

159

CIENCIAS  
JURÍDICAS Y SOCIALES



METODOLOGÍA  
PARA INVESTIGACIONES  
DE ALTO IMPACTO  
EN LAS  
CIENCIAS SOCIALES Y JURÍDICAS

EDITORES/ DIRECTORES

Karla Annet Cynthia Sáenz López  
Francisco Javier Gorjón Gómez  
Marta Gonzalo Quirga  
Cástor Miguel Díaz Barrado



Diseño de cubierta: Marta Gonzalo Quirga

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

UNIVERSIDAD REY JUAN CARLOS

- Editorial: Dykinson
- ISBN: 978-84-9031-267-4
- Páginas: 306
- Fecha de la edición: Enero 2013
- Edición: 1ª ed.
- <http://www.editorialdykinson.com/libros/metodologia-para-investigaciones-de-alto-impacto-en-las-ciencias-sociales-y-juridicas/9788490312674/>

# **METODOLOGÍA PARA INVESTIGACIONES DE ALTO IMPACTO EN LAS CIENCIAS SOCIALES**

## **Editores/Directores**

**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN**

Dra. Karla Annet Cynthia Sáenz López

Dr. Francisco Javier Gorjón Gómez

**UNIVERSIDAD REY JUAN CARLOS**

Dra. Marta Gonzalo Quiroga

Dr. Cástor Miguel Díaz Barrado

# **METODOLOGIA PARA INVESTIGACIONES DE ALTO IMPACTO EN LAS CIENCIAS SOCIALES**

## **Autores**

**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN**

Dra. Mónica Blanco Jiménez

Dra. Paula Villalpando Cadena

Dr. Joel Mendoza Gómez

Dra. Karla Annett Cynthia Sáenz López

Dr. Francisco Javier Gorjón Gómez

Dra. Martha del Pilar Rodríguez García

Dr. Klender Aimer Cortez Alejandro

Dr. Mohammad Hosein Badii Zabeh

Dr. Juan Rositas Martínez

Dr. Gustavo Alarcón Martínez

Dr. José Zaragoza Huerta

Dr. Luis Gerardo Rodríguez Lozano

Dra. Karla Rodríguez Burgos

**UNIVERSIDAD REY JUAN CARLOS**

Dra. Marta Gonzalo Quiroga

Dra. Sagrario Morán Blanco

Dra. Nuria Alonso

Dra. Ma. Ángeles Cano Linares

Dra. Cintia Dfaz-Silveira Santos

Dra. Gabriela C. Cobo del Rosal Pérez

Dr. Jesús-María Navalpotro Sánchez-Peinado

Dr. David Trillo



**UANL**

UNIVERSIDAD AUTÓNOMA  
DE NUEVO LEÓN  
Secretaría de Extensión y Cultura

# ÍNDICE

1. Nociones científicas del protocolo de investigación. Mónica Blanco Jiménez y Paula Villalpando Cadena.
2. La teoría en la investigación científica. Marco teórico, modelos y medición. Joel Mendoza Gómez.
3. Métodos cualitativos aplicados al derecho. Karla Annett Cynthia Sáenz López. Francisco Javier Gorjón Gómez. Marta Gonzalo Quiroga.
4. Nuevos métodos cuantitativos aplicados a las Ciencias Sociales. Martha del Pilar Rodríguez García. Klender Aimer Cortez Alejandro.
5. Investigación cuantitativa. Diseño, técnicas, muestreo y análisis cuantitativo. Karla Eugenia Rodríguez Burgos.
6. Cierre y evaluación de la investigación. Juan Rositas Martínez y Gustavo Alarcón Martínez.
7. La investigación del derecho y la ética. Luis Gerardo Rodríguez Lozano. José Zaragoza Huerta.
8. Metodología e investigación científica en las ciencias sociales. Especial referencia a la disciplina de las relaciones internacionales. Sagrario Morán Blanco.
9. Limitaciones de la aplicación e interpretación de modelos econométricos como herramienta de investigación social. Nuria Alonso. David Trillo.
10. Metodología y técnica de investigación en relaciones internacionales y derecho internacional público. Una aplicación concreta: la elaboración, dirección y evaluación de trabajos fin de máster, impartido en modalidad on line. Ángeles Cano Linares.
11. Internet, una herramienta imprescindible para las investigaciones socio-jurídicas del siglo XXI. Cintia Díaz-Silveira Santos.
12. Multidisciplinaridad en ciencias sociales. Historia-Arte-Derecho. Gabriela C. Cobo del Rosal Pérez. Jesús-María Navalpotro Sánchez-Peinado.



# Capítulo 10

## LIMITACIONES DE LA APLICACIÓN E INTERPRETACIÓN DE MODELOS ECONOMÉTRICOS COMO HERRA- MIENTA DE INVESTIGACIÓN SOCIAL

Dra. Nuria Alonso, Dr. David Trillo

### Introducción

Las cuestiones de método en economía han sido objeto de un amplio debate a lo largo del tiempo. La economía se incardina en el ámbito de las ciencias sociales, puesto que se trata de una disciplina en la que no es posible realizar experimentos controlados y extraer leyes generales que permitan enunciar teorías universales. Por ese motivo caben distintas escuelas de pensamiento y diferentes aproximaciones para analizar fenómenos importantes como la inflación, el crecimiento económico o los efectos de los impuestos, entre otros muchos. En el momento actual además del predominio en revistas de investigación de estudios basados en teorías neoclásicas es general el uso de modelos econométricos como método de contrastación de los resultados de las investigaciones. El presente artículo pretende reflexionar sobre los límites en el uso de la econometría y sobre puntos metodológicos relevantes a la hora de interpretar la información obtenida a partir de los citados modelos. Para ilustrar el problema desarrollaremos los argumentos al hilo de un modelo utilizado para evaluar la vinculación entre el rendimiento de una acción cotizada y el de un índice global de bolsa.

### 1. Los datos y la información no revelada

En general en las investigaciones en economía es frecuente encontrar estudios que confunden la realidad con los datos disponibles. Hay que reflexionar sobre este problema

que a menudo viene provocado por la prisa en la realización de las investigaciones o la presión por acceder a una acreditación y a las plazas docentes en la universidad (cuestiones habituales en nuestros días, que el propio sistema de evaluación alienta). La realidad social se compone de fenómenos complejos; pensemos en un fenómeno como el desempleo, que está lógicamente vinculado al ciclo económico pero además depende de cómo esté regulado su acceso en cada país, de la protección al trabajador en el mercado de trabajo, de la propia idiosincrasia del trabajador o satisfacción social por el hecho de desempeñar un puesto concreto, de la sobreeducación o la escasa cualificación del trabajador y de otros muchos fenómenos como el propio tipo de empleo o la raramente cuestionada capacidad gerencial de los empresarios. Esa realidad del desempleo se intenta luego recoger con un indicador como los desempleados que se registran en las oficinas del INEM o la tasa de desempleo estimada a través de la Encuesta de Población Activa. Es evidente que los datos recogen parcialmente la realidad, de hecho la economía irregular coloca a gran parte de los trabajadores que sobreviven en la tesitura de tener que combinar todas sus fuentes de renta. Por ese motivo no deberíamos considerar que el problema es idéntico para países con una cifra de paro compatible con el pleno empleo del 5% como el Reino Unido que para países como España en el que en plena etapa de crecimiento del 1996 al 2007 nunca tuvo una cifra inferior al ...

También es importante definir exactamente que significa de forma práctica cada indicador económico que se utilice para estudiar la realidad. Por ejemplo cómo se miden fenómenos como la inflación, la producción o qué se entiende por masa monetaria concretamente.

