, N. (1690). *A Discourse of Trade.*

Barker, W., McNelis, P., & Nickelsburg, J. (2013). Currency Exchange Rates. In C. I. Books, *Economics for Investment Decision Makers: Micro, Macro, and International Economics* (pp. 465-526).

Barsky, R., & Summers, L. (1988). Gibson's Paradox and the Gold Standard. *Journal of Political Economy, 96*(3), 528-550.

Baum, C. F., & Thies, C. F. (1989, Oct). The Term Structure of Interest Rates and the Demand for Money during the Great Depression. *Southern Economic Journal, 56*(2), 490-498.

Berle, A., & Pederson, V. (1934). *Liquid Claims and National Wealth.* New York: The Macmillan Company.

Bernanke, B. (1990). On the Predictive Power of Interest Rates and Interest Rate Spreads. *NBER Working Paper Series*.

Bezemer, D. (2010). Understanding financial crisis through accounting models. *Accounting, Organizations and Society*.

Birner, J. (2002). *The Cambridge Controversies in Capital Theory: A study in the logic of theory development.* Routledge.

Black, F. (1990). Mean Reversion and Consumption Smoothing. *Review of Financial Studies, 3*(1), 107-114.

Blaug, M. (1997). *Economic Theory in Retrospect.* Cambridge: Cambridge University Press.

Bliss, C. (1975). *Capital Theory and the Distribution of Income.* North-Holland Publishing Company.

Bliss, C., Cohen, A., & Harcourt, G. (2005). *Capital Theory.* Edward Elgar Publishing.

Böhm-Bawerk, E. (1895). The Positive Theory of Capital and Its Critics I. *Quarterly Journal of Economics, 9*(2), 113-131.

Böhm-Bawerk, E. v. (1881). Whether Legal Rights and Relationships are Economic Goods. In *Shorter Classics of Eugen v. Böhm-Bawerk.* Libertarian Press.

Böhm-Bawerk, E. v. (1888). *The Positive Theory of Capital.* London: Macmillan & Co.

Braun, E. (2013). Carl Menger's Contribution to Capital Theory. *TUC Working Papers in Economics*(11).

Brigham, E., & Ehrhardt, M. (2008). *Financial Management: Theory & Practice.* Thomson South-Western.

Brown, H. (1913). The Marginal Productivity vs. The Impatience Theory of Interest. *Quarterly Journal of Economics*(27), 630-650.

Brunnermeier, M., Gorton, G., & Krishnamurthy, A. (2012). Risk Topography. In D. Acemoglu, & M. Woodford, *NBER Macroeconomics Annual 2011* (Vol. 26, pp. 146-176). University of Chicago Press.

Buiter, W. (2008). Can Central Banks Go Broke? *CEPR Policy Insight No. 24,* 1-12.

Burmeister, E. (2002). A Retrospective View of Hicks' Capital and Time: A Neo-Austrian Theory. *Working Paper*.

Cachanosky, N., & Lewin, P. (2014). Roundaboutness Is Not a Mysterious Concept: A Financial Application to Capital Theory. *Review of Political Economy*.

Cachanosky, N., & Lewin, P. (2018). Hayek's Average Period of Production: One Step Forward and a Missed Opportunity. *Working Paper*, 1-20.

Campbell, J., & Shiller, R. (1987). Cointegration and Tests of Present Value Models. *Review of Economic Studies, 58*(3), 495-514.

Cantillon, R. (1755). *Essai sur la Nature du Commerce en Général.* London: Fletcher Gyles.

Caplan, B. (2005). The Myth of Time Preference. *Working Paper*.

Chamley, C. (1986). Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica*, 607-622.

Chang, R., & Velasco, A. (2001). The 1997-98 Liquidity Crisis: Asia versus Latin America. *Central Bank of Chile Working Paper*(120).

Clark, J. B. (1888, May). Capital and Its Earnings. *Publications of the American Economic Association, 3*(2), 9-69.

Clark, J. B. (1891). Distribution as Determined by a Law of Rent. *Quarterly Journal of Economics, 5*(3), 289-318.

Clark, J. B. (1895, April). The Origin of Interest. *Quarterly Journal of Economics, 9*(3), 257-278.

