Distributed Lags using Elastic-Net regularization for Market Response Models: focus on predictive and explanatory capacity

Journal name: Journal of Marketing Analytics

Andrés Martínez[a,b]*, Alfonso Salafranca[c], Ana E. Sipols[d], Clara Simon de Blas[d], Daniel van Hengel[a]

[a]MDML, Trompenburg 1, 1181 NM Amstelveen, The Netherlands; [b]Coolblue BV, Weena 664, 3012 CN Rotterdam, The Netherlands; [c] Havas Media Group, Eloy Gonzalo 10, 28010 Madrid, Spain; [d]Universidad Rey Juan Carlos, Av. Tulipán, 28933 Mostoles, Spain

E-mail of the correspondence author:

Andrés Martínez: jandres.martinez@protonmail.com

Distributed Lags using Elastic-Net regularization for Market Response Models: focus on predictive and explanatory capacity

Abstract

For many decades, considerable research has been conducted on Market Response models. Mostly without any attempts to validate the results in strictly predictive tasks and often ignoring if the methods comply with the underlying assumptions and conditions, like the method's ability to outline the broadly accepted effects of advertising actions. This work presents an enhanced method for market response models consistent with the underlying assumptions of such. Our method is based on Distributed Lag Models with the novelty of introducing regularization in its estimation, a cross-validation framework, and hold-out testing, next to present an empirical manner of extracting its effects. This approach allows the construction of models in an exploratory and simple manner, unlocking the possibility of extracting the underlying effects and being suitable for large samples and many variables. Last, we conduct a practical example using real-world data, accompanied by an unprecedented set of empirical explainability assessments next to a high level of predictive capability in similar circumstances to how it would be used for decision-making in a corporate setup.

Key words: Distributed Lag Model, advertising, lagged effects, machine learning, prediction

1. Introduction

Since the 1960s, researchers have focused intensively on finding answers to identify and quantify success factors and explain incremental sales volumes through model-based (or model-dependent) approaches (Borden, 1964; Sethi, 1977; McCarthy, 1978; Tellis, 1988). Despite the consensus on the general theoretical dynamics expected from the primary factors driving sales-marketing effectiveness (Vakratsas & Ambler, 1999; Tellis, 2006), these dynamics are often ignored in the studies. Moreover, there is no agreement on how best to address this problem in practical terms, as little or no effort has been put into testing its usefulness beyond the model's in-sample fitting and incredibly neglecting the predictive capacity.

Marketing science and practice are going through an analytics disruption boosted by the increasing availability of data, the strong presence of digital marketing, the important role of social media channels, and the continuous flow of new tools and methods arising for marketing analytics (Moorman, 2016; Verhoef et al., 2016; Iacobucci et al., 2019). However, there are still many questions that need to be answered and issues that should be clarified regarding marketing analytics, especially considering their widespread use and fast-paced development (Iacobucci et al., 2019). A large number of review articles underline the state of marketing research related to marketing analytics, regarding the need for more integrative empirical studies in the primary marketing areas, as well as the formulation of comprehensive theoretical models (Iacobucci et al., 2019). Nevertheless, marketers still need to be equipped to extract underlying insights from their data to measure, track, understand, and interpret the marketplace (Berger et al., 2019). As Midgley et al. (2017) indicated, no one theory or set of theories justifies the use of a particular method that can be comprehensive, especially when dealing with the complex nature of Marketing related phenomena and the increasingly large quantity of data. Therefore, some suggest embracing a broader diversity of methods and analytics to tackle the challenge of equipping marketers with effective and practical tools (Petrescu & Krishen, 2019).

In marketing, 'effectiveness' refers to describing how advertising actions drive brand awareness and sales. Marketing Mix Modeling (MMM) (Borden, 1964) or Market Response Models are methods to quantify the effect of sales-marketing efforts and ultimately forecast outcomes of different configurations. The philosophy underlying model-based approaches is that historical data contains valuable information to enhance our understanding of the studied phenomenon. Besides, embedding these dynamics in a mathematical formula results in a toolbox for predicting how consumers might respond in the future and how to best plan marketing variables (Tellis & Zufryden, 1995). It should be noted that MMMs are regression based on a limited amount of aggregated observational data, and such models are typically constructed based on correlational measures. As such, causality can not be inferred if no further analyses are conducted, like an evaluation on purely predictive tasks or certain conditions are met. Some works suggest that Bayesian methods offer some improvement on this subject (Chan & Perry, 2017), but these imply that researchers possess prior knowledge (1) about what the actual drivers are and (2) the link between these and the studied phenomenon. Paradoxically both points constitute part of the problem and are (very) prone to change over time (e.g., due to competitors' actions, qualitative factors, etc.). Thus, depending on the circumstances, an exploratory frequentist approach is more appropriate than the optimistic Bayesian methods regarding the prior knowledge that scholars or practitioners possess of the phenomenon under study.

In this work, we propose a methodology for marketing response consistent with the underlying assumptions and conditions, from an exploratory perspective using Distributed Lag Models (DLagMs) (see Demirhan, (2020) for more details) in a regularized framework called

Elastic-Net (Zou & Hastie, 2005). A key difference between previous works using DLagsMs is, on the one hand, the usage of regularization in the estimation phase, on the other, a cross-validation strategy and hold-out testing. These additions yield several relevant benefits. First and foremost, we overcome the shortcomings that the traditional estimation of DLagsMs has: the problem of estimating in the presence of considerable multicollinearity between lagged and current values of the same variables and the identification of the number of lagged variables needed. Furthermore, regularization facilitates the search for a parsimonious model, which is achieved by finding the most predictive features among a large number of variables that potentially drive the phenomenon within the estimation process. The process of selecting the features relies on utilizing regularization on a cross-validation framework. Consequently, not much prior knowledge of the true link between the target and input variables is needed, so avoiding the traditional back-and-forth manual work in selecting what features will take part in the model. We put particular emphasis on extracting the commonly discussed patterns of advertising response, which are broadly discussed in the literature but rarely extracted. And last but not least, we assess the results not only on how close they are to the accepted assumptions but also by testing on predictive tasks.

To demonstrate the benefits, we conduct a practical example using a real-world dataset consisting of measurement of Brand Awareness (AW) collected weekly for six years and commonly used metrics to represent the advertisement exposure in television media. Our methodological contribution is the combination of the following elements: (1) using the Elastic-Net regularization for estimating the parameters of the DLagM, easing the model definition greatly and estimation phases; (2) a model-agnostic method is proposed to empirically extract the effects of the relevant drivers, especially relevant when the model complexity is high and (3) an evaluation approach with a focus on the predictive and explanatory ability rather than solely discussing the goodness of fit and theoretical assumptions. Some may see the results of our real-world example as not the best representation of the current reality because the dataset does not contain online media channels. However, the methodological framework presented allows future investigations in the Marketing field utilizing more recent records samples. Similarly, it is also applicable to other domains.

2. About the Marketing Analytics field

Many proposals for aggregated marketing responses (brand awareness or sales volumes) are found in the literature. While the objective is the same, there are differences in terms of assumptions and underlying conditions across the methods, like whether or not the method can capture all the effects of the advertising, ultimately impacting the findings and practical usability. Regrettably, it is difficult nowadays to identify the adequate method as little or no effort has been put into testing its usefulness beyond the model's in-sample fitting and sometimes analyzing how close the results match the expected dynamics. Often these dynamics are forced upfront by setting constraints and performing separated estimations to fix intermediate outcomes, overly complicating the process and, most importantly, changing the intended exploratory purpose of the work to mainly confirmatory. Some studies claim that the current marketing modeling literature produces many new proposals oblivious to the problems associated with their causal inferences and without any attempts to validate the results in strictly predictive tasks. Thereby, the value of such proposals to aid decision-making is diminished (Ehrenberg et al., 2000; Chan & Perry, 2017).

A model is a formal quantitative representation of the theories and hypotheses of how a specific phenomenon arises, aiming to offer a certain level of explanation of this complex system (Lauenroth, 2003). Models are built to approximate these processes and apply inferential statistical methods to test the relationships between the assumed causes and effects (Sarstedt & Danks, 2022). When it comes to validation, and especially in the Marketing field, the rule is that researchers' primary focus is on (a) assessing whether the model fits the data and (b) whether the model coefficients are significant and in the direction expected based on their hypotheses. Less often, analyses are conducted on whether or not the underlying dynamics depicted by the model are in line with the theory, but very rarely, the evaluation phase includes testing the model's predictive ability. Sarstedt & Danks (2022) show that a model with a certain degree of explanatory power can produce vastly different levels of predictive power and vice versa. Surprisingly, still, many works are prolific in deriving practical recommendations, which inherently result from a predictive scenario despite that they were never proven (Sarstedt & Danks, 2022).

Predictive assessment means applying the estimated model generated from a sample at hand to make predictions on other observations not used during estimation. These observations must be (strictly!) kept separate from the primary sample utilized for the model estimation, or they can be collected at a future time or even in another context (Shmueli et al., 2016). In this work, we state that there are very few practical benefits of conducting assessments on the model explainability and how aligned the results are with our hypotheses and broadly accepted assumptions if we do not conduct checks on the predictive model capacity.

With the advent of the internet, traditional broadcast communication channels such as Television, Radio, and Print are no longer the dominant sources of information for consumers, as they have drifted towards social media channels and/or virtual communities for information exchange and relationship-building (Hair et al., 2010). As a consequence, many businesses have shifted their advertising expenditure toward digital media, though multiple studies show that traditional media remain effective (Danaher et al., 2013; Danaher, 2021). The availability of more granular data records collected from online media channels has pushed forward a new trend focusing more on short-term action performance. This tendency has triggered some concerns regarding the conclusions about the effectiveness of Marketing actions, as the growing use of short-term metrics and the development of campaigns aligned to them are extremely damaging developments in this domain[1]. Without entering into this debate, it seems not complicated to develop a set of plausible arguments to justify the presence of short-term and mid/long-term components in the realm of advertising. Thus, it seems reasonable to say that ignoring longer-term effects will negatively impact the validity of the results, implying that approaches that neglect the importance of such dynamics are not the best fit for marketing response.

Patterns of advertising response

Seven important patterns are discussed in the literature: current, carryover (also known as Advertising's adstock), shape, competitive, dynamic, content, and media effects (see Tellis, (2006) for more details). The current effect of advertising is the change in sales occurring at the same time when a batch of messages (pulses) is broadcasted. The carryover effect is the portion of the effect that occurs during the time frame after a pulse of advertising. Shape denotes the effect of changes on the target variable (e.g., sales volume) in response to the increasing intensity of advertising. In a competitive market, advertising effectiveness tends to be reduced by the actions of other brands' advertising, and this dynamic is known as the competitive effect. Dynamic effects are those that impact not only contemporaneously but also in posterior time frames; among these are the *carryover* effects and the so-called *wear-in-wear-out*. Content effects are the response to qualitative changes in the message. Last, media effects are the differences in advertising response

[1] Binet & Field, (2017). Media in Focus: Marketing Effectiveness in the Digital Era

due to the media channel used. Ideally, an advertising model should be capable of capturing these seven effects, though only a few have come close (Tellis, 2006). As far as our knowledge goes, DLagM is considered one of the few methods capable of depicting the above effects and will be shown empirically in later sections of this work.