Otra cuestión importante es que los datos proceden de un único procedimiento de obtención y son los que están disponibles en cada momento. Es decir, no son ninguna muestra aleatoria. Una muestra es aleatoria cuando en un momento determinado del tiempo podemos extraer diferentes valores de la variable que queremos estudiar mediante la realización de un experimento controlado. En un supuesto de análisis del crecimiento agrícola, se puede observar el crecimiento de una planta en un determinado terreno y condiciones climatológicas, con un nivel de abono determinado y con el agua como principal input. Con ese punto de partida un experimento controlado consistiría, por ejemplo, en tomar muestras en 100 terrenos de igual dimensión e iguales condiciones de producción. De esa manera se puede evaluar la estabilidad de los resultados relativos al crecimiento de las plantas en los distintos terrenos. Además se conoce el tiempo de realización de los procesos.

En las ciencias sociales en general y en economía en particular no es posible realizar este tipo de experimentos controlados; una variable importante como el producto interior bruto es el resultado de una única medición, y no se conocen con exactitud ni la ley

que rige el proceso (o factores fundamentales de crecimiento) ni el tiempo en el que el mismo se concreta. Cuando se dice que el Producto Interior Bruto crece por el efecto del aumento del stock de capital, suponiendo que los datos permitan sostener esta afirmación y traducir esta realidad, realmente habría que plantearse en qué plazo se produce ese efecto porque en la práctica es fácil encontrar estudios que toman datos contemporáneos para analizar estas relaciones.

Adicionalmente, en las relaciones sociales existen otros factores ocultos que influyen en los fenómenos, que no son identificables pero subyacen. Además de los problemas inherentes a la estadística de las variables disponibles, habría que realizar un gran esfuerzo por reforzar el trabajo de campo para obtener nuevas vías de estudio de la realidad. En cualquier caso lo que no debería hacerse en la investigación social es obviar las limitaciones de los datos y dar conclusiones sin matizar este extremo. Realmente de las investigaciones sociales que hacen uso instrumental de la estadística lo que se va a obtener siempre es un conjunto de resultados cuyas conclusiones serán abiertas y podrán encajar en diferentes corrientes de pensamiento económico, sociológico o de otro tipo.

## 2. El modelo

Los modelos son representaciones esquemáticas de la realidad que hacen un uso instrumental de las matemáticas, lo que les confiere una coherencia interna: las propiedades del modelo se derivan de las hipótesis de partida. Las estimaciones deberían ser coherentes con las hipótesis, porque si no el trabajo no estaría correctamente realizado desde una perspectiva científica.

El modelo que sirve de base para todo el desarrollo de la econometría es el modelo de regresión lineal es:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Donde  $x_t$  e  $y_t$  son las realizaciones muestrales de las variables X e Y. El subíndice t se utiliza en el análisis de series en el tiempo y en ese caso se estaría contando con una base de datos de diferentes periodos de tiempo, años, días, semanas... para ambas variables analizadas. Cuando el análisis es de sección cruzada se seleccionan datos de diferentes individuos, regiones o instituciones para un momento determinado del tiempo. Los datos con los que se estima el modelo pueden estar en sus variables originales de medida o transformados en tasas o en cualquier otra función de tipo logarítmico o similar.

Un punto importante, que actúa de limitación en la investigación, es que los inputs o  $x_t$  son números fijos, es decir no dependen de otro proceso o varían cada año en función

del mismo. Si por ejemplo se intenta explicar el crecimiento en función de la capacidad de exportación y en el modelo que sustenta la estimación consideramos que la variable exportaciones depende a su vez de la evolución del tipo de cambio (variable aleatoria), habría que utilizar un método de estimación adecuado al supuesto de que las variables explicativas son también aleatorias. En caso de que se utilice el modelo de mínimos cuadrados ordinarios para estimarlo, hay que asumir el presupuesto de que las exportaciones son un número fijo, con todas las limitaciones teóricas que implica este presupuesto.

Otra cuestión que suele explicarse poco en los trabajos econométricos es la razón por la que se utiliza una determinada transformación en los datos originales. Por ejemplo si se ha optado por tomar logaritmos o por calcular tasas o primeras diferencias. Estas transformaciones no deberían ser una cuestión azarosa sino que deberían estar fundamentadas en que el modelo estimado esté correctamente especificado. En el apartado siguiente se analizan las prevenciones que habría que tener en este sentido.

La aleatoriedad del modelo de regresión viene dada por el término de error  $\varepsilon_t$ . Esta variable aleatoria generalmente se supone con esperanza nula y varianza constante y con covarianza nula.

$$E(\varepsilon_t) = 0 \quad \text{Var}(\varepsilon_t) = \sigma^2 \quad E(\varepsilon_t, \varepsilon_{t-j}) = 0 \quad \forall j \neq 0$$

Los desarrollos econométricos a partir del modelo estimado por mínimos cuadrados ordinarios (MCO) se han centrado en establecer correcciones vinculadas a los dos últimos supuestos: la heterocedasticidad, en los supuestos en varianza no constante, y la autocorrelación, cuando la covarianza no es nula, lo que implica que existen relaciones entre las perturbaciones aleatorias en fechas distintas.

De los presupuestos del modelo se deducen sus propiedades. Una de las propiedades básicas es que el parámetro estimado  $\hat{\beta}$  es un estimador insesgado de la variable aleatoria  $\beta$ . Esto quiere decir que la esperanza del beta estimado  $\hat{\beta}$  coincide con el valor de  $\beta$ . Es un tema importante porque eso ocurre en el infinito, es decir si estimásemos múltiples veces  $\beta$  con diferentes extracciones muestrales de  $y_t$  en media el valor sería el de la variable. Sin embargo, en la práctica solo es posible realizar una única estimación de  $\beta$  que en general difiere de su verdadero valor.

Una forma de ilustrar que los parámetros estimados no coinciden con los parámetros del modelo o que el estimador de la varianza de las perturbaciones aleatorias no coinciden con el verdadero valor de la varianza de las perturbaciones, es simular un modelo en el que se establezca un valor cualquier de  $\alpha$  y  $\beta$ , una serie preestablecida de datos  $x_t$  y en el que adicionalmente se genera una muestra de  $t$  datos de una variable aleatoria normal

de media cero y desviación típica dada. Con eso se obtiene un “modelo verdadero” en el que la variable dependiente es función de las variables independientes elegidas para la ecuación objeto de simulación. A continuación, se estimaría un modelo de regresión lineal por mínimos cuadrados ordinarios tomando la  $y_t$  generada y la  $x_t$  original.