Clark, J. B. (1899). *The Distribution of Wealth.* New York: Macmillan.

Cobb, C. W., & Douglas, P. H. (1928). A Theory of Production. *The American Economic Review, 18*(1), 139-165.

Cohen, A. (2003). The Hayek/Knight Capital Controversy: The Irrelevance of Roundaboutness or Purging Processes in Time? *History of Political Economy, 35*(3), 469-491.

Cohen, A. (2006). The Kaldor/Knight controversy: Is capital a distinct and quantifiable factor of production? *European Journal of History of Economic Thought*, *13*(1), 141-161.

Cohen, A. (2011). Is equilibrium enough and was Stigler wrong? Value theory in the Böhm-Bawerk / Fisher controversies. *HOPE Center Working Paper, No. 2011-01*.

Cohen, A. J. (2007). The Mythology of Capital or of Static Equilibrium?: The Böhm-Bawerk/Clark Controversy.

Cohen, A., & Drost, H. (1996). Böhm-Bawerk's Letters to J.B. Clark: A pre-Cambridge controversy in the theory of capital. In P. Arestis, G. Palma, & M. Sawyer, *Capital Controversy, Post Keynesian Economics and the History of Economic Thought: Essays in Honour of Geoff Harcourt, Vol I* (pp. 75-87). Routledge.

Culbertson, J. (1957). The Term Structure of Interest Rates. *The Quarterly Journal of Economics*, *71*(4), 485-517.

Damodaran, A. (2008). What Is the Riskfree Rate? A Search for the Basic Building Block. *Working Paper*.

De Soto, H. (2000). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else.*

Desai, M. (1991). Kaldor Between Hayek and Keynes, or: Did Nicky Kill Capital Theory? In E. Nell, & W. Semmler, *Nicholas Kaldor and Mainstream Economics: Confrontation or Convergence?* Palgrave Macmillan.

Duesenberry, J. (1965). The Portfolio Approach to the Demand for Money and Other Assets. In U.-N. B. Research, *The State of Monetary Economics* (pp. 9-31). NBER.

Edelberg, V. (1933). The Ricardian Theory of Profits. *Economica*(39), 51-74.

Eichengreen, B., & Mitchener, K. (2003). The Great Depression as a credit boom gone wrong. *BIS Working Papers*(137).

Elgin, C., & Birinci, S. (2016). Growth and Informality: A Comprehensive Panel Data Analysis. *Journal of Applied Economics, 19*(2), 271-292.

Ely, B., & Vanderhoff, V. (1991). Lessons Learned from the S&L Debacle: The Price of Failed Public Policy. *Institute for Policy Innovation Report No. 108*.

Endres, A. (1997). *Neoclassical Microeconomic Theory: The Founding Austrian Version.* London: Routledge.

Estrella, A., & Mishkin, F. (1996). The Yield Curve as a Predictor of U.S. Recessions. *Current Issues in Economics and Finance, 2*(7).

Fama, E. (1971). Risk, Return, and Equilibrium. *Journal of Political Economy, 79*(1), 30-55.

Fama, E. (1980). Banking in the Theory of Finance. *Journal of Monetary Economics*, 39-57.

Fama, E. (2013). Does the Fed Control Interest Rates? *The Review of Asset Pricing Studies, 3*(2), 180-199.

Feenstra, R., Inklaar, R., & Timmer, M. (2015). The Next Generation of the Penn World Table. *American Economic Review, 105*(10), 3150-3182.

Fisher, I. (1907). *The Rate of Interest.*

Fisher, I. (1911). *The Purchasing Power of Money: Its Determination and Relation to Credit, Interest, and Crises.*

Fisher, I. (1912). "The Equation of Exchange" for 1911, and Forecast. *The American Economic Review, 2*(2), 302-319.

Fisher, I. (1929, October 16). Fisher Sees Stocks Permanently High. *The New York Times*.

Fisher, I. (1930). *The Theory of Interest.* Augustus M. Kelly Publishers.

Folsom, R., & Gonzalez, R. (2005). Dimensions and Economics: Some Answers. *The Quarterly Journal of Austrian Economics, 8*(4), 45-65.