When assessing the goodness of a model, typically, the following two aspects are important: (a) accuracy of prediction on future data, as it is difficult to defend a model that predicts poorly; (b) interpretation of the model (Zou & Hastie, 2005). In the context of market response models, we will extend point (b) with the model under assessment and should be able to capture the patterns of advertising response.

3. Background

Some examples of works addressing sales-marketing responses are based on Bayesian regressions (Brown, 1986; Bass et al., 2007), Bayesian hierarchical methods, multivariate linear regressions (Havlena & Graham, 2004), the Kalman Filtering as is the case of Naik et al. (1998), Support Vector Machines (SVM) (Viaene et al., 2001) Artificial Neural Networks (ANNs) (Zhang et al., 2009; Guido et al., 2011), nonlinear dynamic time series (Huffaker & Fearne, 2019), dynamic linear model combining multiple submodels for accounting for underlying patterns of advertising (Bruce et al., 2012), Distributed Lag Models (Bass & Clarke, 1972; Clarke, 1976; Weinberg & Weiss, 1982; Rufino, 2008; Mulchandani et al., 2019). Regrettably, it is the rule that these studies solely address in-sample evaluation and, in less often, discuss theoretical assumptions, so they fail to provide sufficient evidence of the validity of their suggestions to practitioners for applying such methods in real-life practice.

Making an effort to validate results on out-of-sample records Rubel & Naik (2017) propose the Robust Dynamic Estimation, and Rutz et al. (2011) suggest a Bayesian version of the Elastic-Net (Li & Lin 2010, Kyung et al. 2010) for modeling indirect effects of paid search advertising. Although there is some proof of validity in these studies for practical purposes, the former requires multiple stages to conduct the estimation, complicating its use by practitioners and increasing the computational cost significantly for large data sets. And the latter work depends on assumptions over some parameter distributions, and yet, what if the correct distribution is different? One might argue that it is unreasonable to expect researchers or practitioners to possess sufficient knowledge and time to estimate parameters for each alternative distribution, especially in a corporate setting.

Most studies use aggregated data primarily in a time series form, though some researchers report that studying ad effectiveness on an individual level provides superior results (Sethuraman et al. 2011). Examining the relative importance of each medium for multiple retailer brands within a product category at the customer level is found in Danaher et al. (2020), where the Tobit model is applied. In their work, some relevant findings in the context of Multimedia and Multichannel environments are found, using, among other features, click-throughs (i.e., touchpoints). Unfortunately, analysis on the customer level with features from the click-stream is not without problems. An important one is that data quality from tracking browsing paths depends on the users' Internet browser privacy settings (e.g., they can block cookie tracking, delete cookies, etc.), complicating the task of extracting correct information. For example, in Danaher et al. (2020), only 64.50% of the records are considered reliable regarding their purchases, along with those who made their first recorded purchase partway through the observed period, potentially biasing the data sample. Besides, recent new regulations to protect users' privacy are drastically changing how cookies are tracked, which is a key element for these approaches. So future works will have to accommodate the new records available.

Much of the recent Market response research focuses primarily on short-term incremental volumes and, to a lesser extent, on the mid and longer-term perspective, like the brand-building effect. While the former is indeed key for short-term decision making, the latter constitutes a crucial part in completing the evaluation of the Return on Investment (ROI), as more strategic budget allocation is a critical success factor in creating equity (Kitchen, 2010), and it fosters consumer brand loyalty and enhances profitability (Tsan-Ming, 2014). Some studies addressing the effects that marketing actions have on the brand, suggest the application of Partial Least Squares (PLS) to explore the mechanisms that online social media channels or virtual communities have on Brand Credibility (Chen & Shupei, 2019). A closely related method to PLS is shown in Bilgin (2018), where Structural Equation Modeling (SEM) is suggested. However, these approaches ignore the importance of the time component nature of the studied phenomenon, as their proposals rely on conducting their analysis over a relatively small sample collected through a questionnaire within an arbitrary time frame.

4. Distributed Lag Models and problem description

In this section the regression class known as Distributed lag models (DLagMs) is briefly discussed and the problem of estimating a time-ordered phenomenon that may be partly or fully driven by distributed lag mechanisms is formalized. Assuming the absence, or minimum, prior knowledge about the true link between potential drivers and the target, the aim is to estimate the underlying mechanisms, including current and carryover effects in the short and mid/long-run.

DLagMs are a class of regression that considers lagged effects in explanatory variables and an autoregressive-like component (Demirhan, 2020), enabling the depiction of certain nonlinear relationships and unlocking the possibility of extracting the expected effects of advertising which show to be consistent with the evidence (Bass & Clarke, 1972). Time series models often involve some notion of distributed lag, and applications of DLagMs estimated as traditionally (i.e., without regularization), are found in a wide range of fields like energy (Csereklyei et al. 2019), agriculture (Berk, 2017; Özsayin, 2017), economics (Belloumi, 2014; Nerudova & Dobranschi, 2019), environmental and medicine related domains (Mohammed et al, 2019; Heaton et al. 2019; Nothdurft & Engel, 2019), and sales-marketing related studies (Bass & Clarke, 1972; Clarke, 1976; Weinberg & Weiss, 1982; Rufino, 2008; Mulchandani et al, 2019).

Let us consider two variables denoted as $□_□$ and $□_□$ with $□ = \{1, 2, \ldots, □\}$ and assume that the relationship between these variables is such as

(1) $$□_□ = □ + □(□_□ + □□_{□-1} + □^2□_{□-2} + \ldots) + □_□ \;□□□h\; |□| < 1$$

The equation (1) indicates that □ is a function of current and past values of □ plus an uncorrelated noise □, where the lag coefficients have a geometrically decaying pattern. Thus, current effect of variable □ in □ is represented by β whereas the total carryover effect is equal to $\frac{\beta}{1-□}$. As this expression contains an infinite number of lagged variables, the so-called Koyck transformation (Koyck, 1954) is often considered which implies subtracting $\theta □_{□-1}$ from (1) to get

(2) $$□_□ = □^* + □□_{□-1} + □□_□ + □^*_□ \;□□□h\; |□| < 1$$

With $\alpha^* = □(1 - □)$ and $□^*_□ = □_□ - □□_{□-1}$. For simplicity and without loss of generality, from now on we refer to these values as α and □. In the realm of time series, the equation (2) is also called the ARMAX model (see Hannan et al. (1980)). The autoregressive part is represented by lagged values of the target (i.e. $\square_{\square-1}$), the moving average part concerns to $\square_{\square-1}$ and the exogenous explanatory variables relies on $\square_{\square}$.

Using the Koyck model framework, a target or dependent variable $\square$ can be estimated based on a set of dependent variables $\square = \{\square_{1,}\square_{2},\ldots,\square_{\square}\}$ as it is shown in equation (3), and it can be seen as an augmentation of the traditional linear model.

$$\widehat{\square_{\square}} \simeq \square + \square_{1}\square_{\square-1} + \sum_{\square=1}^{\square} \blacksquare\, \square_{\square}\square_{\square,\square}\ \square\square\square h\ \square = \{1,2,..,\square\} \tag{3}$$

The higher the value of θ_1 with $0 \leq \square_1 < 1$, the larger the time frame where the effect on the target variable remains (i.e. carryover effect). A particular case is when $\theta_1 = 0$, indicating the absence of a carryover effect and so the estimation is equivalent to a traditional linear regression with solely current effects as drivers.

Although this model looks relatively simple, due to the autoregressive component in the form of a lagged value on the target variable, its mathematics can be quite complex and less straightforward to analyze than one might assume (Clarke, 1976; Franses & van Oest, 2004). The Koyck model should be considered within the context of two important limitations. On the one hand, it can depict only carryover effects with smooth decay and strictly decreasing, in other words, the decay could not have a hump-like form nor a non monotonic decay; on the other, the same carryover shape is assumed for all dependent variables in $\square$.

A general form of DLagMs comprises multiple lagged values in both the dependent variable and independent variables, as shown in equation (4).

$$\widehat{□_□} \simeq □ + \sum_{□=1}^{□} ■ □_□ □_{□-□} + \sum_{□=1}^{□} ■ \sum_{h=0}^{□_□} ■ □_{□,h} □_{□,□-h} \quad □□□h\ □ = \{1,2,..,□\} \qquad (4)$$

In equation (4), the target variable is estimated with an autoregressive-like form of order □, and lagged transformations of order $□_□$ are applied to explanatory variables $□_□$ for all $□ \in □$. This overcomes Koyck's model limitation as it can capture a whole range of carryover effects and allows depicting distinct carryover shapes, but two challenges are to be faced when using this model. These are, performing the estimation in the presence of considerable multicollinearity between lagged and current values of the same variables, and identifying the number of lagged variables needed.

4.1. Formal problem description

Let $□_□$ be a variable measuring a certain phenomenon collected at □ discrete equidistant time instances $□ = \{□_1, □_2, \ldots, □_□\}$, ordered chronologically such that $□_{□1} < □_{□2}$ for all $□_1 < □_2$. And let □ be a matrix with dimension $(□, □)$, containing □ observations collected in □ from □ variables considered as potential drivers of the phenomenon of interest.

The link between the target and input variables is initially unknown. So the model definition, meaning the features which should be included in the so-called design matrix X, constitutes part of the problem. In order to eliminate intermediate steps required for selecting and constructing features, which in a way is like constraining the underlying forces driving the phenomenon upfront; we set the problem as finding a solution that selects and estimates features at the same time, from a set of potential drivers. Last, but not least, the resulting estimation should be able to explain the phenomenon retrospectively, but equally important is its ability to make future inferences.

Given that the underlying dynamic driving the phenomenon is unknown, the model has to be general enough to capture a wide range of distinct effects' shapes. Thus, it seems the most suitable model is the general form of a DLagMs (equation (4)), and the problem under consideration can be stated as follows:

Problem 1 (Identifying and estimating the model). Given $\square$ observations ordered chronologically in equidistant time instances indexed by $\square \in \square = \{\square_1, \square_2, \ldots, \square_\square\}$, of a target variable $\square_\square$ and $\square$ potentially explanatory features comprised in the matrix $\square$, created from arbitrary lagged transformations of the target plus contemporaneous and arbitrary lagged values of the rest of the characteristics, estimate a model based on a general form of a Distributed Lag Model (DLagM), able to identify what inputs are relevant, outline the effects within a broad range of shapes, and ultimately capable of making reliable future inferences.

