Assessing the effect of advertising expenditures upon sales: a Bayesian structural time series model

Víctor Gallego1, Pablo Suárez-García2, Pablo Angulo3 and David Gómez-Ullate1,2,4

1 Institute of Mathematical Sciences, ICMAT-CSIC
2 Facultad de Ciencias Físicas, Universidad Complutense de Madrid
3 ETSIN, Universidad Politécnica de Madrid
4 Department of Computer Science, School of Engineering, Universidad de Cádiz

emails: victor.gallego@icmat.es, pasuares@ucm.es, pablo.angulo@upm.es, david.gomez-ullate@icmat.es

Abstract

We propose a robust implementation of the Nerlove–Arrow model using a Bayesian structural time series model to explain the relationship between advertising expenditures of a country-wide fast-food franchise network with its weekly sales. Thanks to the flexibility and modularity of the model, it is well suited to generalization to other markets or situations. Its Bayesian nature facilitates incorporating a priori information (the manager’s views), which can be updated with relevant data. This aspect of the model will be used to present a strategy of budget scheduling across time and channels.

Key words: Bayesian structural time series, Sales forecasting, Risk management, Budget allocation

1 Introduction

It is widely acknowledged that a firm’s expenditure on advertising has a positive effect on sales and other aspects [1,2,3,4]. However, the exact relationship between them remains a moot point, see e.g. [5] for a broad survey. Since the seminal work of Nerlove and Arrow [6], several models have been proposed to pinpoint this relationship, although consensus on the best approach has not been reached yet. Two diverging model building schools seem to dominate the marketing literature [7]: a priori models rely heavily on intuition and are derived from general principles, although usually with a practical implementation on mind...
(the aforementioned [6], or [8] and [9] inter alia) and statistical or econometric models, which usually start from a specific dataset to be modelled (see [1] for a review). In this work we will rely mostly on the first type of models since, as we will show, their formulation, interpretation and implementation adapt seamlessly to the state-space or structural time series models, although it will be nevertheless applied to a real dataset.

Bayesian structural time series models [10], in turn, have positioned themselves in the past few years as very effective tools not only for analysing marketing time-series, but also to throw light into more uncertain terrains like causal impact, incorporating a priori information into the model, accommodating multiple sources of variations or making variable selection. Although the origins of this formalism can be traced back to the 1950’s in engineering problems of filtering, smoothing and forecasting, first with Wiener [11] and specially with Kalman [12], these problems can also be understood from the perspective of estimation in which a vector valued time series \{X_0, X_1, X_2, \ldots\} that we wish to estimate (the latent or hidden states) is observed through a series of noisy measurements \{Y_0, Y_1, Y_2, \ldots\}. This means that, in the Bayesian sense, we want to compute the joint posterior distribution of all the states given all the measurements [13]. The ever-growing computing power and release of several programming libraries in the last few years like [14, 15] have in part alleviated the difficulties in the mathematical underpinning and computer implementation that this formalism suffers, making these methods broadly known and used. This family of models have been used successfully to infer causal impact of marketing campaigns [16], select variables and nowcast consumer sentiment [17], or for other economic time series models like unemployment [10].

In this paper we use the formalism of Bayesian structural time-series models to formulate a robust model that links advertising expenditures with weekly sales. Due to the flexibility and modularity of the model, it will be well suited to generalization to various markets or situations. Its Bayesian nature also adapts smoothly to the issue of introducing outside or a priori information (the manager’s views), which can be updated according with the posterior distribution of the estimated parameters. This aspect will be used to present an optimal budget allocation strategy across time and channels. Last, the formulation of the model allows for non-gaussian innovations of the process, which will take care of the heavy-tailedness of the distribution of sales increments.

The paper is organized as follows: after a brief review of the most usual marketing models and the formalism of the structural time series in section 2, we define the model to be used to fit the data. The experimental setup will be detailed in section 3. Section 4 provides a discussion in which alternative models will be compared and also possible uses of this model in the industry will be sketched. A brief summary and ideas for further research will be detailed in section 5.
2 Theoretical background and model definition

2.1 The Nerlove-Arrow model

Numerous formulations of aggregate advertising response models exist in the marketing literature, e.g. [7]. The model of Nerlove and Arrow [6] is parsimonious and is considered as a standard in the quantitative marketing community. We shall use it as a starting point.

In this model, advertising expenditures are considered to affect the present and future character of output and, hence, the present and future net revenue of the investing firm. The idea is to define an “advertising stock” called goodwill \( A(t) \) which seemingly summarizes the effects of current and past advertising expenditures over demand. Then, the following dynamics is defined for the goodwill.

\[
\frac{dA}{dt} = qu(t) - \delta A(t),
\]

where \( u(t) \) is the advertising spending rate (e.g., dollars or gross rating points (GRPs) per week), \( q \) is a parameter that reflects the advertising quality (effectiveness coefficient) and \( \delta \) is the decay or forgetting rate. The goodwill then increases linearly with the advertisement expenditure but decreases also linearly due to forgetting.