Frankel, D. (1998). Transitional Dynamics of Optimal Capital Taxation. *Macroeconomic Dynamics*, 492-503.

Friedman, M. (1977). Time Perspective in Demand for Money. *The Scandinavian Journal of Economics, 79*(4), 397-416.

Friedman, M., & Schwartz, A. (1987). *Money in Historical Perspective.* University of Chicago Press.

Fullarton, J. (1845). *Regulation of Currencies of the Bank of England.* New York.

Garrison, R. (1990). Austrian Capital Theory: The Early Controversies. In B. Caldwell, *Carl Menger and His Legacy in Economics* (pp. 133-154). Durham and London: Duke University Press.

Gesell, S. (1918). *The Natural Economic Order*.

Glasner, D. (1989). *Free Banking and Monetary Reform.* Cambridge University Press.

Glasner, D. (1992). The Real-Bills Doctrine in the Light of the Law of Reflux. *History of Political Economy, 24*(4), 867-894.

Glock, J. (2017). The "Riefler-Keynes" Doctrine and Federal Reserve Policy in the Great Depression. *Working Paper*.

Goldstein, M., & Turner, P. (2004). *Controlling Currency Mismatches in Emerging Markets.* Washington, D.C.: Institute for International Economics.

Goodhart, C., & Perotti, E. (2015). Maturity Mismatching Stretching: Banking Has Taken a Wrong Turn. *CEPR Policy Insight No. 81*.

Gorton, G., & Penacchi, G. (1990). Financial Intermediaries and Liquidity Creation. *Journal of Finance, 45*, 49-72.

Grant, J. (2010, September 17). Over to you, H. Parker Willis. *Grant's Interest Rate Observer*.

Greenwood, R., & Vayanos, D. (2008). Bond Supply and Excess Bond Returns. *NBER Working Paper No. 13806*.

Groenewegen, P. (1999). Reflections on the Centenary of John Bates Clark. *History of Economics Review*(30), 82-88.

Gunning, J. (2005). Praxeological vs Positive time Preference: Ludwig von Mises's Contribution to Interest Theory. *History of Economic Ideas, 13*(1), 105-120.

Gunning, J. (2008). Mises vs. Rothbard on Time Preference and Interest: The Future Prospect of neo-Austrian Growth Theory. *Working Paper*, 1-27.

Gunning, P. (2004). Non-Material Austrian Capital Theory and Loan Markets. *Working Paper*.

Hallerbach, W. (1999). Duration & Dimension. *Working Paper*.

Harcourt, G. (1972). *Some Cambridge Controversies in the Theory of Capital.* Cambridge University Press.

Harrod, R. (1939). An Essay in Dynamic Theory. *The Economic Journal, 49*(193), 14-33.

Hausman, D. (1981). *Capital, Profits, and Prices: An Essay in the Philosophy of Economics.* New York: Columbia University Press.

Hayek, F.A. (1929). *Monetary Theory and the Trade Cycle.* London: Jonathan Cape.

Hayek, F.A. (1931). *Prices and Production.*

Hayek, F.A. (1931, August). Reflections on the Pure Theory of Money of Mr. J.M. Keynes. *Economica*, 270-295.

Hayek, F.A. (1934). On the Relationship between Investment and Output. *Economic Journal, 44*, 207-231.

Hayek, F.A. (1941). *The Pure Theory of Captial.*

Hayek, F.A. (1955). *The Counter-Revolution of Science.* New York: Free Press.

Hayek, F.A. (1973). *Law, Legislation, and Liberty (Vol I).* University of Chicago Press.

Hayek, F.A. (1976). *The Denationalization of Money*. IEA.

Hayek, F.A. (1978). Coping with Ignorance. *Imprimis 7*(7).

Hayek, F.A. (1983, June 11). The Austrian critique. *The Economist*, pp. 45-48.

Haymond, J. (2000). Are MMMFs money? *The Quarterly Journal of Austrian Economics, 3*(4), 53-68.

Hénin, P. (1986). *Macrodynamics: Fluctuations and Growth.* Routledge.

Hicks, J. (1939). *Value and Capital.* Clarendon Press.

Hicks, J. (1965). *Capital and Growth.* The Clarendon Press.