5. Our proposal

To address **Problem 1**, we will cast it as a supervised machine learning problem where the estimation is done by regularization in a cross-validation framework, which enables us to perform the features selection based on their predictive power. By doing so, it overcomes the two shortcomings of the traditional estimation of DLagsMs (see the previous section) and dramatically reduces the estimation time in large data samples. Regularization basically means adding a penalty in the loss function affecting the model's coefficients, resulting in shrinking the coefficients with little or no impact in terms of fitting improvement during the estimation process. Two famous regularization techniques are Ridge and the Lasso, proposed by Hoerl & Kennard (1988) and Tibshirani (1996), respectively. These methods are typically used for reducing the input's dimensionality in other fields and are especially suitable when this matrix is very sparse (i.e., with many zeros). Regarding predictive performance, there is no clear advantage between Ridge and Lasso (Tibshirani, 1996; Fu, 1998), but if the variable selection is the main focus, Lasso is more appealing because of its more minimalist representations (Zou & Hastie, 2005).

A more suitable regularization method for situations where the number of variables is not very large and in the presence of potentially highly correlated features, as is the case of the type of problems discussed here, is known as Elastic-Net (Zou & Hastie, 2005). This method can be seen as a generalization of the Lasso, and similarly to it, simultaneously does automatic variable selection and continuous shrinkage by mixing Lasso and Ridge penalties, and is capable of selecting groups of correlated variables. Elastic-Net achieves the same performance when the Lasso is at its best, while improving its results under the circumstances highlighted above. Elastic-net solves the problem shown in equation (5) over a grid of λ values.

$$\min_{\beta_0,\beta} \frac{1}{N} \sum_{i=1}^{N} w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda\left[\frac{(1-\alpha)\|\beta\|_2^2}{2} + \alpha\|\beta\|_1\right] \tag{5}$$

With □ being the number of observations in the sample used for the estimation, $\square(\square_{\square}, \square_0 + \square^{\square}\square_{\square})$ the negative log-likelihood contribution for observation □, □ refers to the weights applied to each observation, and α defines the combination between Lasso and Ridge applied on the loss function with $0 \leq \square \leq 1$. Finally, λ controls the overall strength of the penalty, which can be the same for all the coefficients or set differently per coefficient.

Resulting values β_0 and β will constitute the mathematical expression that performs best at representing the studied phenomenon. α and weights applied per observation $\square_{\square}$ are user parameters set upfront. The overall strength of the penalty λ can be set as one scalar (not recommended), as a sequence of values which is set either by the user, or in some implementations, selected over a sequence of values proposed. The final selection of a single λ is done by means of cross-validation (Stone, 1974; Allen, 1974) that is a technique very often used in the field of Machine Learning. The goal of cross-validation is to test the model's ability to make predictions over unseen records in the estimation process and reduce the so-called overfitting or selection bias (Cawley & Talbot, 2010).

With the above notions, the problem under consideration can be stated as follows:

Problem 2 (Reformulation as supervised learning problem). Estimate the values $\square_{\square}$ from the features vectors in $\square$ representing potential drivers of the phenomenon on current and future time instances (i.e. DLagM form), by means of Elastic-Net regularization and cross-validation.

For solving problem 2, we utilize an implementation for fitting the entire Elastic-Net regularization path with Gaussian response, available in a R-package called *Glmnet* (Friedman et al. 2010). This implementation enables the setting of bounds on the space where coefficients are searched and establishing distinct penalty factors on the parameters' shrinkage, allowing the user to modify the algorithm outcome if needed (further details in later sections). The algorithm uses cyclical coordinate descent in a path-wise fashion as described in Friedman et al. (2010).

5.1. Effects extraction

Given that all drivers are in an additive form, one might think that individual and aggregated effects could be extracted analytically in a fairly simple manner. Unfortunately, this will not be an easy task because the presence of an autoregressive component implies that the effect in a given time instance is calculated taking into consideration prior and contemporaneous effects. To illustrate this, let us consider the expression $0.4 + 0.5\square_{\square-1} + 0.2\square_{1,\square} + 0.1\square_{2,\square}$ that estimates variable $\square_{\square}$

within a range of time instances $\square \in \square = \{1, 2, \ldots, \square\}$. In this expression, effects can be computed as shown in Table 1.

Table 1: Simple example of effects extraction analytically

Time instance $\square$	Autoregressive *	Effect of var. $\square_1$	Effect of var. $\square_2$
1	0.4	$0.2\square_{1,1}$	$0.1\square_{2,1}$
2	0.6	$0.2\square_{1,2} + 0.1\square_{1,1}$	$0.1\square_{2,2} + 0.05\square_{2,1}$
...	...	...	...
n	$0.4\sum_{\square=1}^{\square} \blacksquare 0.5^{(\square-1)}$	$0.2\sum_{\square=1}^{\square} \blacksquare 0.5^{(\square-\square)}\square_{1,\square}$	$0.1\sum_{\square=1}^{\square} \blacksquare 0.5^{(\square-\square)}\square_{2,\square}$

* The autoregressive component includes the intercept and $\square_0$ and for simplicity is assumed to be equal to 0

It can be seen in Table 1 that in the presence of an autoregressive component, the effect of exogenous variables is a function of contemporaneous and autoregressive coefficients, whereby there is a lagged effect regardless of the existence of lagged transformations in exogenous variables. While in some cases effects can easily be expressed analytically as is the case of the example above, it becomes very complex when the autoregressive order is greater than 1 and lagged transformations exist in exogenous variables. To ease this task, we propose an approach that could be seen as inverse engineering the resulting model.

Let the estimation of a variable $\square_\square$ be based on $\square$ explanatory variables as $\square(\square, \square) = \hat{\square}_\square \simeq \square + \sum_{\square=1}^{\square} \blacksquare \square_\square \square_{\square-\square} + \sum_{\square=1}^{\square} \blacksquare \sum_{h=0}^{\square_\square} \blacksquare \square_{\square,h} \square_{\square,\square-h}$, with target and inputs indexed chronologically in $\square \in \square = \{1, 2, \ldots, \square\}$ and $\square + \sum_{\square=1}^{\square} \blacksquare (\square_\square + 1) = \square$. The idea is to recreate a configuration where effects can easily be calculated empirically, such that the extracted shape is a homothetic-like transformation of the original one. Afterwards, by converting them to percentages, original effects can be extracted from the estimation $\hat{\square}_\square$.

The process consists of computing the estimated values iteratively within the following configuration: (1) □ instances previous to □ $= 1$ are initialized to zero, (2) intercept is subtracted from the estimation in every iteration and (3) estimated values are used in the autoregressive component instead of original ones.

Within the above setting, the autoregressive component will depend only on the coefficients $\theta_{□}$ □□□h □ $= \{1, 2, \dots,$ □$\}$. And effects from the (□ $-$ □) exogenous variables (current and lagged) will depend on the coefficients $\theta_{□}$, and coefficients $\beta_{□,h}$ times the values of the exogenous variables. This procedure means that the original design matrix □ is transformed to a new matrix $□^{(*)}$where lagged values of the target are replaced with predicted values with the autoregressive component initially set to 0 and subtracting the intercept while the rest inputs remain unchanged. The outcome of this process is denoted as $\hat{□}_{□}^{(*)}$ and although it is a bad estimation of the original phenomenon in total terms (i.e. actual versus fitted values are not close), it serves as a simple way to isolate the driving forces in next steps. Within the same premises, $\hat{□}_{□}^{(*□)}$ can be computed from a matrix $□^{(*□)}$ constructed by setting columns in $□^{(*)}$ to zero referring to variables indexed by □. In this way, the proportion of the effect from variables indicated by indexes in □ can be calculated as $□□□□□□□^{(□)} = (\hat{□}_{□}^{(*)} - \hat{□}_{□}^{(*□)}) / \hat{□}_{□}^{(*)}$. Finally, the effect is transformed to fit the original values as $□□□□□□^{(□)} = \hat{□}_{□} \times □□□□□□□^{(□)}$ where $\hat{□}_{□}$ are the values estimated by the original input matrix □. It is to be noted that the effect of the intercept is distributed proportionally among all drivers' effects. The process is shown in Table 2.

Table 2: Process to empirically extract the effects regardless of the DLagM's complexity

Sequence of steps in the procedure

Inputs

Estimated model:

$$f(Y,X) = \alpha + \sum_{j=1}^{p} \theta_j y_{t-j} + \sum_{i=1}^{m} \sum_{h=0}^{q_i} \beta_{i,h} x_{i,t-h}$$

Original variables:

$$Y = [y_{t-1}, y_{t-2}, \dots, y_{t-p}]$$

$X = [x_1, x_2, \dots, x_b]$ with $b = (k - p)$

$n = N$(total number of observations)

Original estimation:

Yestim = f(Y, X)

User parameter of indexes denoting the variables to extract the effect

$$v = (1, 2)$$

Initialization

Y1 = vector of zeros of length n+p

Y2 = vector of zeros of length n+p

Loop

for i in 1 to n-p

X1_i = X[i,]

X2_i = X[i,]

X2_i[,v] = 0

Y1[i+p,]= max(f(Y1[i:(i+p-1),], X1_i) - α, 0)

Y2[i+p,]= max(f(Y2[i:(i+p-1),], X1_i) - α, 0)

end

Output

$peffect^{(v)}$= (Y1[(p+1):n] - Y2[(p+1):n]) / Y1[(p+1):n]

$effect^{(v)}$= Yestim × $peffect^{(v)}$

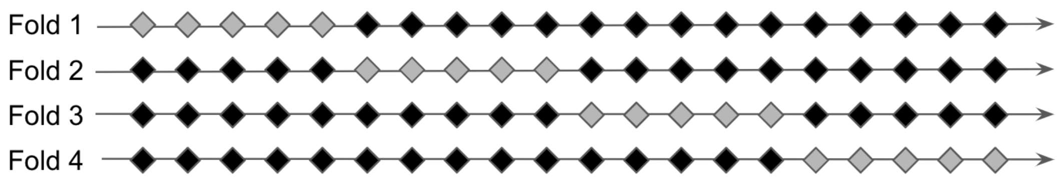
5.2. Model training, evaluation and testing

Instead of examining parameters' statistical significance based on theoretical assumptions on how the underlying dynamic should be, we focus on the predictive ability for outlining effects driving the phenomenon. The following points are in scope of our assessment: (a) variability explained historically measured by the R^2 metric, (b) the variability explained for predictive tasks, and (c) the procedure's ability to extract true effects over a simulated data, where the underlying dynamic of the phenomenon is known and will serve as 'ground truth'.

When results are assessed by analyzing the model fitting on records used in the estimation, we refer to it as an in-sample evaluation (point (a)). This is the case when the mathematical expression is applied to interpolate values within the time range selected for constructing the model. A more appropriate assessment of the model's predictive capacity is when the assessment is conducted on records strictly from future time instances, such that those records were never used for constructing the estimation (point (b)), and if using a moving window it is known as walk-forward testing.

The so-called overfitting (i.e when in-sample fitting is good, but much worse on new records), is minimized using a cross-validation process known as K-fold (Hastie et al. 2009). K-fold cross-validation consists of dividing the sample into K disjoint subsets (folds), then training the model K times with all folds but one each time. In each iteration, the subset not included in the training is used for evaluation, and the final outcome is extracted by averaging these. To avoid an overoptimistic evaluation when records are in the form of a time series, instead of splitting the records among folds randomly, the selection will be done in a way that they are divided in consecutive time instances (see Figure 1). In this way, the time dynamic in the fold for evaluating is only partially contained in those used for training.