Several extensions and modifications have been proposed to this simple model: it can include a limit for potential costumers [8], a non-linear response function to advertise expenditures [9], wear-in and wear-out effects of advertising [18], interactions between different advertising channels [19], etc. Still, for most of the tasks, the Nerlove-Arrow model remains as a simple and solid starting point.

2.2 Bayesian structural time series models

Structural time series models or state-space models provide a general formulation that allows a unified treatment of virtually any linear time series model through the general algorithms of the Kalman filter and the associated smoother. Several handbooks [20, 21, 13, 22] discuss this topic in depth, so we will not develop the theory further; we will however present the most salient features that concern our modelling problem. For further details, the reader is encouraged to check the aforementioned handbooks.

The state-space formulation of a time series consists of two different equations: the state or evolution equation which determines the dynamics of the state of the system as a first-order Markov process — usually parametrized through hidden or latent variables — and an observation or measurement equation which links the latent state with the observed state. Both equations are also affected by uncontrollable noise. Denoting by \( \theta_t \) the \( m \times 1 \) state vector describing the inner state of the system, by \( G_t \) the \( m \times m \) matrix that generates the dynamics, by \( H_t \) a \( m \times g \) matrix and by \( \epsilon_t \) a \( g \times 1 \) vector of serially uncorrelated disturbances with mean zero and covariance matrix \( W_t \), the system would evolve according to equation...
\[ \theta_t = G_t \theta_{t-1} + H_t \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, W_t). \]

These states \((\theta_t)\) are not generally observable, but are linked to the observation variables \(Y_t\) through the observation equation:

\[ Y_t = F_t \theta_t + \epsilon'_t \quad \epsilon'_t \sim \mathcal{N}(0, V_t), \]

where:

- \(Y_t \in \mathbb{R}\) is the observed value at timestep \(t\).
- \(F_t\) is the \(1 \times m\) matrix that links the inner state to the observable.
- \(V_t \in \mathbb{R}^+\) is the variance of \(\epsilon'_t\), the random disturbances of the observations.

The specification of the state-space system is completed by assuming that the initial state vector \(\theta_0\) has mean \(\mu_0\) and a covariance matrix \(\Sigma_0\) and it is uncorrelated with the noise. The problem then consists of estimating the sequence of states \(\{\theta_1, \theta_2, \ldots\}\) for a given series of observations \(\{y_1, y_2, \ldots\}\) and whichever other structural parameters of the transition and observation matrices. State estimation is readily performed via the Kalman filter; different alternatives however arise when structural parameters are unknown. In the classical setting, these are estimated using maximum likelihood. In the Bayesian approach, the probability distribution about the unknown parameters is updated via Bayes Theorem. If exact computation through conjugate priors is not possible, the probability distributions before each measurement are updated by an approximate procedure such as Markov chain Monte Carlo (MCMC) \([10]\).

### 2.3 Model specification

#### 2.3.1 State-space formulation

The continuous-time Nerlove-Arrow model must be first cast in discrete time so as to formulate our model in state-space. From equation (1), we get:

\[ A_t = (1 - \delta) A_{t-1} + qu_{t-1} + \epsilon_t \]

where \(A_t\) is the goodwill stock, \(u_t\) is the advertising spending rate, \(q\) is the effectiveness coefficient and the random disturbance \(\epsilon_t\) captures the net effects of the variables that affect goodwill that cannot be modelled explicitly. This discrete counterpart of Nerlove-Arrow is a distributed-lag structure with geometrically declining weights, i.e., a Koyck model \([23, 24]\). Since in our setting the model includes the effect of \(k\) different channels in the goodwill, we shall modify the previous equations accordingly to:
\[ A_t = (1 - \delta)A_{t-1} + \sum_{i=1}^{k} q_i u_i(t-1) + \epsilon_t \]

Following the notation from equations (2) and (3), the discrete-time Nerlove-Arrow model in state-space form reads as follows:

**Evolution equation:**

\[ \theta_t = G_t \theta_{t-1} + H_t \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, W_t), \]  

(4)

where

\[
\theta_t = \begin{bmatrix} A_t \\ q_1 \\ \vdots \\ q_k \end{bmatrix}, \quad G_t = \begin{bmatrix} (1 - \delta) & u_1(t-1) & \cdots & u_k(t-1) \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad H_t = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\]

Note that the \( q_i \) are constant over time and that the matrix \( G_t \) depends on the known inversion levels at timestep \( t - 1 \) and on an unknown parameter (\( \delta \)) to be estimated from the data.

**Observation equation:**

\[ Y_t = F_t \theta_t + \epsilon'_t \quad \epsilon'_t \sim \mathcal{N}(0, V_t) \]  

(5)

where \( Y_t \) are the observed sales at time \( t \) and \( F_t = [1, 0, \ldots, 0]^T \).