Hicks, J. (1973). *Capital and Time: A Neo-Austrian Theory.*

Hicks, J. (1989). *A Market Theory of Money.* New York: Oxford University Press.

Homer, S., & Sylla, R. (1996). *A History of Interest Rates (Updated Fourth Edition).* Rutgers University Press.

Hoppe, H. (1994). How is Fiat Money Possible?—or The Devolution of Money and Credit. *Review of Austrian Economics, 7*(2), 49-74.

Horwitz, S. (2011). Contrasting Concepts of Capital: Yet Another Look at the Hayek-Keynes Debate. *The Journal of Private Enterprise, 27*(1), 9-27.

Howden, D. (2009). *Financial Asset Pricing Under Knightian Uncertainty.* Madrid.

Huerta de Soto, J. (1998). A Critical Note on Fractional-Reserve Free Banking. *Quarterly Journal of Austrian Economics, 1*(4), 25-49.

Huerta de Soto, J. (2006). *Money, Bank Credit, and Economic Cycles.* Ludwig von Mises Institute.

Huerta de Soto, J. (2008). *The Austrian School: Market Order and Entrepreneurial Creativity.* Edward Elgar.

Huerta de Soto, J. (2008). *The Theory of Dynamic Efficiency.* Routledge.

Huerta de Soto, J. (2010). *Socialism, Economic Calculation and Entrepreneurship.* New Thinking in Political Economy.

Huerta de Soto, J. (2011). A brief note on economic recessions, banking reform and the future of capitalism. In O. Dejuán, E. Febrero, & M. Marcuzzo, *The First Great Recession of the 21st Century: Competing Explanations* (pp. 33-41). Edward Elgar.

Hülsmann, J. (2000). Introduction to Capital & Production by Richard von Strigl. In R. Von Strigl, *Capital & Production* (pp. vii-xxviii). Ludwig von Mises Institute.

Ize, A. (2005). Capitalizing Central Banks: A Net Worth Approach. *IMF Working Paper*.

Jevons, W. (1871). *The Theory of Political Economy.*

Johnson, H. (1982). *On Economics and Society: Selected Essays.* University of Chicago Press.

Jordan, J. (2017). Rethinking the Monetary Transmission Mechanism. *Cato Journal, 37*(2), 361-384.

Kaldor, N. (1937). Annual Survey of Economic Theory: The Recent Controversy on the Theory of Capital. *Econometrica*, *5*(3), 201-233.

Kaldor, N. (1938). On the theory of capital: A rejoinder to Professor Knight. *Econometrica, 6*(2), 163-176.

Kaldor, N. (1939). Capital Intensity and the Trade Cycle. *Economica, 6*(21), 40-66.

Kaldor, N. (1957). A Model of Economic Growth. *The Economic Journal, 67*(268), 591-624.

Keynes, J. (1930). *A Treatise on Money.* New York: Harcourt Brace & Company.

Keynes, J. (1936). *The General Theory of Employment, Interest and Money*. Palgrave Macmillan.

Keynes, J. (1937). The General Theory of Employment. *The Quarterly Journal of Economics, 51*(2), 209-223.

Kihwan, K. (2006). The 1997-98 Korean Financial Crisis: Causes, Policy Response, and Lessons. *The High-Level Seminar on Crisis Prevention in Emerging Markets.* Singapore.

Kirzner, I.M. (1996). *Essays on Capital and Interest.*

Kirzner, I. M. (1960). *The Economic Point of View.* Institute for Human Studies.

Klein, P., & Foss, N. (2012). *Organizing Entrepreneurial Judgment: A New Approach to the FIrm.* Cambridge University Press.

Klüh, U., & Stella, P. (2008). Central Bank Financial Strength and Policy Performance: An Econometric Evaluation. *IMF Working Paper*.

Knight, F.H. (1921). *Risk, Uncertainty and Profit.* New York: Sentry Press.

Knight, F.H. (1931). Professor Fisher's Interest Theory: A Case in Point. *Journal of Political Economy, 39*(2), 176-212.

Knight, F.H. (1934). Capital, Time and the Interest Rate. *Economica, 1*(3), 257-286.

Knight, F.H. (1935). Professor Hayek and the Theory of Investment. *Economic Journal, 45,* 77-94.