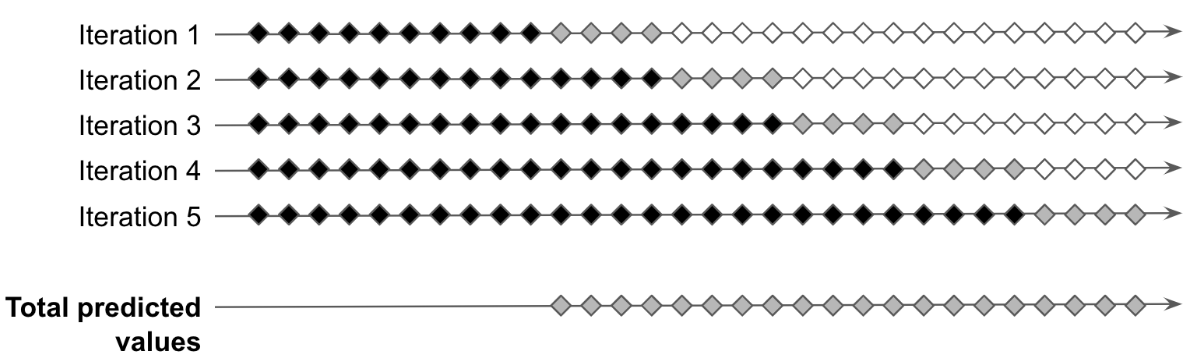
Figure 1 Time-consecutive data splitting for cross-validation for K equals to 4



Caption: In black are records indexes used for training and gray for validation.

The predictive performance is addressed by conducting walk-forward testing; for such a procedure, the time horizon for collecting future-holdout records has to be set and it should be representative of how the model is to be used. For instance, if the studied phenomenon is needed for planning purposes on a monthly basis, and the time instances are aggregated weekly, a representative horizon could be 4 instances in the future. In this context, a representative test can be established by iteratively estimating the model over an increasing range of records by 4 instances, 4 predictions will be made in each iteration, and finally measuring how close the predictions are from the actuals. The outcome of this procedure is a time series of future values by concatenating all tranches of length four (see Figure 2)

Figure 2 Diagram by the testing procedure in tranches of length four (example of 5 iterations)



Caption: In black, records indexes used for fitting the model, in gray for testing (hold-out / out-of-sample), white indicates records indexes not used in each iteration.

The model's ability to depict the true underlying dynamic is done with the help of synthetic records. We simulate a data set to be used as 'ground truth' to compare the outlined effects from our proposal against the dynamic established for creating these records. Although the simulated records are generated by setting arbitrary underlying forces driving the phenomenon, these forces ideally should be somehow consistent with reality. In particular, if the variable under study is in the field of advertisement, to simulate the records one should consider introducing some of the effects discussed in literature: current, carryover (also known as Advertising adstock), shape, competitive, dynamic, content, and media effects (see Tellis, (2006) for more details).

5.3. Note on the potential existence of endogeneity

If an input variable, observed or unobserved, that is not included in our models is related to a variable in the design matrix, we might find the so-called endogeneity problem. Concretely, in market response models, there can be situations when unobserved demand shocks are captured in the errors being correlated with ad spending, which wrongly causes brand awareness/recall to improve. Another known problem is that if companies allocate more ad spend to better selling brands (which also have higher awareness/recall), then some might argue that causality is reversed. In cases when these potential issues are required to be tackled, one could include Instrumental Variables (IV) and proceed with the estimation in two steps, similarly to Xiao & Xu (2015). However, the advantages of this method for depicting the true IVs require more investigation. Other studies on extracting the true relationships in the presence of some invalid instruments using a similar regularization as in this work are Kang et al. (2016) and Windmeijer et al. (2018); though, these studies are based on the Lasso method, while in this work, Elastic-Net is preferred.

For the present work, addressing these two potential issues will be considered outside the scope, as our main focus is on presenting a market response model method that combines predictive validity and explicability by empirically extracting the effects depicted by the model. Further analyses and improvements could be made on the basis of this proposal.

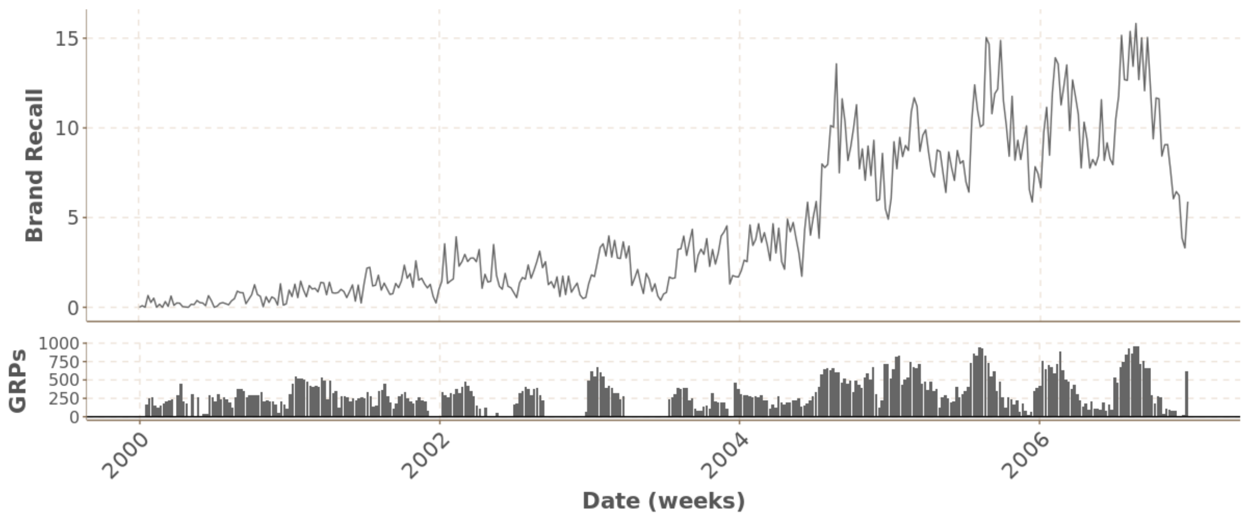
6. Dataset description

We apply the developed framework for estimating and predicting the Brand Recall (BR) (Keller, 1993) in the adult Spanish population, using records from a large company working in the financial and insurance sector with headquarters in central Europe. This company already has a long and well established business in several central European countries and entered the Spanish market in the year 2000. Data was provided through Vértice Sistemas S.L. previous agreement with the media planner which will remain anonymous.

Brand Recall (BR) together with *Brand Recognition* (BRg) (Belch, 2017) are measures commonly used to represent Brand Awareness, and it represents the ability of consumers to recognize and recall a brand in different situations (Aaker, 1996). BR denotes how often a certain targeted population correctly generates the brand name from memory when prompted by a product category, and it is our target variable. The data set consists of two different sources; the first contains daily GRPs ranging from 18/01/2000 to 31/12/2006 (dd/mm/yyyy) achieved by broadcasting advertisements on TV, characteristics related to the type and duration of messages, time frames (i.e. primetime, etc) and the features related to the TV channel. The second one contains the Brand Recall collected weekly by telephone interviews from 2000 to 2006.

Features related to the media planning configuration will not be in the scope of this study; only BR and GRPs are considered. The reason is because characteristics related to the type and duration of messages, time frames when broadcasted, and other features in the sample, are potential drivers for explaining variations on GRPs but indirect drivers of the studied phenomenon. In other words, we are not interested in how to achieve GRPs but its effect on BR. Aggregating from daily to weekly by summing GRPs (assigning to Monday), the number of observations is N=363; the two variables of interest are plotted in Figure 3.

Figure 3: Data sample visualization (Brand Recall & GRPs)



During the whole period under study, none or very few efforts were put in other media channels besides the traditional ones (i.e. TV, radio, etc.), with the advertisement investment in television the predominant by far (no further details were disclosed).

7. Experiment settings

In this section, we describe the features creation for constructing the model and the settings to proceed with the estimation of the parameters using Elastic-Net.

7.1. Features Creation

Original data records include the weekly BR as target variable and GRPs (sum of 7 days) as the explanatory variables. Next, we proceed extending the set of inputs that might drive this phenomenon; this is done by generating a set of hypotheses regarding the forces which may explain Brand Recall variations and materialized in the form of variables included in the matrix □.

7.1.1. *Lagged features on the target and GRPs*

Without prior knowledge of the true link between the target and input variables, features representing mechanisms related to the distributed lags should not be selected as is traditionally done in regression problems. Since the intention is not to confirm our knowledge about this phenomenon, but instead discover the mechanisms from an exploratory perspective, the idea is to construct a design matrix that contains a large enough number of plausible potential drivers such that the underlying dynamic can be depicted via selection of a subset of them with the help of regularization.

As media planning often is decided on a monthly basis, we create 4 lagged features of the target variable (BR) for constituting the autoregressive part of the model. The decision of the order of lagged features is more uncertain in the case of GRPs, but given the exploratory nature of our study, we simply decided to arbitrarily create 8 of these features. Moreover, as it is expected that a linear increase in advertising exposure does not have a similar effect on brand awareness, but each increment in advertising exposure causes a progressively lesser effect on awareness (i.e. diminishing returns effect), features referring to GRPs are transformed to the logarithm of original values plus one.

7.1.2. *Structural changes as potential qualitative changes*

A simple visual analysis of Figure 3 suggests that there are forces other than GRPs driving BR, as there is not a constant linear dependency relationship between the former and the latter over the whole time frame. Specifically, large sudden increases occur from mid 2004 onwards. Despite not knowing the actual drivers of these changes, one can focus on identifying when these happen and construct input variables to estimate those effects regardless of what the true cause is. In our dataset, these changes in the form of a *step* (i.e. time series level is shifted either up or down) may be the consequence of implementing new advertising actions, so a plausible hypothesis is that they correspond to qualitative variations on the advertising message and/or a change in the theme. We will identify weeks where structural changes occur following the approach suggested by Zeileis et al. (2003) which is available in an R-package called *Strucchange* (Zeileis et al. 2002).

The break points where there are step changes are identified at 2001/12/24, 2004/03/08 and 2005/07/11 (yyyy/mm/dd). These are extracted from a regression using the lagged GRPs features (log transformed) as shown in equation (6). Therefore, three *Step* variables will be added to the design matrix.

(6)
$$BR_t = \alpha + \sum_{i=0}^{8} \log \beta_i \, GRP\_i_t + \epsilon_t$$

i denotes the order of lagged features in weeks and so GRP_0 refers to contemporaneous effects and this variable captures the current effect of GRPs, whereas those belonging to GRP_[1-8] represent effects on BR from the GRPs that occurred 1 to 8 weeks ago. The inclusion of structural changes combined with the non-linear effects as decay and wear-in/out, enables the outlining of changes in these latter effects as a consequence of a variation in the theme of the ad or due to other qualitative changes, and so overcoming the critique to Distributed Lag Models reported by Bass et al. (2007).