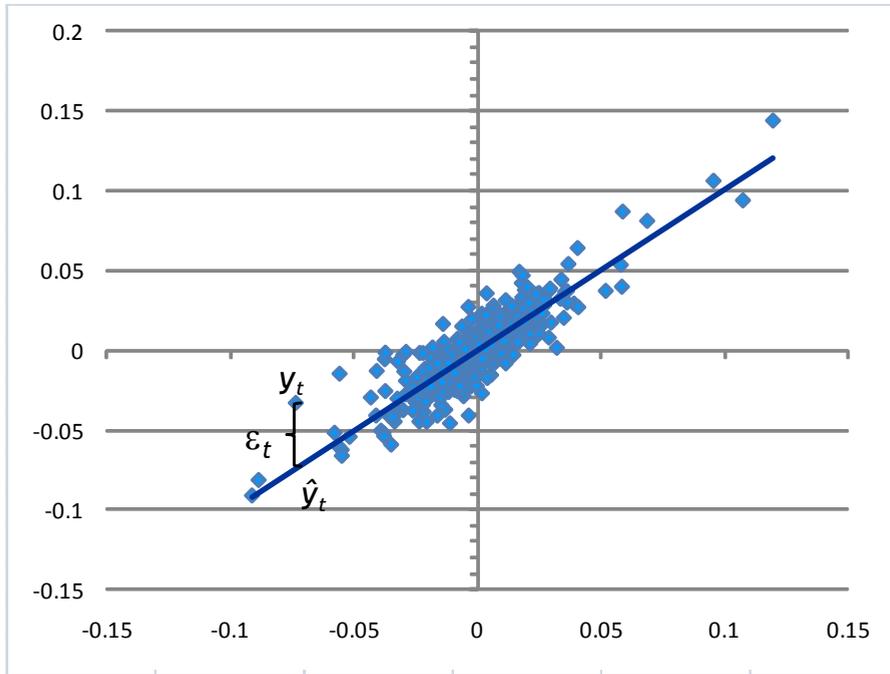
Con este ejercicio se puede comprobar que los parámetros que resultan de una estimación de, por ejemplo de 200 datos, no coinciden con los que se han establecido en el modelo de referencia, si bien los estimadores obtenidos son próximos a los verdaderos valores. La razón reside en la aleatoriedad del término de error. Si se repitiese el experimento cada vez con una muestra distinta extraída de una variable aleatoria normal para simular el término de error, cada vez que estimemos  $\alpha$  y  $\beta$  saldrán números diferentes, y lo mismo ocurrirá con la estimación de la varianza de las perturbaciones aleatorias. El valor medio de los estimadores obtenidos sí será muy similar al verdadero valor de los parámetros objeto de estimación, cuando el número de veces que se simula el modelo es grande.

Otro tema interesante es mantener los datos elegidos para  $\alpha$ ,  $\beta$  y  $x_t$  para generar un modelo de referencia pero estimarlo presuponiendo una varianza mayor para las perturbaciones del modelo. Si se amplía la varianza, el modelo tendrá un  $R^2$  menor.

El  $R^2$  es un indicador del grado de ajuste de los datos a la recta de regresión

$$R_1^2 = 1 - \frac{SR}{ST} = 1 - \frac{\sum_{i=1}^N \hat{\varepsilon}_t^2}{\sum_{i=1}^N (y_t - \bar{y})^2} = 1 - \frac{\sum_{i=1}^N (y_t - \hat{y}_t)^2}{\sum_{i=1}^N (y_t - \bar{y})^2}$$

**Gráfico1:**  
**Gráfico de dispersión**  
**(relación entre las rentabilidad del IBEX y de una acción)**



Gráficamente, en un modelo con una única variable explicativa, el numerador del cociente que aparece restando en la ecuación anterior será cero si el dato del cuadrado de la suma de desviaciones de variable dependiente respecto a su estimación<sup>□</sup> y en ese caso el  $R^2$  es igual a la unidad. En el gráfico 1 cuanto más cercanos estén los datos a la recta de regresión más cercano a uno será el  $R^2$  y cuanto más lejanos a la recta peor será la capacidad explicativa del modelo. Esto liga con lo señalado en el párrafo anterior, en el que se proponía un ensayo consistente en ampliar la varianza de un modelo simulado, puesto que si aumenta la varianza de las perturbaciones aleatorias peor ajuste tiene ese modelo.

Otro tema fundamental es que el parámetro estimado  $\hat{\beta}$  o estimador MCO es aquel que minimiza la suma de los cuadrados de los residuos. Esta propiedad hace que la varianza del estimador dependa igualmente de la varianza de los residuos. El intervalo de confianza viene definido por esa varianza, de manera que si los residuos plantean problemas de especificación no debería hacerse inferencia a partir de los  $\hat{\beta}$ . En el apartado siguiente se plantea que una salida de datos de un paquete

econométrico no debería interpretarse hasta tener la seguridad de que los residuos cumplen con las hipótesis del modelo.

Otro tema relevante es el del número de datos utilizados para la estimación, dado que cuanto más grande sea la muestra, más cercanos serán los valores a los reales, en la medida en que será más difícil que una observación extrema modifique notablemente el valor del parámetro.

### **3. Un paso previo a la interpretación de los resultados**

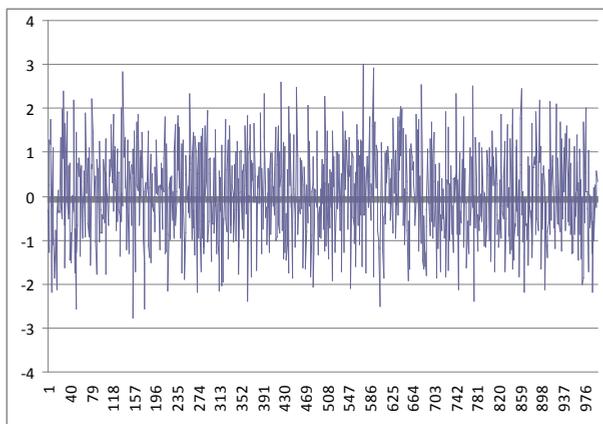
Es habitual que en los trabajos de investigación se presenten las estimaciones como “prueba” de la solidez de las tesis planteadas. La interpretación directa de los resultados de una salida de un paquete econométrico carece de sentido, previamente hay que realizar un análisis preliminar para validar si el modelo especificado cumple con las hipótesis básicas del modelo planteado para la estimación. Como hemos visto, el modelo básico de estimación por mínimos cuadrados ordinarios plantea unas hipótesis muy concretas y constituye de alguna manera el método seminal cuyas transformaciones han dado lugar a extensiones y nuevos modelos en economía.

Un punto fundamental que hay que analizar es si los residuos que se derivan de la estimación son estacionarios en media y varianza. Si la hipótesis inicial del término de error era que seguía una distribución de esperanza nula y desviación típica constante, hay que comprobar si no es rechazable que la realización muestral de esa variable aleatoria, los residuos estimados, sea compatible con la hipótesis inicial.

Uno de los problemas que estamos viviendo en el campo de la investigación en estos momentos es la facilidad de uso del software econométrico y la formación de econometría a partir de esos paquetes estadísticos. Los modelos, como hemos visto, no soportan cualquier estimación, pero los programas informáticos ofrecen igualmente resultados a cualquiera que sepa ejecutar unos mínimos comandos; es el investigador el que debe conocer bien el modelo teórico y corregir las estimaciones si se quiere realizar el trabajo de una manera “limpia” metodológicamente hablando.

Como se ha apuntado anteriormente, la simulación de una variable aleatoria de media cero y desviación típica uno es una referencia para comparar los residuos “ideales o verdaderos” con los estimados. El gráfico que se obtiene de la extracción de una muestra de 1000 datos de una variable normal  $N(0,1)$  es el siguiente:

**Gráfico 2:**  
**Muestra de 1000 datos generados para una distribución normal N(0,1)**



Conociendo la gráfica de la distribución “ideal” para los residuos es posible su comparación con la grafica de los residuos resultantes de una estimación.