Knight, F.H. (1935). The Ricardian Theory of Production and Distribution. *The Canadian Journal of Economics and Political Science, 1*(1), 3-25.

Knight, F.H. (1938). On the Theory of Capital: In Reply to Mr. Kaldor. *Econometrica, 6*(1), 63-82.

Knight, F.H. (1944). Diminishing Returns from Investment. *Journal of Political Economy, 52*(1), 26-47.

Knoll, K., Schularick, M., & Steger, T. (2017). No Price Like Home: Global House Prices, 1870-2012. *American Economic Review, 107*(2), 331-353.

Kurz, H. (2015). Capital Theory, Crises, and Business Cycles: The Triangular Debate between Hayek, Keynes, and Sraffa. *Journal of Reviews on Global Economics*(4), 186-191.

Kydland, F., & Prescott, E. (1990). Business Cycles: Real Facts and a Monetary Myth. *Federal Reserve Bank of Minneapolis Quarterly Review, 14*(2), 3-18.

Lachmann, L.M. (1937). Uncertainty and Liquidity Preference. *Economica, 4*, 295-308.

Lachmann, L.M. (1938). Investment and Costs of Production. *American Economic Review, 28*, 469-481.

Lachmann, L.M. (1941). On the Measurement of Capital. *Economica, 8*, 361-377.

Lachmann, L.M. (1956). *Capital and Its Structure.* Sheed Andrews and McMeel, Inc.

Lachmann, L.M. (1976). On Austrian Capital Theory. In E. Dolan, *The Foundations of Modern Austrian Economics.* Liberty Fund.

Lachmann, L.M. (1977). *Capital, Expectations and the Market Process: Essays on the Theory of the Market Economy.* Sheed Andrews and McMeel, Inc.

Lachmann, L.M. (1956). *Capital and Its Structure.* Bell & Sons.

Laidler, D. (1980). Adam Smith's Monetary Economics. *Economic Working Papers*.

Lavoie, M. (2000). The Reflux Mechanism in the Open Economy. *White Paper*, 1-40.

Lee, H. (1999). Korea's 1997 Financial Crisis: Causes, Consequences and Prospects. *Agenda, 6*(4), 351-363.

Leontief, W. (1953). Domestic Production and Foreign Trade: The American Capital Position Re-Examined. *Proceedings of the American Philosophical Society, 97*(4), 332-349.

Levhari, D. (1965). A Nonsubstitution Theorem and Switching of Techniques. *Quarterly Journal of Economics*, 98-105.

Lewin, P. (1977). Hayek and Lachmann and the Complexity of Capital. In R. Garrison, *The Elgar Companion to Hayek.* Edward Elgar.

Lewis, C., & Pain, N. (2015). Lessons from OECD forecasts during and after the financial crisis. *OECD Journal: Economic Studies, 2014*, 9-39.

Lieber, P. (1931). The Commercial and Financial Future of America. In H. Bodfish, *History of Building and Loan in the United States.* Chicago: U.S. Building and Loan League.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics, 47*(1), 13-37.

Lowe, P. (2015). Address to Thomson Reuters' 3rd Australian Regulatory Summit. Sydney, Australia.

Lucas, R.J. (1976). Econometric Policy Evaluation.

Lucas, R.J. (1990). Why Doesn't Capital Flow from Rich to Poor Countries? *The American Economic Review, 80*(2), 92-96.

Luckett, D. (1959). Professor Lutz and the Structure of Interest Rates. *The Quarterly Journal of Economics, 73*(1), 131-144.

Lutz, F. (1940). The Structure of Interest Rates. *Quarterly Journal of Economics, 55*, 36-63.

Macaulay, F. (1938). *The Movements of Interest Rates. Bond Yields and Stock Prices in the United States since 1856.* New York: National Bureau of Economic Research (NBER).

Machlup, F. (1935). Professor Knight and the "Period of Production". *The Journal of Political Economy, 43*(5), 577-624.

Machlup, F. (1982). *Lecture on Theory of Human Capital.* Islamabad: Pakistan Institute of Development Economics.

Malkiel, B. (2003). Passive Investment Strategies and Efficient Markets. *European Financial Management, 9*(1), 1-10.