7.1.3. *Increasing acquisition of broadcast spaces during Christmas*.

Other potential sources of variations not evident just by looking at the time series of our example, are related to the increasing demand for slots to broadcast ads in certain periods. For instance, one might expect a certain level of influence when direct competitors and other unrelated brands compete for time to broadcast their messages, as it may happen during the Christmas period.

In this practical example, three binary features representing competition during Christmas will be considered: (a) from 10th to 31st of December which depending on the year covers 3 or 4 full weeks, (b) weeks ranging from 15th to 31st of December (2 or 3 weeks), and (c) from the 20th to 31st of December (1 or 2 full weeks).

7.1.4. ***Features in the design matrix constituting the candidate model definition***

The final set of features in the design matrix □ to estimate the model is described in Table 3, and the regression model with intercept and Gaussian response is stated in equation (7).

Table 3: Features considered in the design matrix

#	Variable name	Type	Description
	Lagged target (autoregressive)		
1-4	BR_[1-4]	Numeric	Lagged features on the target
	Current and lagged GRPs		
5-13	GRPs_[0-8]	Numeric	Contemporaneous and Lagged features on the GRPs
	Steps: qualitative changes		
14-16	step_[1-3]	Binary	Indicates the presence of a potential qualitative changes
	Competition effect during Christmas		
17-19	Christmas_[1-3]	Binary	Indicates the presence of potential effects during Christmas period

$$\square\square_{\square} = \square + \sum_{\square=1}^{4} \blacksquare\, \square_{\square}\square\square\_\square_{\square} + \sum_{h=0}^{8} \blacksquare\, \square_{h}\square\square\square\_h_{\square} + \sum_{\square=1}^{3} \blacksquare\, \square_{\square}\square\square\square\square\_\square_{\square} + \sum_{\square=1}^{3} \blacksquare\, \square_{\square}\square h\square\square\square\square\square\square\square\_\square_{\square} + \square_{\square} \quad (7)$$

7.2. Algorithm's parameters tuning and estimation

The estimation is performed with an Elastic-Net penalty with the sequence of *lambdas* internally suggested by *Glmnet*. The optimal value is selected via cross-validation such that it achieves the minimum loss within the cross-validation (aliased 'lambda.min'). The loss function for cross-validation is the Mean Square Error (MSE), and the penalty mix parameter *alpha* is set to 0.5 (default), meaning that Lasso and Ridge penalties are equally considered (alpha=1 means lasso penalty, alpha=0 ridge penalty). Values in the input matrix are standardized prior to fitting the model using the built-in function in *Glmnet*.

To prevent conceptual misalignments on the effects' signs, restrictions are introduced on the value limits of the coefficients. An example of misalignment is that effects from GRPs should be either positive or zero, but never negative. Additionally, distinct penalties factors affecting coefficients are established in a way that contemporaneous effects from GRPs and lagged effects of the target variable are more prone to be part of the expression resulting from the estimation (i.e., we emphasize capturing the lagged component if it exists). A penalty value of 0 on a parameter's shrinkage implies no shrinkage, and these variables are always included in the model, whereas larger values demand greater predictive power in terms of fitting improvement to qualify as part of the model. In this example, the algorithm is configured as follows: (1) coefficients reflecting the competition effect during the Christmas period are constrained to be lower or equal to zero, while the rest of the coefficients are restricted to be greater or equal to zero, and (2) coefficients related to the autoregressive component and contemporaneous GRPs have a shrinkage penalty of 1 with the remainder inputs having a penalty 10 times larger. In summary, the estimation procedure requires stronger predictive evidence from all inputs that are not part of the autoregressive or contemporaneous GRPs component, and coefficients' signs are restricted to what is conceptually expected. With the above setting, the procedure will be prone to construct a mathematical expression as a function of the lagged target variables and current effects from GRPs, while other features will take part in the expression if there is strong evidence of improving the fitting.

Given the relatively large number of restrictions established on the algorithms, some might argue that there are many highly informative priors in the form of restrictions, which is, to a certain extent, similar as it is done with Bayesian approaches. However, it is worth noting that the "priors" utilized here are far less restrictive than assuming a certain distribution per parameter, which implies constraining not only the type of distribution but also the mean and the standard deviation. Moreover, it seems more reasonable to expect practitioners to determine bounds on the minimum and/or maximum value of the parameters than knowing to what class of distribution every single parameter belongs, plus its mean and standard deviation.

8. Results

The software scripts for conducting this analysis have been created in the R language for statistical computing and graphics (R Core Team, 2019) version 3.6.2, visualizations are done with the package *ggplot2* (Wickham, 2016).

The mathematical expression resulting from the estimation consists of the coefficients distinct to zero presented in Table 4. Variables GRPs_[1-8] and Christmas_3 yielded coefficients that shrunk to 0 and therefore are not part of the expression that best estimates this phenomenon. P-values are not calculated, as in a regularization framework variables are selected based on their predictive value in the cross-validation instead of its p-value. The proportion of the variance explained, measured by the R^2, is equal to 0.9215 (~92%).

Table 4: Non-zero coefficient values after the estimation process

Variable name	Coefficient	Value[1,2]
(Intercept)	α	-0.7289
Lagged target (autoregressive)		
BR_1	θ_1	0.3763
BR_2	θ_2	0.1832
BR_3	θ_3	0.1229
BR_4	θ_4	0.1417
Current and lagged GRPs		
GRPs_0	β_0	0.1948
Steps: qualitative changes		
step_1	γ_1	0.4631
step_2	γ_2	0.9050
step_3	γ_3	1.4198
Competition effect during Christmas		
Christmas_1	δ_1	-0.1708
Christmas_2	δ_2	-0.3748

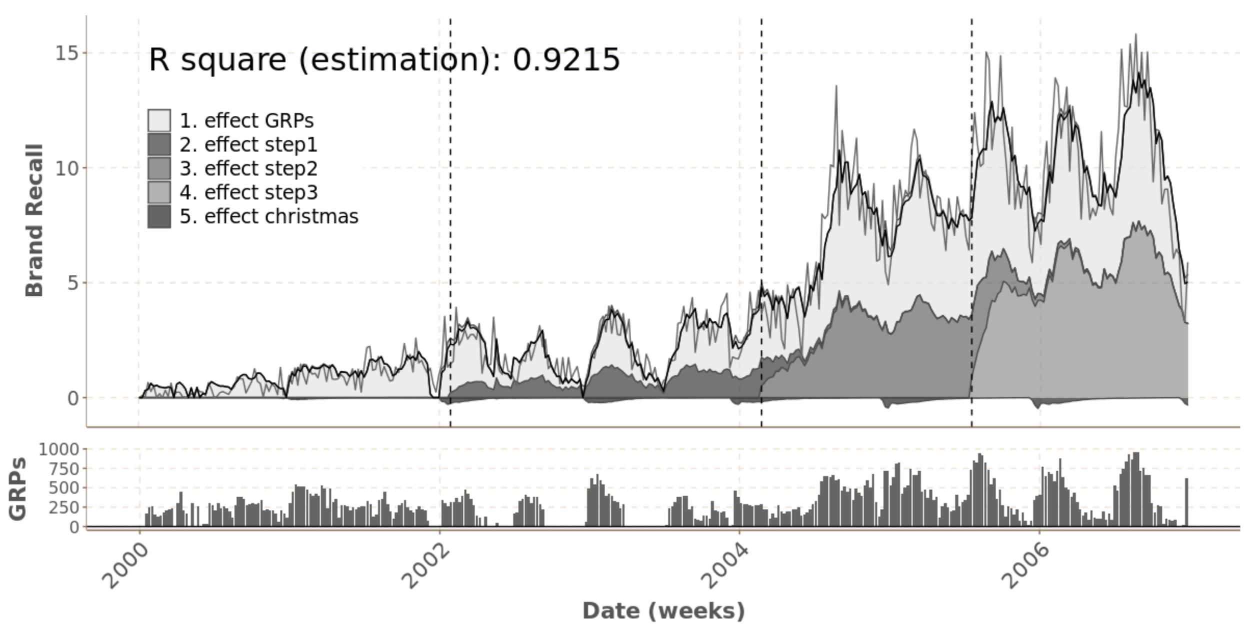
[1] Rounded to the fourth decimal digit
[2]To reproduce the exact same results, the seed of the pseudo-random generation process has to be fixed as the estimation procedure contains a random component.

In spite of the outcome showing that the lagged GRPs are not part of the expression, the presence of an autoregressive component, as it is seen in Table 4, indicates that the effects of exogenous variables are functions of both their coefficients and are autoregressive, whereby non-linearity exists as a consequence of distributed lags as discussed in Section 5.1. Thus, although the lagged variables from GRPs were considered irrelevant by the estimation procedure, its effect still is distributed (i.e., Brand Recall is driven by past values of GRPs). An implication of having only GRP_0 as part of the estimation output is that the same decay form affects all external explanatory variables (i.e., GRPs, Steps, Christmas) in this dataset. Instead of interpreting the concrete values of each parameter in this section, we suggest a much more insightful and easy-to-understand pathway, which is to present the effects in the next section

8.1. Fitting and effects extraction

Effects are extracted empirically following the approach described in Section 5.1, and together with the model fitting are visualized in Figure 4. Effects referring to the Christmas period are considered conjointly and the same with variables of GRPs.

Figure 4: Actuals, model fitting and effects in stacked form. GPRs at the bottom as reference.

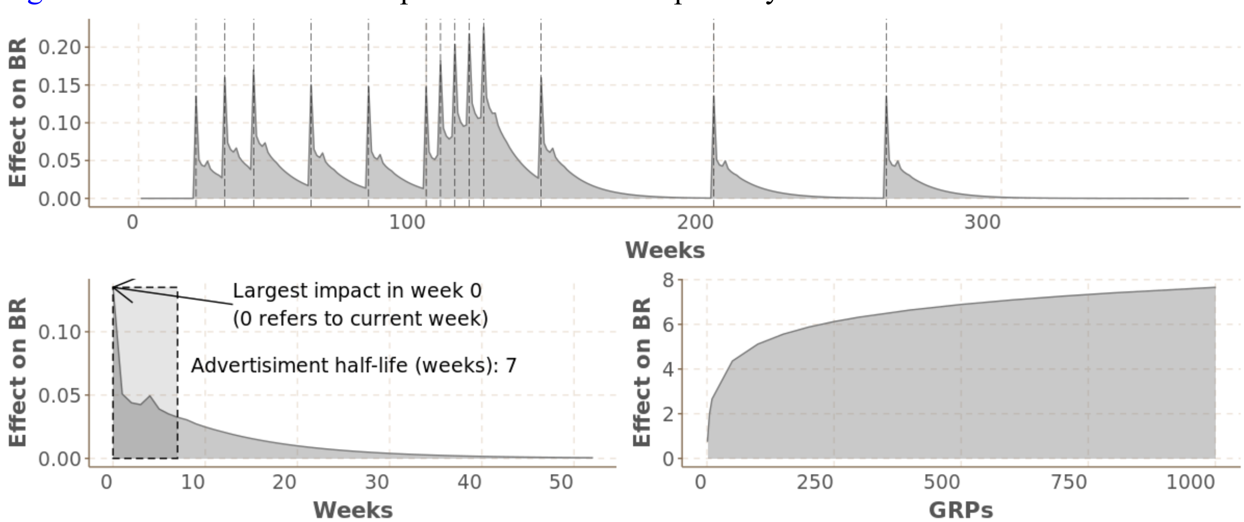


Caption: Gray line refers to actual values, black thicker line represents the model fitting. Different colors denote different effects. Vertical dotted lines indicate the structural changes found to construct the Step features.