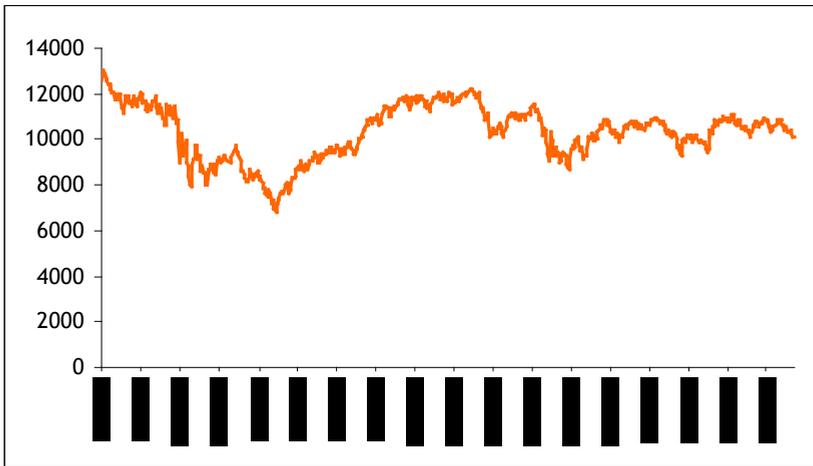
Para ello hay varias aproximaciones siguiendo la metodología Box-Jenkins de análisis de series temporales. Lo más intuitivo y preliminar es realizar un gráfico con los residuos estimados. En la tabla 1 se presentan los resultados de una regresión del IBEX respecto al precio de las acciones de Telefónica. Es una aproximación al beta de una acción, si bien incorrectamente especificada por motivos que se comentan en adelante.

**Tabla 1:**  
**Regresión del IBEX respecto al precio de las acciones de Telefónica**

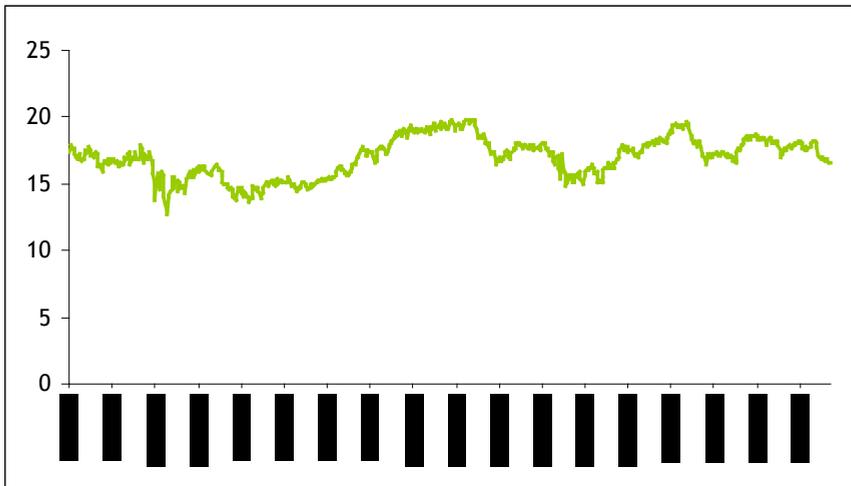
Dependent Variable: TEL				
Method: Least Squares				
<b>Telefónica con Ibx en nivel</b>				
Sample: 1 752				
Included observations: 752				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6,310	0,269	23,444	0,000
IBEX	0,001	0,000	39,870	0,000
R-squared	0,679	Mean dependent var		16,971
Adjusted R-squared	0,679	S.D. dependent var		1,475
S.E. of regression	0,836	Akaike info criterion		2,481
Sum squared resid	523,651	Schwarz criterion		2,494
Log likelihood	-930,963	F-statistic		1.589,583
Durbin-Watson stat	0,029	Prob(F-statistic)		0,000

Los resultados obtenidos no tienen validez porque existe una estructura no estacionaria de los residuos, y en consecuencia no habría que seguir interpretando ningún resultado, ni t-ratios ni significatividad de los coeficientes por supuesto (esos t-ratios también están condicionados por la naturaleza no estacionaria de los datos originales). Si las series de partida no son estacionarias, se produce el fenómeno de autocorrelación por el crecimiento idéntico de medias en el tiempo en las series originales en el tiempo, es decir que cuando una variable aumenta la otra también. En el supuesto que estamos utilizando como ejemplo, si seleccionásemos subperíodos de tiempo desde el origen hasta el último dato y sacáramos la media de las variables del modelo en cada subperíodo, dichas medias serían variables en el tiempo y con tendencias de crecimiento similares (Véase gráfico 3 y gráfico 4). En el caso de otras variables macroeconómicas el efecto es aun más claro porque suelen tener tendencias crecientes de largo plazo, como en el caso de la producción o del empleo.

**Gráfico3:**  
**Evolución del IBEX**



**Gráfico 4:**  
**Evolución de la cotización de Telefónica**



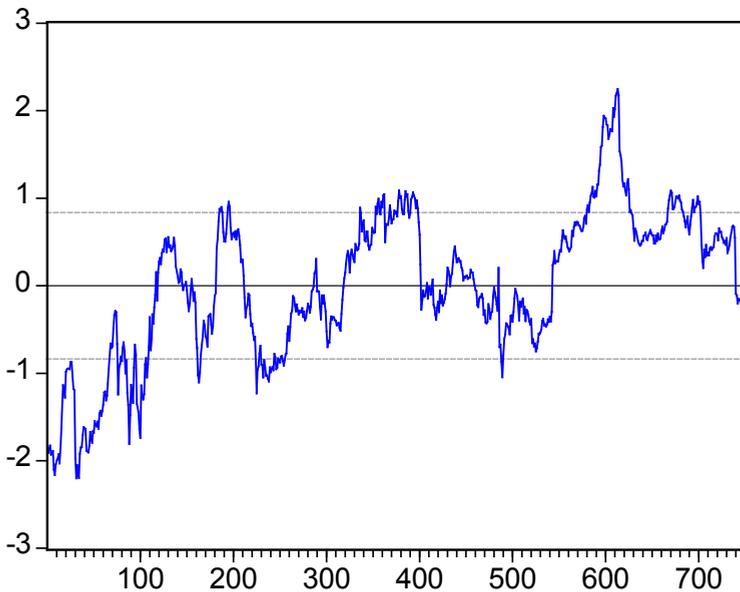
Además, la autocorrelación se puede analizar fácilmente a partir de la expresión del coeficiente de correlación entre dos variables. El parámetro  $\hat{\beta}$  puede expresarse en función del coeficiente de correlación, de manera que la interpretación que se de al mismo es igualmente válida para el parámetro.

$$\rho = \frac{\sum_{t=1}^N (x_t - \bar{x})(y_t - \bar{y})}{\sigma_x \sigma_y}$$

Lo relevante es el análisis del numerador de esta expresión porque las desviaciones típicas de las variables analizadas son siempre positivas y actúan como un factor de ponderación. En el numerador ocurre que en dos series no estacionarias si aumenta o disminuye una serie por encima de su media, la otra hará lo mismo la mayor parte de las veces, dando lugar a un resultado de correlación positiva y alta.

En la estimación de la tabla 1 del principio del apartado se han tomado las series originales en nivel. Si las variables originales no son estacionarias el modelo genera este tipo de gráficos de residuos:

**Grafico 5:**  
**Residuos de la estimacion del beta de Telefonica (IBEX)**



Además del análisis gráfico de los residuos, otros instrumentos que permiten evaluar la estacionariedad de los residuos son las funciones de autocorrelación y autocorrelación parcial.

La función de autocorrelación (en adelante ACF) permite analizar la relación que existe entre el residuo estimado y el de los periodos anteriores al analizado.

Si se define la varianza de los residuos como

$$\hat{\gamma}_0 = \hat{\sigma}^2 = \frac{\sum_{t=1}^N (\hat{\varepsilon}_t - \bar{\varepsilon})^2}{N}$$

donde  $\hat{\varepsilon}_t$  son los residuos estimados  $\bar{\varepsilon}$  es su media

La función de autocorrelación es

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \text{ donde } \hat{\gamma}_k \text{ es la covarianza de los errores entre un periodo y el de otro}$$

periodo anterior en el tiempo. Un valor positivo implica que el residuo está vinculado a otro momento temporal, ya sea dependiendo de los datos un mes anterior o tres, etc.