Mancini, L., Ranaldo, A., & Wrampelmeyer, J. (2013). Liquidity in the Foreign Exchange Market: Measurement, Commonality, and Risk Premiums. *The Journal of Finance, 68*, 1805-1841.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business, 36*(4), 394-419.

Mandelbrot, B. (2004). *The Misbehavior of Markets: A Fractal View of Financial Turbulence.* Basic Books.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91.

Marx, K. (1894). *Capital.*

McLeay, M., Radia, A., & Ryland, T. (2014). Money creation in the modern economy. *Bank of England Quarterly Bulletin 2014 Q1*, 14-27.

Mehrling, P. (2000). Understanding Fischer Black. *Working Paper*.

Menger, C. (1871). *Principles of Economics.*

Menger, C. (1888). Zur Theorie des Kapitals. *Jahrbücher fúr Nationalökonomie und Statistik*, 1-49.

Menger, C. (1892). On the Origins of Money. *Economic Journal 2,* 239-55.

Merlo, P. (2016). Consequences of the Absence of Monotonicity of the NPV Function to the Assessment of the Effectiveness of Investment Projects. *Engineering Economics, 27*(1), 39-46.

Mill, J.S. (1848). *Principles of Political Economy.*

Mises, von, L.H.E. (1912). *The Theory of Money and Credit.*

Mises, von, L.H.E. (1940). *Nationalökonomie.*

Mises, von, L.H.E. (1949). *Human Action.*

Mitchell, W. (1923). The Institutional Basis for the Shiftability Theory of Bank Liquidity. *The University Journal of Business, 1*(3), 334-356.

Mitchener, K., & Richardson, G. (2012). Shadowy Banks and Financial Contagion during the Great Depression: A Retrospective on Friedman and Schwartz. *Working Paper*.

Modigliani, F., & Miller, M. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review, 48*(3), 261–297.

Modigliani, F., & Sutch, R. (1966). Innovations in Interest Rate Policy. *The American Economic Review, 56*(1/2), 178-197.

Morton, J. (1956). *Urban Mortgage Lending: Comparative Markets and Experience.* Princeton: Princeton University Press.

Moulton, H. (1918). Commercial Banking and Capital Formation: I. *Journal of Political Economy, 26*(5), 484-508.

Moulton, H. (1921). Possibilities of a General-Survey Course in Finance. *Journal of Political Economy, 29*(5), 368-394.

Munn, C. (1982). *Scottish Banking Practice: Banking in Scotland.* Edinburgh: The Institute of Bankers in Scotland.

Noland, M. (2007). South Korea's Experience with International Capital Flows. In S. Edwards, *Capital Controls and Capital Flows in Emerging Economies: Policies, Practices and Consequences* (pp. 481-528). University of Chicago Press.

Northrup Buechner, M. (1976). Frank Knight on Capital as the Only Factor of Production. *Journal of Economic Issues, 10*(3), 598-617.

Osborne, M. (2012). A note on Macaulay’s formula for duration. *Working Paper*.

Osborne, M. (2014). *Multiple Interest Rate Analysis: Theory and Applications.* Palgrave Macmillan.

Osborne, M., & Davidson, I. (2016). The Cambridge capital controversies: contributions from the complex plane. *Review of Political Economy*, 1-20.

Palyi, M. (1936). *Liquidity.*

Pasinetti, L. (1966). Changes in the Rate of Profit and Switches of Techniques. *Quarterly Journal of Economics*, 503-517.

Pastor, L., & Stambaugh, R. (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy, 111*(3), 642-685.

Perelman, M. (1989). *Keynes, Investment Theory and the Economic Slowdown: The Role of Replacement Investment and q-Ratios.* Palgrave Macmillan.

Phelps, E. (1961). The Golden Rule of Accumulation: A Fable for Growthmen. *The American Economic Review*, 638-643.

Piketty, T. (2014). *Capital in the Twenty-First Century.* Harvard University Press.

Piketty, T., & Zucman, G. (2014). Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010. *The Quarterly Journal of Economics*, *129*(3), 1255-1310.

Pyle, D. (1995). The U.S. Savings and Loan Crisis. In R. Jarrow, V. Maksimovic, & W. Ziemba, *Handbooks in Operations Research and Management Science: Finance* (pp. 1103-1125). Elsevier.