It is estimated that the competitive effect during Christmas is not very large. A possible explanation could be that not only do competitors increase the number of actions during this period but also the company's object of this study (in the data set message's duration during Christmas increases 5% and broadcasting in PrimeTime slots is 7% more frequent). Large impact is seen from variables related to assumed qualitative changes in messaging (i.e. Steps_[1-3]), implying that in the context of this study Brand Recall variability is largely driven by qualitative factors, whereas effect from GRPs is the main force driving brand awareness during the whole time frame. Patterns known as current, shape, carryover and the wear-in wear-out of advertising are embedded in Figure 4, these are later isolated and shown in Figure 5.

Decay and wear-in wear-out are outlined by creating GRPs to pulses of unitary value on a set of arbitrary time instances (weeks) in the input matrix. To extract the progressively lesser effect on awareness (i.e. diminishing returns effect), we proceed similarly as when constructing the decay shape, but instead of using a single pulse of unitary value, the procedure is applied multiple times by increasing the pulse's value from 1 to 1000 GRPs.

Figure 5: Three of the relevant patterns extracted empirically



Caption: Subfigure at the top represents wear-in wear-out, decay is shown at the bottom-left, at the bottom-right the non-linear effect on the target as response to increasing intensity of advertising is presented.

The decay function is seen as the lifespan that a message has in terms of Brand Recall. Some studies report that the half-life of advertising ranges from 7 to 12 weeks (Leone, 1995) and according to our experience, industry practitioners report shorter values typically around 2 to 5 weeks. Although this topic is fundamental to the management of a company, the duration and scope of the advertising effects have not been precisely determined within the academy (Peterson & Jeong, 2010; Wang, 2008). In our study, half-life is estimated to be 7 weeks (including the week where the ad is broadcasted).

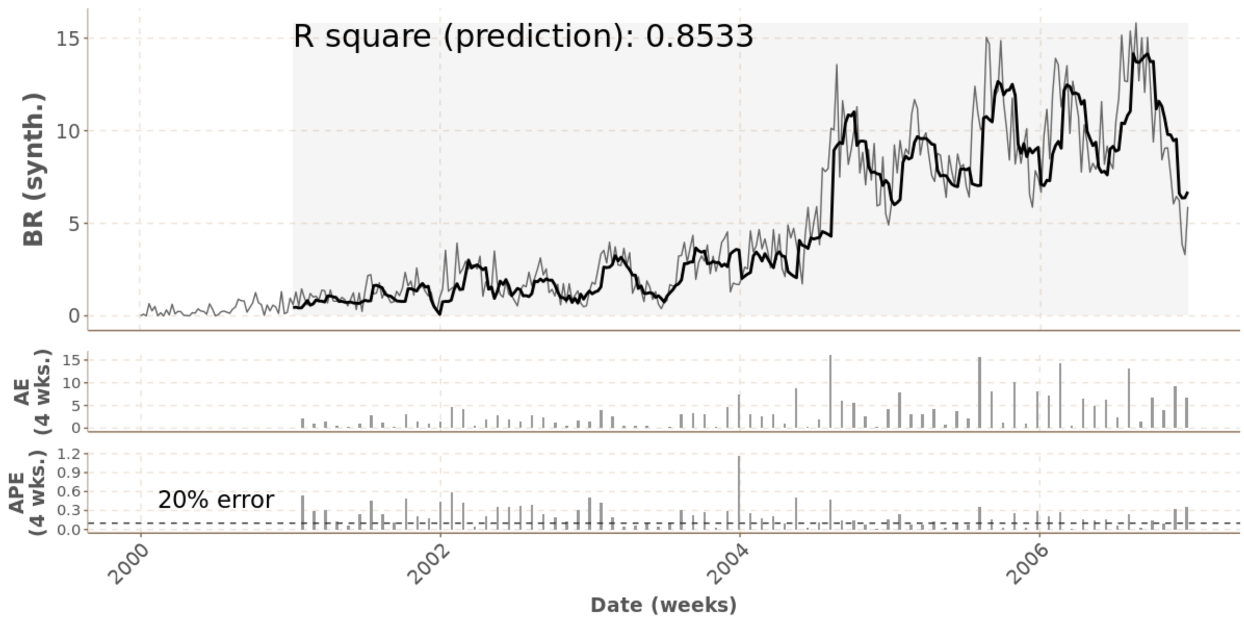
8.2. Testing on prediction

We examine the performance on strictly predictive tasks over future hold-out records, which provide insights about potential spurious relationships present in the model and the so-called overfitting problem. If the explained variability achieved in-sample differs greatly from the yielded in future-holdout records, the model is disqualified as representative of the system under study, as it does not properly represent the true forces driving the phenomenon.

The test consists of iteratively estimating the model over an increasing range of records of 4 instances (weeks) and predicting for subsequent 4 time instances (see Section 5.2**,** Figure 2) starting from the 54th week. Afterwards, all batches of predicted values are concatenated in a single time series and compared with actuals in terms of explained variance by the R^2. For predicting 4 instances ahead and avoiding the so-called *data-leakage*, predicted values will be used in the autoregressive component. More specifically, the first prediction is computed using past values of BR, second prediction is calculated based on latter prediction plus the previous 3 actual values and so on until the forth prediction, wherein the autoregressive component will be composed by 3 predictions and an actual value of BR. Future values of explanatory variables are also required for predicting, meaning that another source of uncertainty takes part in the process. In this test, future GRPs values, as well as binary variables of Steps and those referring to Christmas, are gathered from the original data sample. Results from this test should be interpreted as: *with a perfect prediction of GRPs in next 4 weeks, indicating qualitative changes we know will occur, and specifying future time periods referring to Christmas, this model is capable of explaining about X% of the variability in terms of Brand Recall*.

Notwithstanding the previous remarks, we state that this experiment is an accurate assessment of the method's predictive ability, as the noise derived from miscalculating future GRPs constitutes part of a different problem. Our results in terms of variance explained by R^2 in predictive tasks is equal to 0.8533, as expected smaller than in estimation (~85% versus ~92%). We consider this drop as not large enough to disqualify the model's abilities for depicting the underlying forces driving Brand Recall. Results are shown in Figure 6.

Figure 6: Visual results of the prediction test



Caption: In the top figure actuals (gray) and predictive values (black) are shown, predicting range is denoted by the shadowed area. Mid and bottom subfigures contain the Absolute Errors (AE) and Absolute Percentage Errors (APE) respectively, in both cases aggregated by month (prediction horizon of 4 weeks).

Percentage errors in prediction are larger when the estimation is performed in chunks of data belonging to the first 4 years and target variable ranges from 0 to 5, being around and below 20% when the estimation is done with a wider range of records (> 4 years), and the target variable larger than 5. It should not be forgotten that percentual errors greatly depend on the records dimension, being more complex to have small errors when values in the target variable are small. In this study, the target is around 2 to 4 orders of magnitude smaller than in other works commonly using sales. We have not found works measuring purely predictive performance as presented here for benchmarking purposes.

8.3. Testing on simulated data

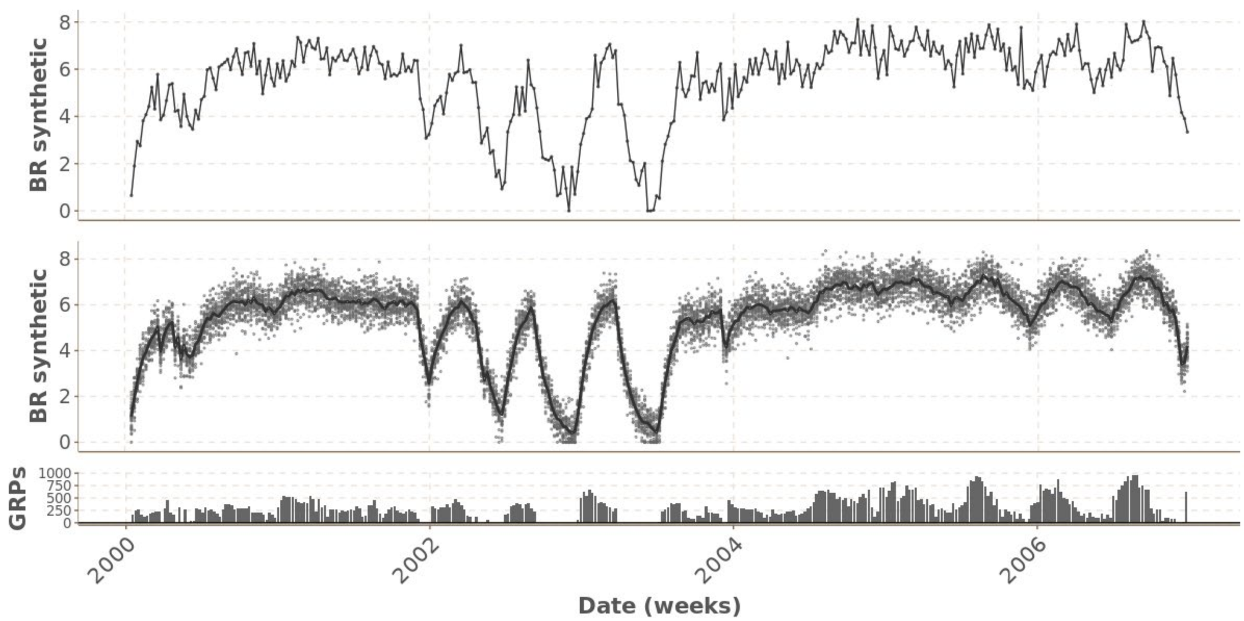
Seeking a further assessment over the ability of the method to depict the true effects, we generate a set of synthetic records to be used as a 'ground truth'. Based on the GRPs in the practical example described before, a variable representing Brand Recall is simulated by fixing arbitrary dynamics acting as current and carry-over effects, and adding a stochastic component to both forces. The test consists of applying our method to explain and predict the values of the synthetic BR, and most importantly to compare the estimated effects depicted versus the true dynamic that generates the records. The mechanisms used for simulation should not be considered as an accurate representation of how BR emerges in a general context, but as a simplistic case of how it could emerge for validation purposes.