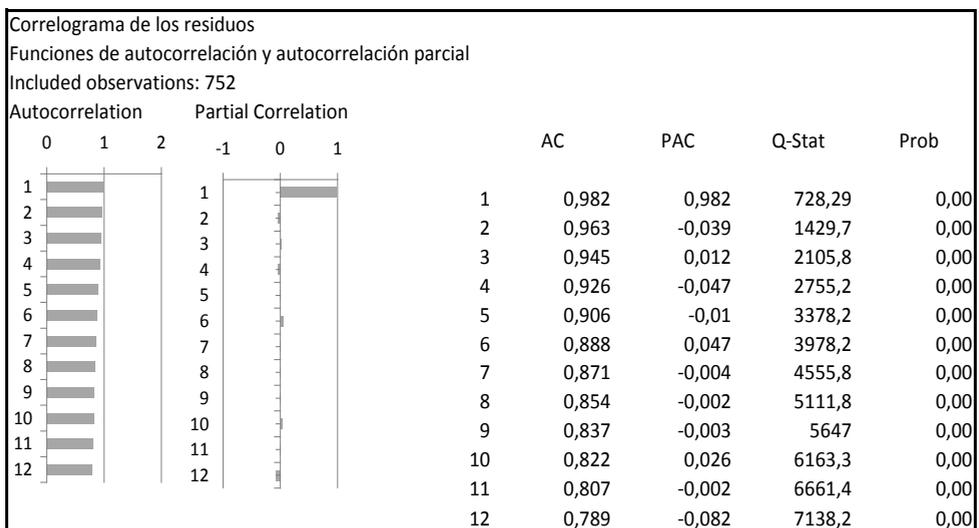
La función de correlación parcial (en adelante PACF) se obtiene de la ecuación:

$$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \varphi_2\varepsilon_{t-2} + \dots + \varphi_k\varepsilon_{t-k} + a_t; \quad a_t \sim N(0, \sigma_a^2) \quad E(a_t, a_{t-j}) = 0 \quad \forall j \neq 0$$

que relaciona los residuos estimados con la suma de cada periodo anterior. En esa ecuación hay un término aleatorio que sigue una distribución normal de esperanza cero y varianza constante. La función de autocorrelación parcial de orden K viene representada por cada  $\varphi_k$ .

Las funciones de autocorrelación en el modelo que se ha utilizado en este apartado para ilustrar este problema serían:

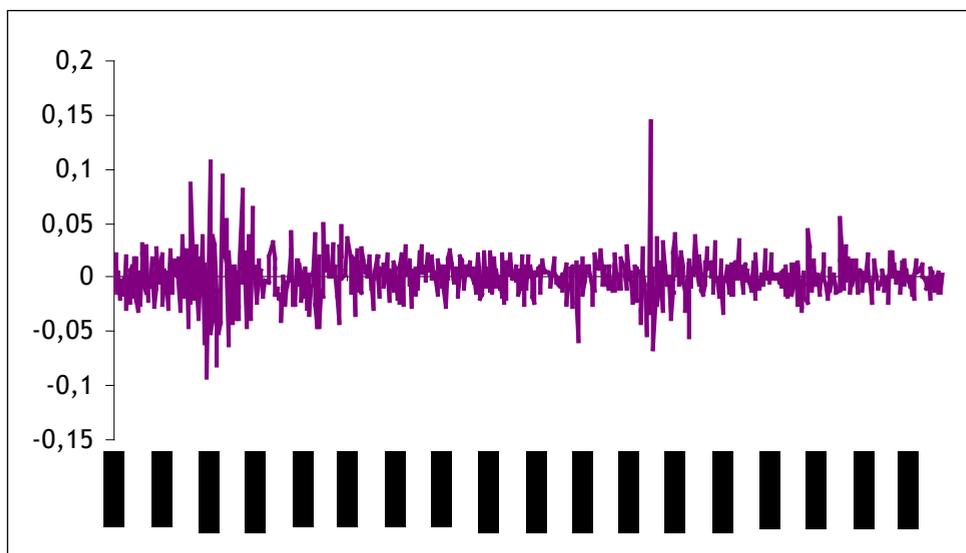
**Tabla 2:**  
**Funciones de autocorrelación vinculadas a la regresión de la tabla 1**



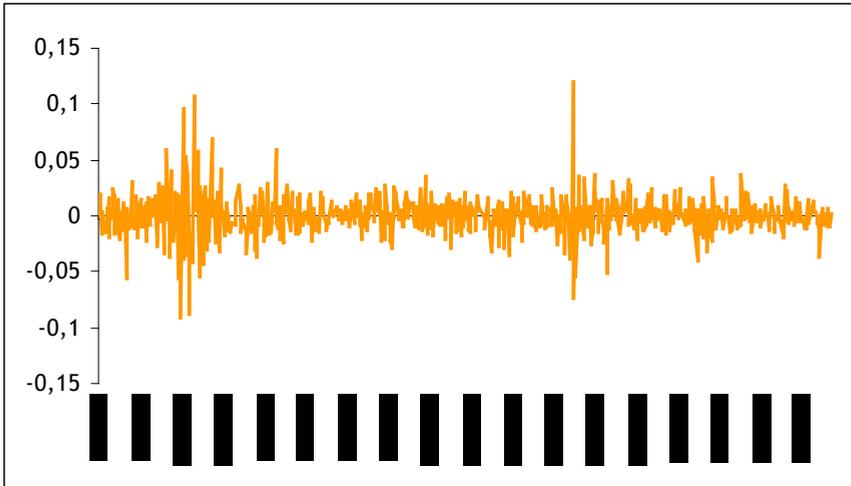
En la tabla 2 se recogen las funciones de autocorrelación de orden uno hasta el 12, lo que permite evaluar la vinculación que existe entre los residuos estimados para el periodo  $t$  y el periodo  $t-k$ ,  $k = 1 \dots 12$ . Si existe correlación de orden uno, gráficamente la función de autocorrelación parcial con unos residuos no estacionarios de este tipo tiene el aspecto de la tabla con una primera barra a la derecha.

A continuación se ha presentado una estimación con las variables transformadas en rentabilidades mediante la tasa de variación diaria. Es una posible transformación que permite corregir el problema de no estacionariedad. Otras opciones son tomar primeras diferencias o el logaritmo de la primera diferencia (en este último caso es equivalente a calcular la rentabilidad). Como en el mundo financiero se manejan las rentabilidades para la estimación de los betas se ha optado por esta modalidad. Sin embargo, la justificación del uso de una u otra transformación debería basarse en que conducen a una mejor especificación del modelo. En gráfico 6 y 7 se observa cómo al calcular la rentabilidad del IBEX y de Telefónica en lugar de las variables en nivel los datos fluctúan en torno a una media constante. Sin embargo, los datos indican que existe algún periodo que puede considerarse atípico y dan muestra de una varianza algo mayor en el comienzo de la serie, que coincide con el momento de máximas turbulencias financieras en 2008.