Rallo, J. (2014). Desigualdad y capitalismo: los errores de Thomas Piketty. *IX Universidad de Verano.* Instituto Juan de Mariana.

Rallo, J. (2016). *Contra la Teoría Monetaria Moderna.* Ediciones Deusto.

Ricardo, D. (1817). *Principles of Political Economy and Taxation.*

Ricardo, D. (1895). *Letters of David Ricardo to John Ramsay McCulloch, 1816-1823.* Macmillan & Co.

Robinson, J.V. (1953). The Production Function and the Theory of Capital. *Review of Economic Studies, 21*(2), 81-106.

Robinson, J.V. (1953). The Production Function and the Theory of Capital. *Review of Economic Studies, 21*(2), 81-106.

Samuelson, P.A. (1945). The Effect of Interest Rate Increases on the Banking System. *The American Economic Review, 35*(1), 16-27.

Samuelson, P.A. (1966). A Summing Up. *The Quarterly Journal of Economics, 80*(4), 568-583.

Schilling, L. (2017). The Value of the Bank under Endogenous Liquidity Risk: The Modigliani-Miller Theorem Revisited. 1-43. Retrieved from https://ssrn.com/abstract=3054659

Schleifer, A., & Vishny, R. (1997). The Limits of Arbitrage. *The Journal of Finance, 52*(1), 35-55.

Schumpeter, J.A. (1908). *Methodological Individualism.* Brussels: European Institute.

Schumpeter, J.A. (1942). *Capitalism, Socialism and Democracy.* Impact Books.

Schumpeter, J.A. (1954). *History of Economic Analysis.* Allen & Unwin Lltd.

SEC (2016). 17 CFR Parts 210, 270, 274. *Investment Company Liquidity Risk Management Programs*, 1-459.

Shackle, G.L.S. (1989). What Did the General Theory Do? In J. Pheby, *New Directions in Post-Keynesian Economics.* Edward Elgar.

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance, 19*(3), 425-442.

Shostak, F. (2000). The Mystery of the Money Supply Definition. *Quarterly Journal of Austrian Economics, 3*(4), 69-76.

Siegel, J. (1998). *Stocks For The Long Run.* New York: McGraw-Hill.

Silvers, J. (1976). Liquidity, Risk and Duration Patterns in Corporate Financing. *Financial Management, 5*(3), 54-64.

Skousen, M. (2007). *The Structure of Production.* New York and London: New York University Press.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* London.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* London: Methuen & Co., Ltd.

Smith, G., & Xu, A. (2017). Stocks Should Be Valued with a Term Structure of Required Returns. *The Journal of Investing, 26*(2), 61-68.

Smith, H. (2000). *Industry Policy in Taiwan and Korea in the 1980s.* Edward Elgar Publishing.

Smith, V. (1936). *The Rationale of Central Banking and the Free Banking Alternative.* Liberty Press.

Smithers, A., & Wright, S. (2002). *Valuing Wall Street.* McGraw Hill Professional.

Solow, R. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics, 70*(1), 65-94.

Solow, R. (1963). *Capital Theory and the Rate of Return.* Amsterdam: North-Holland Publishing Company.

Solow, R. (1970). *Growth theory.* New York: Oxford University Press.

Soros, G. (2007). *The Alchemy of Finance.* Wiley.

Spitznagel, M. (2011). The Dao of Corporate Finance, Q Ratios, and Stock Market Crashes. *Universa Working White Paper*, 1-11.

Spitznagel, M. (2012). The Austrian and the Swan: Birds of a Different Feather. *Universa Working Paper*.

Spitznagel, M. (2013). *The Dao of Capital: Austrian Investing In a Distorted World.* Hoboken, New Jersey: John Wileys & Sons, Inc.

Sproul, M. (2010). The Law of Reflux. *MPRA Paper No. 24813*, 1-16.

Sraffa, P. (1932). Dr. Hayek on Money and Capital. *The Economic Journal, 42*(165), 42-53.

Sraffa, P. (1960). *Production of Commodities By Means of Commodities.* Cambridge University Press.