Brand Recall is simulated by setting GRPs current effect with value 0.2 and a decay functional form as in equation (8) with $\rho = 0.2$, meaning that the so-called advertisement half-life is 4 weeks (current week plus 3 weeks after) and the largest impact on BR occurs in week 0.

(8) $$□(□□□,□) = □□□ \cdot □□□(-□□)\ □□□h\ □ \in □ = \{1, 2, \ldots, □\}\ □□□\ 0 < □ < 1$$

The autoregressive component order is set to 1 (i.e. AR(1)) and its effect is constructed from values arising out of the decay shape in equation (8) within time instances ranging from 1 to D. In our experiment, D=40 (40 weeks) is selected, as it is long enough such that the decay shape is fully defined and embedded in the simulation. Next, a random noise is added to this decay following a Normal Distribution with a mean of 0 and a standard deviation equal to 4% of the decay value. The lag effect with noise is applied to GRPs per time instance, concretely, we map every individual value of GRPs to a time ordered vector of length 40. Then, all decay sequences are aggregated, summing per time instance, and a random noise is added to represent the 'measurement error' when collecting the Brand Recall. This latter noise is generated from a Normal Distribution with a mean of 0 and standard deviation equal to the maximum value between 0.5% and 8% of the synthetic BR before this step. The simulated BR is the result of previous steps and negative values are imputed with 0. The synthetic data is created without introducing changes in the form of steps, nor effects related to particular time periods such as Christmas. Graphical results of the simulation are shown in Figure 7.

Figure 7: Simulated Brand Recall. GPRs at the bottom as reference.



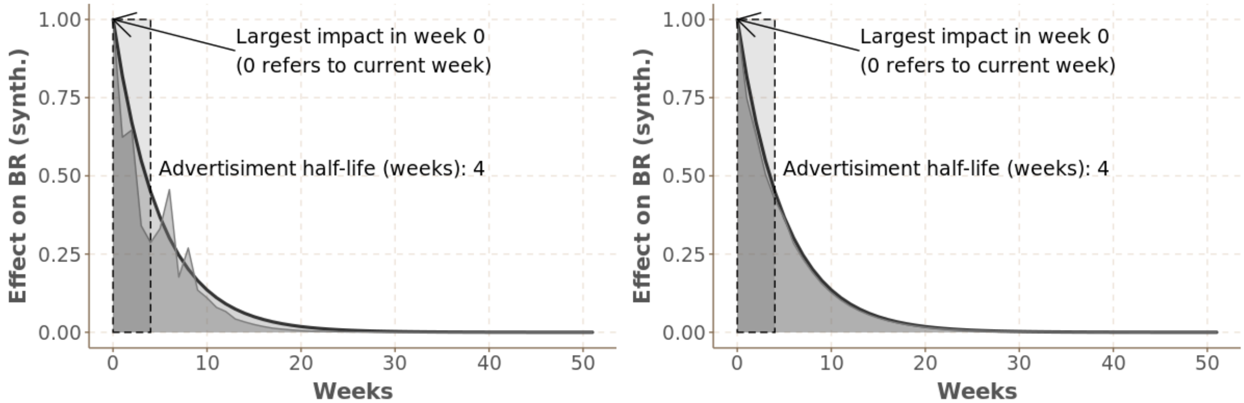
Caption: At the top a single simulation of the time series, the figure in the middle is the average time series of 30 simulations together with the values simulated per time instance represented by dots in gray color.

Exactly the same settings as with the real-world data sample are applied (see Section 7): a set of lagged features on the target variable and GRPs are created, as well as binary features intending to capture potential effects during Christmas. Structural changes will be searched following an identical procedure and added to the design matrix if found. Last but not least, the algorithm's parameters are set equally as when using the real-world dataset. For a detailed assessment, the analysis is conducted over a single simulation and over the averaged time series of 30 simulations (top and middle plots in Figure 7). The reason for conducting the analysis twice is because data records from a single time series might not be the best representation of this phenomenon due to the random noise.

As proof of the capacity to distinguish irrelevant effects, (a) in both cases no structural changes are identified as relevant by the method, whereby these variables are shrunk to zero during the model training. (b) Coefficients related to the Christmas period are also shrunk to zero or very close to zero, and so is its effect. The method applied to a single simulated time series yields values for the R^2 metric of ~0.9145 in-sample and ~0.9084 in prediction; when applied to the averaged time series of 30 independent simulations, values are ~0.9969 in both in-sample and prediction. The exact coefficient values resulting from the estimation are not relevant for this test and therefore not reported.

Decay comparison shows to be close to what is used in the simulation, advertisement half-life and estimated largest impact of GRPs matches the ‘ground truth’ (i.e. 4 weeks). Nevertheless, discrepancies are found in the decay extracted from a single simulation: a shape with humps rather than a smooth decreasing form. The same comparison from applying the method to the averaged time series demonstrates a very accurate extraction of this force (see graphs in Figure 8) implying that with low levels of noise, the method is capable of precisely depicting what is expected.

Figure 8: Decays extracted empirically versus decay used in the simulation



Caption: Gray line denotes the empirical effect extracted and thicker black line the true effect. Left figure is the comparison using the decay shape extracted from a single simulation, the right figure is the comparison with the extraction from the averaged time series, both effects match almost perfectly in the latter.

With the results presented, we corroborate the method's ability to actually represent the true underlying dynamic of a phenomenon driven by current and lagged effects. Nonetheless, one has to keep in mind that the true dynamic might not be outlined totally accurately in high noise level circumstances.

9. Conclusions

In this work, we have presented an approach based on Distributed Lag Models (DLagMs) for estimating phenomena involving the notion of distributed lags in an exploratory manner, consistent with the underlying conditions and assumptions of Marketing Response models. Our proposal uses Elastic-Net regularization and a cross-validation framework to estimate the model parameters, which overcomes the problems of the traditional estimation method. Moreover, we suggested a procedure to extract the model effects empirically to ease this task when models contain autoregressive components, as in the case of DLagMs. We presented a set of results obtained in a circumstance similar to how the method would be used in a company for decision-making purposes, which is crucial for demonstrating its practical usability. Last, we corroborate the method's ability to accurately outline the commonly discussed effects of advertising next to showing high predictive power.

This study should be considered within the context of several limitations. First, further investigation is needed using sample records comprising both traditional and online media channels. Second, a more extensive benchmark is required by comparing different methods using the same data sample and with the evaluation framework described here. Third, the goodness of the method briefly suggested for coping with endogeneity remains an open question for future investigations.

References

Allen, D. M. (1974). "*The Relationship between Variable Selection and Data Augmentation and a Method for Prediction*". Technometrics, 16(1), 125–127. https://doi.org/10.2307/1267500

Bass, F., Bruce, N., Majumdar, S. & Murthi, B.P.S. (2007), *"Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising-Sales Relationship"*. Marketing Science 26 (2) 179-195 https://doi.org/10.1287/mksc.1060.0208

Bass, F., & Clarke, D. (1972), "*Testing Distributed Lag Models of Advertising Effect*". *Journal of Marketing Research*, 9(3), 298-308. doi:10.2307/3149541

Belch, G.E. (2017), "*Advertising and promotion : an integrated marketing communications perspective*". *Belch, Michael A. (Eleventh ed.)*. New York, NY. ISBN 978-1-259-54814-7. OCLC 968690198.

Belloumi, M. (2014), "*The relationship between trade, FDI and economic growth in Tunisia: An application of the autoregressive distributed lag model*". *Economic Systems*, Volume 38, Issue 2, 269-287. https://doi.org/10.1016/j.ecosys.2013.09.002.

Berger, J., A. Humphreys, S. Ludwig, W.W. Moe, O. Netzer, and D.A. Schweidel. 2019. "*Uniting the tribes: Using text for marketing insight*". Journal of Marketing. https://doi.org/10.1177/00222 42919873106.

Berk, A. (2017), "*The analysis of the relationship between sunflower production and its price by using the Koyck model in Turkey*". CUSTOS AGRONEGOCIO ONLINE. 2017; 13(4):42–53.

Binet, L., & Field, P. (2017). "*Media in focus: Marketing effectiveness in the digital era*". Institute of Practitioners in Advertising.

Bilgin, Y. (2018), "*The Effect Of Social Media Marketing Activities On Brand Awareness, Brand Image And Brand Loyalty*", *BMIJ*, 6(1): 128-148 doi: http://dx.doi.org/10.15295/bmij.v6i1.229

Borden, N. H. (1964), "*The concept of the marketing mi*x". *Journal of advertising research*, 4 (2), 2–7.

Brown, G. (1986), "*Modelling advertising awareness*". *Journal of the Royal Statistical Society*. Series D (The Statistician) 35(2), 289–299.

Bruce N.I., Foutz N.Z. & Kolsarici C. (2012) "*Dynamic Effectiveness of Advertising and Word of Mouth in Sequential Distribution of New Products*". Journal of Marketing Research. 2012;49(4):469-486. doi:10.1509/jmr.07.0441

Cawley, G. & Talbot, N. (2010), "*On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation*". Journal of Machine Learning Research. 11. 2079-2107.

Chen L. & Shupei Y. (2019), "*Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media*", Journal of Interactive Advertising, 19:1, 58-73, DOI: 10.1080/15252019.2018.1533501

Clarke, D. G. (1976), "*Econometric measurement of the duration of advertising effect on sales*". Journal of Marketing Research, 13, 345–357.

Csereklyei, Z., Ancev, T., & Qu, S. (2019), "*The effect of wind and solar power generation on wholesale electricity prices in Australia*". Energy Policy. 131. 358-369. 10.1016/j.enpol.2019.04.007.

Danaher, P. (2021), "*Advertising Effectiveness*" Impact at JMR, (January), Available at: https://www.ama.org/2021/01/26/advertising-effectiveness/

Danaher, Peter J. and Tracey S. Dagger (2013), "*Comparing the Relative Effectiveness of Advertising Channels: a Case Study of a Multimedia Blitz Campaign*" Journal of Marketing Research, 50 (4), 517–34.

Danaher P, Danaher T, Smith M & Loaiza-Maya R (2020). "*Advertising Effectiveness for Multiple Retailer-Brands in a Multimedia and Multichannel Environment.*" Journal of Marketing Research. 2020;57(3):445-467. doi:10.1177/0022243720910104

Demirhan, H. (2020), "*An R package for distributed lag models and ARDL bounds testing*". PLoS ONE. 15(2): e0228812. https://doi.org/10.1371/journal.pone.0228812

Ehrenberg, A., Barnard, N. & Sharp, B. (2000), "*Problems With Marketing's 'Decision' Models*". ANZMAC 2000. Visionary Marketing for the 21st Century: Facing the Challenge

Franses, Ph. H. B. F, & van Oest, R.D. (2004), "*On the econometrics of the Koyck model*". (No. EI 2004-07). Report / Econometric Institute, Erasmus University Rotterdam. Retrieved from http://hdl.handle.net/1765/1190

Friedman, J., Hastie, T. & Tibshirani, R. (2010), "*Regularization Paths for Generalized Linear Models via Coordinate Descent*". Journal of Statistical Software, 33(1), 1–22. http://www.jstatsoft.org/v33/i01/.