**Gráfico 6:**  
**Evolución de la rentabilidad diaria del IBEX**



**Gráfico 7:**  
**Evolución de la rentabilidad diaria del IBEX**



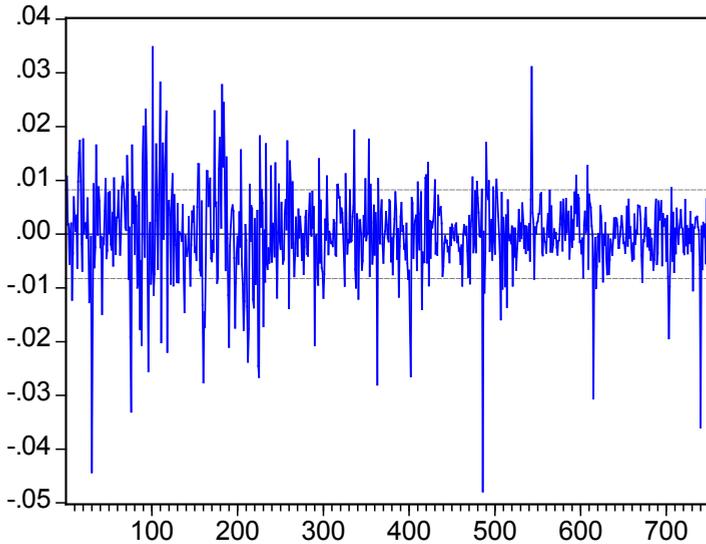
Si se estima el modelo de regresión con la rentabilidad de telefónica como variable independiente y la del IBEX como variable dependiente se obtiene la siguiente información:

**Tabla 3:**  
**Estimación del beta de la rentabilidad de Telefónica (IBEX)**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0,000	0,000	0,584	0,560
RIBEX	0,771	0,015	50,764	0,000
R-squared	0,775	Mean dependent var		0,000
Adjusted R-squared	0,774	S.D. dependent var		0,017
S.E. of regression	0,008	Akaike info criterion		-6,754
Sum squared resid	0,051	Schwarz criterion		-6,742
Log likelihood	2541,480	F-statistic		2576,962
Durbin-Watson stat	1,794	Prob(F-statistic)		0,000

Se puede observar que la grafica de residuos que se deriva de la estimación tiene una estructura similar a la presentada al principio del apartado, como resultado de una extracción muestral generada a partir de una distribución de probabilidad normal de media cero y desviación típica constante.

**Gráfico 8:**  
**gráfico de residuos a partir de la estimación de la tabla 3**

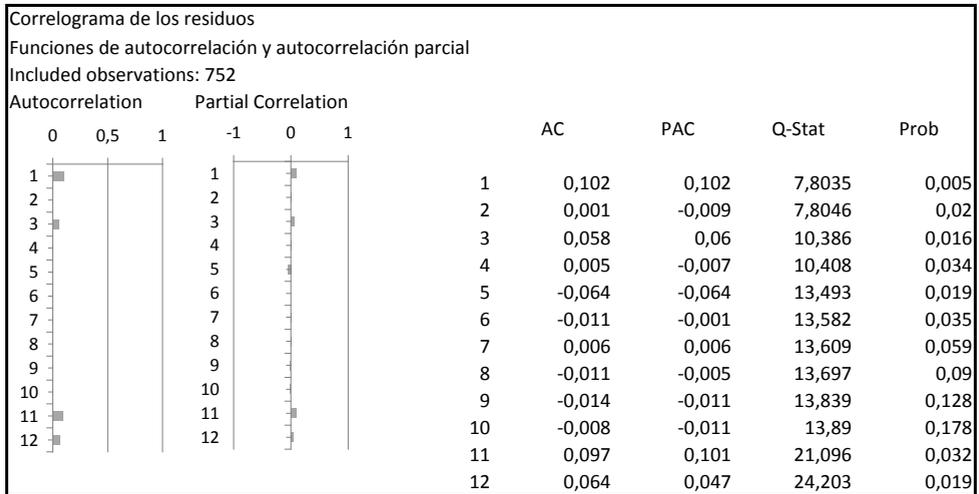


El gráfico ACF y PACF de la tabla 4 no presentan una figura que muestre tan claramente autocorrelación como en el supuesto anterior; además en este caso los valores extremos de los residuos pueden generar una apariencia de falsa correlación<sup>1</sup>.

---

1 El estadístico Durwin-Watson permite analizar si esos residuos tienen autocorrelación de orden uno, para modelos con constante. En cualquier caso, hay dos cuestiones relevantes en relación con esta prueba, la primera es que solo sirve para analizar la autocorrelación entre los residuos estimados para el año  $t$  y el  $t-1$ , lo cual no permite analizar otro tipo de autocorrelación. La segunda afecta en general a los test de hipótesis, y que es que están pensados para muestras grandes y carecen de sentido para estimaciones con pocos datos.

**Tabla 4:**  
**Resultados de la regresión de la rentabilidad del IBEX**  
**respecto a la de Telefónica**



En el caso de las series temporales es habitual que los residuos sigan teniendo estructura aunque se hayan hecho las transformaciones oportunas, como la primera diferencia o la transformación en rentabilidades y, en ese caso, es necesario identificar el proceso estocástico subyacente.

#### 4. Interpretación de los resultados:

En la tabla 3 del apartado anterior el parámetro ligado a la rentabilidad de telefónica es significativo pero la constante no. En ese caso lo lógico es reestimar quitando la constante<sup>2</sup>.

<sup>2</sup> La estimación de un modelo sin constante tiene implicaciones porque en ese escenario las distintas fórmulas de cálculo del  $R^2$  ya no son coincidentes. El  $R^2$  definido en este capítulo puede ser negativo y por tanto ya no estará comprendido entre cero y uno.

**Tabla 5**  
**Estimación del modelo sin constante**

Dependent Variable: RTEL				
Method: Least Squares				
Rentabilidad de Telefónica con Ibex				
Sample: 1 752				
Included observations: 752				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RIBEX	0.771	0.015	50.784	0.000
R-squared	0.774	Mean dependent var		0.000
Adjusted R-squared	0.774	S.D. dependent var		0.017
S.E. of regression	0.008	Akaike info criterion		-6.756
Sum squared resid	0.051	Schwarz criterion		-6.750
Log likelihood	2541.309	Durbin-Watson stat		1.794

Vamos a estimar el modelo suponiendo que el término de error se considera dependiente del error en t-1 más un término de perturbación aleatoria que sigue una distribución normal sin autocorrelación. El nuevo modelo ~~corregido~~ sería el siguiente:

$$y_t = \alpha + \beta x_t + \varepsilon_t \text{ donde}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + a_t; a_t \sim N(0, \sigma_a^2) \quad E(a_t, a_{t-j}) = 0 \quad \forall j \neq 0$$

**Tabla 6**  
**Estimación de un modelo AR(1) sin constante**

Dependent Variable: RTEL				
Method: Least Squares				
<b>Rentabilidad de Telefónica con Ibx y AR1</b>				
Sample: 1 752				
Included observations: 752				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RIBEX	0,778	0,015	51,533	0,000
AR(1)	0,104	0,036	2,865	0,004
R-squared	0,777	Mean dependent var		0,000
Adjusted R-squared	0,777	S.D. dependent var		0,017
S.E. of regression	0,008	Akaike info criterion		-6,765
Sum squared resid	0,050	Schwarz criterion		-6,753
Log likelihood	2542,341	Durbin-Watson stat		1,999

Realizadas las correcciones oportunas para que el modelo estimado sea congruente con las hipótesis de partida se pueden interpretar los resultados de la tabla 6. La inferencia se basa en la distribución previamente establecida para el término de error<sup>3</sup>. Con un 95% de nivel de confianza el valor de la distribución t de la tabla<sup>4</sup> se utiliza para rechazar la hipótesis nula de que el parámetro sea cero; se rechaza si supera el valor en tablas en función del número de grados de libertad, y en ese caso se dice que la variable es significativa. De forma aproximada si el parámetro es mayor que dos la variable es significativa.