Stigler, G. (1941). *Production and distribution theories.* New York: Macmillan.

Stigler, G. (1952). The Ricardian Theory of Value and Distribution. *Journal of Political Economy*, *60*(3), 187-207.

Stigler, G. (1985, December). Frank Hyneman Knight. *Working Paper No. 37*.

Stiglitz, J. (1974). The Cambridge-Cambridge Controversy in the Theory of Capital: A View from New Haven: A Review Article. *Journal of Political Economy, 82*(4), 893-903.

Strigl, von, R. (1934). *Capital and Production.*

Swan, T. (1956). Economic Growth and Capital Accumulation. *The Economic Record, 32*(2), 334-361.

Szabo, N. (2016, December 18). *Weigh and deliver: compensation and the evolution of law and money.* Retrieved November 5, 2017, from Unenumerated: http://unenumerated.blogspot.com/2016/12/weigh-and-deliver-compensation-and.html

Taleb, N.N. (2010). *The Black Swan: The Impact of the Highly Improbable.* Random House Digital, Inc.

Taleb, N.N. (2012). Why It Is No Longer a Good Idea to Be in The Investment Industry. *NYU Working Paper*.

Taleb, N.N. (2015). *Silent Risk, Lectures on Probability.* DesCartes Publishing.

Taleb, N.N., & Pilpel, A. (2004). On the Unfortunate Problem of the Nonobservability of the Probability Distribution. *Working Paper*, 1-38.

Taleb, N.N., & Tapiero, C. (2010). Risk externalities and too big to fail. *Physica A 389*, 3503-3507.

Taylor, T. (1980). *An Introduction to Austrian Economics.*

Tiwari, K. (1994). Single Versus Multiple Discount Rates in Investment Theory. *Journal Of Financial And Strategic Decisions, 7*(2), 19-42.

Tobin, J. (1963). Commercial Banks as Creators of "Money". *Cowles Foundation Discussion Paper no. 159*.

Tobin, J. (1985). Neoclassical Theory in America: J.B. Clark and Fisher. *American Economic Review*(75), 28-38.

Tobin, J., & Brainard, W. (1976). Asset Markets and the Cost of Capital. *Cowles Foundation Discussion Paper no. 427*.

Turner, D. (2016). The Use of Models in Producing OECD Macroeconomic Forecasts. *OECD Working Paper*, 3-22.

Van Winden, M. (2002). *Nederland, de schatkist van Europa.* Amsterdam: Balans.

Vargo, S., & Lusch, R. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing, 68*, 1-17.

Vayanos, D., & Vila, J. (2009). A Preferred-Habitat Model of the Term Structure of Interest Rates. *NBER Working Paper No. 15487*.

Watkins, M. (1919). Commercial Banking and Capital Formation: A Criticism. *Journal of Political Economy, 27*(7), 578-605.

White, L.H. (1999). *The Theory of Monetary Institutions.* Basil Blackwell.

White, L.H. (2012). *The Clash of Economic Ideas.* New York: Cambridge University Press.

White, L.H. (2016). Hayek on Capital and Interest before The Pure Theory of Capital. In F. Hayek, *Capital and Interest.* Routledge.

Wicksell, J.G.K. (1898). *Interest and Prices: A Study on the Causes of Regulating the Value of Money.* New York: Sentry Press.

Williams, J. (1938). *The Theory of Investment Value.*

Wright, S. (2004). Measures of Stock Market Value and Returns for the US Nonfinancial Corporate Sector, 1900-2002. *Review of Income and Wealth*.

Xie, C., Bagozzi, R., & Troye, S. (2008). Trying to prosume: toward a theory of consumers as co-creators of value. *Journal of the Academy of Marketing Science, 36*(1), 109-122.

Yoshitomo, M., & Ohno, K. (1999). Capital-Account Crisis and Credit Contraction: The New Nature of Crisis Requires New Policy Responses. *ADB Institute Working Paper*.

Yuki, T. (2015). Reconsideration of Gesell's Critique of Marx's Capital Theory. *Uno Newsletter*, *2*(15), 1-24.

Zarlenga, S. (1994). *Refutation of Menger's Theory of the Origin of Money.* American Monetary Institute Charitable Trust.

Index