Fu, W. (1998), "*Penalized regression: the bridge versus the lasso*". J. Computational Graph. Statist., 7, 397–416.

Guido, G., Prete, M. & Miraglia, S. & Mare, I. (2011), "*Targeting direct marketing campaigns by neural networks*". Journal of Marketing Management. 27. 992-1006. 10.1080/0267257X.2010.543018.

Hair, N., Moira, C., & Shapiro, M. (2010), "*Toward a Classification System of Relational Activity in Consumer Electronic Communities: The Moderators' Tale*". Journal of Relationship Marketing, 9 (1), 54–65.

Hannan, E. J., Dunsmuir, W. T., & Deistler, M. (1980). "*Estimation of vector ARMAX models*". Journal of Multivariate Analysis, 10(3), 275-295.

Hastie, T. , Tibshirani, R. , & Friedman, J. (2009), "*The elements of statistical learning: Data mining, inference, and prediction*". Second Edition. *Springer*.

Havlena, W. J. & Graham, J. (2004), "*Decay effects in online advertising: Quantifying the impact of time since last exposure on branding effectiveness*". *Journal of Advertising Research* 44(4), 327–332.

Heaton, M., Olenick, C., & Wilhelmi, O. (2019), "*Age-specific distributed lag models for heat-related mortality*". Environmetrics. 30:e2561. https://doi.org/10.1002/env.2561

Hoerl, A. & Kennard, R. (1988), "Ridge regression". *Encyclopedia of Statistical Sciences*, vol. 8, pp. 129–136. New York: Wiley

Huffaker, R. & Fearne, A. (2019), "*Reconstructing systematic persistent impacts of promotional marketing with empirical nonlinear dynamics*". PLOS ONE. 14(9). https://doi.org/10.1371/journal.pone.0221167 PMID: 31532779

Iacobucci, D., Petrescu, M., Krishen, A. et al. "*The state of marketing analytics in research and practice*". J Market Anal 7, 152–181 (2019). https://doi.org/10.1057/s41270-019-00059-2

Kang H., Zhang A., Cai T. T., and Small D. S. (2016), "*Instrumental Variables Estimation with some Invalid Instruments and its Application to Mendelian Randomization*", Journal of the American Statistical Association, 111, 132–144.

Keller, K.L. (1993), "*Conceptualizing, Measuring, and Managing Customer-Based Brand Equity*". Journal of Marketing. 57 (1): 1–22. doi:10.2307/1252054. ISSN 0022-2429. JSTOR 1252054.

Kitchen, P.J. (2010), "*Integrated Brand Marketing and Measuring Returns*". Palgrave Macmillan, London. https://doi.org/10.1057/9780230297340_1

Koyck, L.M. (1954). “*Distributed Lags and Investment Analysis*”. North Holland Publishing Company, Amsterdam, pp: 21-50.

Kyung, M., J. Gil, M. Ghosh, G. Casella. 2010. “*Penalized regression, standard errors, and Bayesian lassos*”. Bayesian Anal. 5(2) 369–412.

Lauenroth, W. K. (2003). “*Models in ecosystem science*”. Princeton University Press.

Leone, R.P. (1995), “*Generalizing what is known about temporal aggregation and advertising carry-over*”. Marketing Science, 14, G141-G150.

Li, Q., N. Lin. (2010). “*The Bayesian elastic net*”. Bayesian Anal. 5(1) 151–170.

McCarthy, J. E. (1978), “*Basic marketing: a managerial approach (6th ed.)*”. Homewood, Il: R.D. Irwin.

Midgley, G., J. Nicholson, and R. Brennan. (2017). “*Dealing with challenges to methodological pluralism: The paradigm problem, psychological resistance and cultural barriers*”. Industrial Marketing Management 62: 150–159.

Mohammed, S.M.A., Guo, P., Haq, I.U., Pan, G. & Khan, A. (2019), “*Do government expenditure and financial development impede environmental degradation in Venezuela?*” PLOS ONE. 14(1):e0210255. doi: 10.1371/journal.pone.0210255. PMID: 30629649; PMCID: PMC6328114.

Moorman, C. (2016). “*Celebrating marketing's dirty word*”. Journal of the Academy of Marketing Science 44 (5): 562–564.

Mulchandani, K., Mulchandani, K. & Attri, R. (2019), "*An assessment of advertising effectiveness of Indian banks using Koyck model*". Journal of Advances in Management Research. 16. 10.1108/JAMR-08-2018-0075.

Naik, P., Mantrala, M. & Sawyer, A. (1998), "*Planning media schedules in the presence of dynamic advertising quality*". Marketing science, 17(3), 214–235.

Nerudova, D. & Dobranschi, M. (2019), "*Alternative method to measure the VAT gap in the EU: Stochastic tax frontier model approach*". PLOS ONE. 14(1):e0211317. https://doi.org/10.1371/journal.pone.0211317 PMID: 30689663

Nothdurft, A. & Engel, M. (2019), "*Climate sensitivity and resistance under pure-and mixed-stand scenarios in Lower Austria evaluated with distributed lag models and penalized regression splines for tree-ring time series*". European Journal of Forest Research. p. 1–23.

Özsayin, D. (2017), "*Investigation of Production and Price Relationship in Cow Milk Production by Koyck Model Approach*". Turkish Journal of Agriculture-Food Science and Technology. 5(6):681–686. https:// doi.org/10.24925/turjaf.v5i6.681-686.1164

Peterson, R. & Jeong, J. (2010), "*Exploring the impact of advertising and R&D expenditures on corporate brand value and firm-level financial performance*". Journal of the Academy of Marketing Science, 1–14

Petrescu, M., & Krishen, A. S. (2019). "*Strength in diversity: methods and analytics*". Journal of Marketing Analytics, 7(4), 203-204. https://doi.org/10.1057/s41270-019-00064-5

R Core Team (2019), "*R: A language and environment for statistical computing. R Foundation for Statistical Computing*", Vienna, Austria. URL http://www.R-project.org/

Rubel, O. & Naik, P.A. (2017) "*Robust Dynamic Estimation*". Marketing Science 36 (3) 453-467 https://doi.org/10.1287/mksc.2016.1010

Rufino, C. (2008). "*Lagged effect of TV advertising on sales of an intermittently advertised product*". DLSU Business & Economics Review, 18(1), 1-12.

Rutz, O. J., Trusov, M. & Bucklin, R. E. (2011). *"Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?"*. Marketing Science 30 (4) 646-665 https://doi.org/10.1287/mksc.1110.0635

Sarstedt, M., & Danks, N. P. (2022). *"Prediction in HRM research–A gap between rhetoric and reality"*. Human Resource Management Journal, 32(2), 485– 513. https://doi.org/10.1111/1748-8583.12400

Sethi, S.P. (1977), "*Dynamic Optimal Control Models in Advertising: A Survey*". Society for Industrial and Applied Mathematics. SIAM Review, Vol. 19, No. 4, pp. 685-725

Sethuraman, Raj, Gerard J. Tellis, and Richard A. Briesch (2011), "*How Well Does Advertising Work? Generalizations from MetaAnalysis of Brand Advertising Elasticities*", Journal of Marketing Research, 48 (3), 457–71.

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). "*The elephant in the room: Evaluating the predictive performance of PLS models*". Journal of Business Research, 69(10), 4552–4564. https://doi.org/10.1016/j.jbusres.2016.03.049

Stone, M. (1974). "*Cross-validatory choice and assessment of statistical predictions*". Journal of the royal statistical society: Series B (Methodological), 36(2), 111-133.

Tellis, G. J. (2006), “*Marketing Mix Model concepts*”. Handbook of marketing research (Chapter 24: Modelling Marketing Mix). methods.sagepub.com

Tellis, G. J. (1988), “*The price sensitivity of competitive demand: A meta-analysis of sales response models*”. Journal of Marketing Research, 15, 331–341.

Tellis, G. J., & Zufryden, F. (1995), “*Cracking the retailer’s decision problem: Which brand to discount, how much, when and why?*”. Marketing Science, 14(3), 271–299.

Tibshirani, R. (1996), “*Regression shrinkage and selection via the lasso*”. J. R. Statist. Soc. B, 58, 267–288

Tsan-Ming, C. (2014), “*Fashion Branding and Consumer Behaviors: An Introduction*”. (chapter 9). Springer. 10.1007/978-1-4939-0277-4_1.

Vakratsas, D. & Ambler, T. (1999), “*How advertising works: What do we really know?*”, *J. Mark.* 63, 26–43. doi: 10.2307/1251999

Verhoef, P.C., E. Kooge, and N. Walk. 2016. “*Creating Value with Big Data Analytics: Making Smarter Marketing Decisions*”. London: Routledge.

Viaene, S., Baesens, B., Van Gestel, T., Suykens, J.A.K., Van den Poel, D., Vanthienen, J., De Moor, B., & Dedene, G. (2001), “*Knowledge discovery in a direct marketing case using least squares support vector machine*”. International Journal of Intelligent System, 16(9),1023–1036.

Wang, F. (2008), “*Reasons for market evolution and budgeting implications*”. Journal of Marketing. 72. 15-30. 10.1509/jmkg.72.5.15.

Weinberg, C. B., & Weiss, D. L. (1982), "*On the Econometric Measurement of the Duration of Advertising Effect on Sales*". Journal of Marketing Research, 19(4), 585–591. doi:10.1177/002224378201900416

Wickham, H. (2016), "*ggplot2: Elegant Graphics for Data Analysis"*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Windmeijer F, Farbmacher H, Davies N, Davey Smith G. "*On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments*". J Am Stat Assoc. 2018;114(527):1339-1350. Published 2018 Nov 13. doi:10.1080/01621459.2018.1498346

Xiao, N. & Xu Q-S. (2015). "*Multi-step adaptive elastic-net: reducing false positives in high-dimensional variable selection*", Journal of Statistical Computation and Simulation, 85:18, 3755-3765, DOI: 10.1080/00949655.2015.1016944

Zeileis, A., Kleiber, C., Krämer, W. & Hornik, K. (2003), "*Testing and Dating of Structural Changes in Practice*". Computational Statistics & Data Analysis. Volume 44, Issues 1–2, 28. Pages 109-123

Zeileis, A., Leisch, F., Hornik, K. & Kleiber, C. (2002), "*strucchange: An R Package for Testing for Structural Change in Linear Regression Models*". Journal of Statistical Software. Volume 7, Issue 2. Pages 1-38. http://www.jstatsoft.org/v07/i02/

Zhang, G., Jansen, B.J., & Spink, A. (2009), "*Identification of factors predicting clickthrough using neural network analysis*". Journal of the American Society for Information Science and Technology, 60(3), 557–570.

Zou, H. & Hastie, T. (2005), "*Regularization and variable selection via the elastic net*", (vol B 67, pg 301, 2005). Journal of the Royal Statistical Society. Series B. 67. 768-768. 10.1111/j.1467-9868.2005.00527.x.