La inferencia estadística es utilizada para establecer regiones de confianza y determinar si las variables estimadas son o no significativas. En otras disciplinas científicas sería extraño que una investigación se presentase sin tomar las precauciones debidas. Por ejemplo, si se pretende explicar en qué medida un medicamento contra el SIDA aumenta la esperanza de vida el investigador se cuidará de dar el parámetro de la estimación acompañado de su intervalo de confianza. Además de la responsabilidad que aparece el buen hacer científico, hay que pensar en que una mala presentación de los resultados podría tener consecuencias sobre la reputación del científico o sobre su futuro en la profesión.

<sup>3</sup> El gráfico resultante de las funciones ACF y PACF confirma la ausencia de autocorrelación.

<sup>4</sup> El cociente entre el valor del parámetro y su desviación típica sigue una distribución t-student.

En economía desgraciadamente es habitual que se comente el valor del parámetro sin más, llegando en ocasiones a utilizarse como una realidad “universal” que permite influir sobre políticas, obviando el hecho de que los estimadores son variables aleatorias. En el mundo económico no debería ser lo mismo que a un inversor le dijeran que va a obtener una rentabilidad del 6% entre el 1% y el 11% o un 6% entre -3% y 9%, como tampoco debería ser igual que un crecimiento de la producción estimada del 2% se sitúe en un rango, entre un 2 y un 6% que entre un 1% y 3%. En la práctica en muchas ocasiones las investigaciones comentan el valor de los parámetros como si fuesen demostraciones de que en la realidad se producen los fenómenos tal cómo se implementan en el modelo, cómo si la teoría pudiese ser refrendada científicamente por la econometría, y además cómo si los estimadores fueran constantes, y cuyo valor coincidiera exactamente con el valor verdadero desconocido.

En esta misma línea de tomar los datos medios que recogen los parámetros como un indicador que no procede de un proceso aleatorio, también es muy frecuente que se interpreten los parámetros como elasticidades, por ejemplo en modelos de crecimiento o en modelos de estimación de demanda; la elasticidad se obtiene despejando de una ecuación determinística y el parámetro estimado ya se ha explicado que depende de las perturbaciones aleatorias.

## **5. Otros problemas de las estimaciones econométricas**

Además de la autocorrelación hay dos problemas relevantes que suelen estar presentes cuando se pretende analizar econométricamente la realidad económica. El primero es el de la multicolinealidad, que se deriva de la naturaleza social de los datos, en un escenario en el que hay más de una variable explicativa del mismo fenómeno y en el que resulta inevitable que una variable guarde relación con otras muchas. Así por ejemplo, un fenómeno como el abandono escolar depende en gran medida del contexto social en el que se mueven los alumnos, que puede ser recogido por el nivel socioeconómico familiar, pero a su vez dicho nivel guarda relación con el nivel educativo de los padres, con las posibilidades de refuerzo en casa y con el espacio o con los materiales que pueden ayudar al alumno a mejorar su resultado. En este terreno no es extraño encontrar estimaciones que pretenden incluir todas las variables disponibles en el análisis, lo cual conduce a una tabla de resultados donde la mayoría de las variables no son significativas. Una cuestión es presentar las pruebas y otra cuestión es que el modelo final intente aislar las variables que hagan que el modelo esté mejor ajustado. Para eso también sería recomendable una selección previa de variables basada en un sencillo análisis de sus correlaciones.

Si las variables independientes están altamente correlacionadas, lo único que podría decirse de esa estimación es que una combinación de los parámetros estimados resulta explicativa de la variable dependiente, sin pretender aislar el efecto individual con gran precisión, puesto que la introducción de una muestra mayor o la sustitución o eliminación de alguna de las variables ya existentes puede hacer cambiar de forma notable el resultado de los parámetros de las que queden en el modelo finalmente.

El segundo problema es el de la heterocedasticidad o varianza no constante en los residuos estimados. Este segundo problema tiene solución en el ámbito de las series temporales a través de la estimación de modelos autorregresivos condicionalmente heterocedásticos (ARCH, Engle (1982) o mediante la estimación robusta de heterocedaticidad de White (1980) en otros supuestos, entre otros métodos.

Otro problema es el de las variables atípicas, pues es frecuente que en los datos aparezcan datos extremos que pueden afectar de manera importante a la estimación. El tratamiento de las variables atípicas y su inclusión como dummies en la estimación es un proceso que debe realizarse de manera adicional al presentado en los dos últimos apartados.

## **6. ¿Lo complejo antes que lo robusto?**

Spanos (2006) revisa en su libro *Econometrics in Retrospect and Prospect* la historia de la econometría. Durante la última etapa del siglo diecinueve y principios del siglo veinte existió un cierto consenso sobre el método hipotético deductivo, mediante el cual sobre unos postulados o premisas iniciales se procede a derivar “leyes económicas”. Gran parte del debate posterior se dirigió a analizar cómo la estadística podía ayudar al proceso inductivo y convertir a la ciencia económica en una ciencia empírica. En esa línea Moore (1908) señaló que “la economía podía convertirse en una ciencia empírica en la que su componente deductiva estuviese suplementada con una adecuada componente inductiva basada en la estadística”, Jevons (1871) era igualmente defensor de la estadística como método de verificación de la economía deductiva. En la actualidad se ha generalizado el uso de las herramientas estadísticas para modelizar cualquier tipo de datos, bien sean series temporales, sección cruzada o datos de panel y los desarrollos de la econometría durante el siglo veinte han ido dirigidos a una mayor sofisticación para estimar modelos teóricos, no a “aprender de los datos sobre los fenómenos económicos observables”.

Anteriormente Leontief realizó una revisión crítica de los avances econométricos

en 1948 y después en 1971. En el periodo que media entre ambos trabajos el autor señala que su optimismo inicial sobre la posibilidades de la econometría como herramienta de análisis y soporte de las teorías económicas se evaporó por “la debilidad y el lento crecimiento de los fundamentos, que claramente no soportan la superestructura proliferante, o podría decirse, la pura teoría económica especulativa (...) La validez de esas herramientas estadísticas depende de la aceptación de presupuestos sobre las propiedades estocásticas del fenómeno que los modelos particulares intentan explicar, presupuestos que rara vez puede ser verificados”. Señala el autor que “en ningún otro campo de la investigación empírica ha sido tan masiva y sofisticada y la maquinaria estadística se ha utilizado con resultados tan indiferentes; sin embargo, los teóricos siguen ensayando modelo tras modelo y los estadísticos matemáticos continúan elaborando procedimientos complicados, uno detrás otro”.

Spanos (2006) señala las limitaciones de la estadística y la econometría como método inductivo treinta años después de lo que analizó Leontief y concluye que la situación es aún peor que antes por la rápida acumulación de indicadores económicos, el uso de paquetes estadísticos y ordenadores personales “que reduce el coste de producir la evidencia empírica”, la econometría aplicada consiste en “pruebas de estudio específico o de periodos concretos que son muy poco confiables y por tanto muy inadecuadas como fundamento para la economía”. (...) “De la misma manera que los teóricos de la economía no sienten la obligación de acudir a la evidencia empírica para desarrollar modelos, motivados por el rigor y la sofisticación matemática o la generalidad, los económetras continúan desarrollando nuevas y sofisticadas técnicas independientemente de que sean apropiadas para los datos económicos”. Spanos señala adicionalmente que la aproximación empírica a la modelización basada en enfoques dominados por teorías económicas suelen fallar en tres aspectos: a) La enorme separación entre la teoría y los datos, b) La estructura probabilística de los datos y c) las diferentes vías en que la inferencia podría estar equivocada. Por ese motivo se tienden a simplificar los presupuestos del modelo, porque cuanto menos restrictivos sean menos susceptible será a errores de especificación.

**Tabla 6:**  
**Relación de autores que intervienen en el desarrollo de la Econometría**

AUTORES	APORTACIÓN
Laplace, Legendre, Gauss (finales XVIII-primer mitad XIX)	Regresión lineal y MCO
Quetelet y Jevons (S. XIX)	Economistas empíricos
Galton, Gosset, Pearson, Fischer (Primera mitad s. XX)	Inferencia estadística
Haavelmo, Koopmans, Klein, (1930-1950)	Modelos de ecuaciones simultaneas
Anderson, Box, Jenkins (1960, 1970)	Series temporales
Amemiya, McFadden, Maddala (1970)	Variable dependiente limitada, Elección probabilística
Baltagi, Maddala, Hsiao, Arellano	Datos de panel
Zellner (años 60)	Econometría bayesiana
Dickey, Fuller, Phillips, Perron, Stock	Raíces unitarias
Granger, Engle	Cointegración
Engle, Bollerslev	Modelos ARCH, GARCH

Un análisis de la historia del desarrollo de los modelos econométricos es ilustrativo del grado de complejidad que han ido adquiriendo los mismos (véase la tabla 6). Como han señalado los autores citados la complejidad en si mismo del instrumento no tiene por qué implicar un mayor carácter científico. Más aún, lo no científico sería que el criterio de publicación en revistas científicas fuese “estar a la moda” o “a la última”. Quizás implicase una mayor aportación una estimación sencilla pero bien especificada y con un buen análisis del trasfondo económico, con un conocimiento de los procesos más allá de lo que hay en las bases de datos.

### **7. Los modelos y las “teorías económicas”**

Por último, una vez hechas las correcciones para evitar los problemas de especificación del modelo, hay que poner el acento en el sentido económico de los resultados: es decir si se está analizando causalidad o correlación entre las variables observadas. Intentar establecer una relación de causalidad es justo el tema en el que confluye la necesidad de una teoría o proceso que explique la realidad y la conexión

que se establece entre los datos.

La revisión crítica del apartado anterior ha dejado constancia en palabras de expertos econométricos la incapacidad de la econometría para dar sustento a teorías universales. Los modelos aunque estén correctamente especificados capturan relaciones positivas o negativas entre variables e inferir si son estadísticamente significativas, pero siempre en el contexto de la muestra de años o de la muestra de datos. Dos o más teorías explicativas de un fenómeno como por ejemplo del crecimiento económico pueden convivir en este estado del arte. También puede darse el caso que dentro de una misma región y con el mismo periodo de datos diferentes investigadores encuentren diferentes modelos de análisis de la realidad.

En el ámbito del crecimiento económico hay una parte de los investigadores que están publicando en la actualidad sobre la base del modelo de Solow y sus ulteriores desarrollos, mientras que otros investigadores utilizan modelos orientados a la demanda de corte postkeynesiano.

Stiglitz (2010) señala al hilo de la crisis económica actual la necesidad de reforma de la actual ciencia económica, cuya corriente de corte neoclásico principal sigue siendo dominante y basándose en un postulado que no resulta compatible con la generación de burbujas y la existencia de crisis económicas, los mercados perfectamente competitivos. Si ese es el debate que hay sobre la mesa no se entiende que las revistas económicas sigan aceptando modelos que se estiman sobre la base de que existe perfecta información y competencia. Una revisión de los análisis de revistas científicas serviría para refrendar que para una parte importante de los investigadores la crisis sencillamente no les afecta para nada. Eso refrendaría como mínimo la posibilidad de otros economistas críticos de avanzar en otra línea de pensamiento. Lo que es realmente paradójico es que se utilicen modelos de equilibrio general competitivo en estudios que avalan las políticas de recortes sociales o de reformas en el sistema de pensiones impulsadas por la crisis económica pero no se haga ninguna crítica a los postulados del modelo.

## **8. Conclusiones**

A lo largo del capítulo se ha intentado abordar el problema que representa el uso de la econometría como método de investigación en economía. Sería recomendable un replanteamiento del uso de la econometría como instrumento de análisis de la realidad económica que implique un análisis que ponga de manifiesto las limitaciones de los presupuestos, que intente corregir los problemas de autocorrelación y he-

terocedasticidad, que simplifique en la medida de lo posible el número de variables para evitar problemas de multicolinealidad. En definitiva esto implica reconocer la debilidad de la herramienta y señalar hasta donde puede ayudar a entender mejor la realidad.

La complejidad no puede ser en si misma sinónimo de refinamiento ni necesariamente ha de ser mejor herramienta para el científico. La estadística en general, y la econometría en particular, no permiten ratificar leyes universales, porque es justo lo que no es posible hacer en las ciencias sociales. La econometría es una herramienta correcta como método de análisis de la realidad, más “fino” que la mera observación de gráficos y tablas. Por todos los argumentos expuestos a lo largo del capítulo, podemos concluir que la econometría es un instrumento de análisis que permite capturar correlaciones entre variables pero no tiene capacidad resolutive para determinar su conexión causal. El terreno de las teorías y las escuelas de pensamiento sigue en ese sentido completamente abierto en el mundo económico.

## BIBLIOGRAFÍA

- Box, G.E.P y Jenkins, G. M. (1976) *Time Series Analysis forecasting and control*, Holden-Day, Oakland, California.
- Dickie, D. A. y Fuller, W. A. (1979): “Distribution of the estimators for autorregressive time series with a unit root” *Journal of the American Statistical Association*, 74 427-431.
- Durbin, J. y Watson, G. S. (1950): “Testing for serial correlation in least squares regression”, *Biométrica*, 37, 409-428.
- Engle R F (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of. United Kingdom inflation. *Econometrica* 50:987-1008, 1982.
- Engle, R.F. and C.W.J. Granger (1987) Co-integration and Error-correction: representation, estimation and testing. *Econometrica* 55, 251-76.
- Granger, C.W.J. y Knewbold, P. (1974): “Spurious regressions in econometrics”, *Journal of Econometrics*, 26, 1045-1066.
- Haavelmo, T. (1943) The statistical implications of a system of simultaneous equations. *Econometrica* 11, 1-12.
- Haavelmo, T. (1944) The probability approach to econometrics. *Econometrica* 12, suppl., 1-115.
- Jevons, W.S. (1871) *The Theory of Political Economy*. London: Macmillan.
- Jevons, W.S. (1874) “The Principles of Science”. London: Macmillan. 54 *Econometrics in Retrospect and Prospect*.
- Moore, H.L. (1908) The statistical complement of pure economics. *Quarterly Journal of Economics* 23, 1-33.
- Morgenstern, O. (1970): *Sobre la exactitud de las observaciones económicas*, Tecnos, Madrid.

- Phillips, P.C.B. (1987): "Time series regression with a unit root". *Econometrica*, 55, 207-301.
- Ruud, P.A. (2000): *An Introduction to Classical Econometric Theory*, Oxford University Press, New York.
- Spanos, Aris. (2006). Econometrics in retrospect and prospect. In *New Palgrave handbook of econometrics, vol. 1*, eds. T. C. Mills, and K. Patterson. London: MacMillan, 3-58.
- Stiglitz, J. (2010): *Caída Libre*. Ed. Taurus.
- Theil, H. (1959): *Agregación lineal de relaciones económicas*, Aguilar, Madrid.