### 2.4 Modularity and additional structure

These models are very flexible in the sense that they can be defined *modularly*, inasmuch as the different hidden states evolve independently of the others (*i.e.* the evolution matrix can be cast in block-diagonal form). This greatly simplifies their implementation and allows for simple building-blocks with characteristic behavior. Typical blocks specify trend and seasonal components — which can be helpful to discover additional patterns in the time series — or explanatory variables that can be added to further reduce the uncertainty in the model and bridge the gap between time series and regression models. Via the *superposition principle* [25] we could include additional blocks in our model:

\[ Y_t = Y_{NA,t} + Y_{R,t} + Y_{T,t} + Y_{S,t} \]

where \( Y_{NA,t} \) corresponds to the discretized Nerlove-Arrow equation, defined in (4) and (5), \( Y_{R,t} \) is a regression component, containing the effects of external explanatory variables \( X_t \), \( Y_{T,t} \) is a trend component or a simpler local level component, and \( Y_{S,t} \) is a seasonal component.
2.4.1 Models for trend

The simplest model is the local level model

\[ Y_{T,t} = \mu_t + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2_{obs}) \]
\[ \mu_t = \mu_{t-1} + \epsilon'_t \quad \epsilon'_t \sim \mathcal{N}(0, \sigma^2_{level}) \]

where a fully Bayesian approach could be adopted by incorporating prior distributions over the variances, allowing for their estimation from the data.

2.4.2 Models for seasonality

Adding a seasonal component is crucial in econometric series, as this kind of data typically exhibits seasonal behavior (e.g., sales skyrocket during the Christmas’ campaign). Thus, we can specify the seasonal component via:

\[ Y_{S,t} = F\theta_t + v_t \]
\[ \theta_t = G\theta_{t-1} + w_t \]

where \( F \) is a \( p \)-dimensional vector and \( G \) is a \( p \times p \) matrix such that

\[
G = \begin{bmatrix}
0 & 0 & \ldots & 0 & 1 \\
1 & 0 & \ldots & 0 & 0 \\
0 & 1 & \ldots & 0 & 0 \\
\vdots & & & & \\
0 & 0 & \ldots & 1 & 0
\end{bmatrix}
\]

and \( F = (1, 0, \ldots, 0, 0) \).

In order to establish the seasonal period \( p \) of the model, we can choose the one that maximizes the loglikelihood, or minimizes the root mean squared error (RMSE) over a validation set. In our dataset, we found that the optimal value is \( p = 52 \) since we are dealing with weekly data for a period of several years.

2.4.3 Regression components. Spike and slab variable selection

To take into account the effects of external explanatory variables, a static regression component can be easily incorporated into the model

\[ Y_{R,t} = X_t\beta + \epsilon_t, \]

where we let the state \( \beta \) be constant in time to get a more parsimonious model. A spike and slab prior [26] is used for the static regression component since it can incorporate prior
information into the model and also facilitate variable selection. This is specially useful for models with large number of regressors, a setting typically encountered in business scenarios.

Let $\gamma$ denote a binary vector that indicates whether or not a particular regressor is included in a regression. More precisely, $\gamma_i = 1$ if and only if $\beta_i \neq 0$. The subset of $\beta$ for which $\gamma_i = 1$ will be denoted as $\beta_\gamma$, and let $\sigma^2$ be the residual variance from the regression part. The spike and slab prior can be expressed as

$$p(\beta, \gamma, \sigma^2) = p(\beta_\gamma | \gamma, \sigma^2)p(\sigma^2 | \gamma)p(\gamma).$$

A usual choice for the $\gamma$ prior is a product of Bernoulli distributions:

$$\gamma \sim \prod_i \pi_i^{\gamma_i}(1 - \pi_i)^{1 - \gamma_i}.$$  

The manager of the firm may elicit these $\pi_i$ in various ways. A reasonable choice when detailed prior information is unavailable is to set all $\pi_i = \pi$. Then, we may specify an expected number of non-zero coefficients by setting $\pi = k/p$, where $p$ is the total number of regressors. Another possibility is to set $\pi_i = 1$ if the manager believes that the $i$-th regressor is crucial for the model. For instance, the manager may impose her business knowledge and experience from past campaigns by setting $\pi_i = 1$ for the key variables. For further information regarding the spike and slab prior specification see [27].

2.5 Model estimation and forecasting

Model parameters can be estimated using Markov Chain Monte Carlo simulation, as described in Chapter 4 of [21] or more specifically in [10]. We follow the same scheme. Let $\Theta$ be the set of model parameters other than $\beta$ and $\sigma^2$. The posterior distribution can be simulated with the following algorithm:

1. Simulate $\theta \sim p(\theta | y, \Theta, \beta, \sigma^2).$
2. Simulate $\Theta \sim p(\Theta | y, \theta, \beta, \sigma^2).$
3. Simulate $\beta, \sigma^2 \sim p(\beta, \sigma^2 | y, \theta, \Theta).$

Repeatedly iterating the above steps gives a sequence of draws $\rho^{(1)}, \rho^{(2)}, \ldots, \rho^{(K)}$ from $p(\Theta, \beta, \sigma^2, \theta)$. In our experiments, we set $K = 4000$ and discard the first 1000 draws to avoid burn-in issues.

In order to sample from the predictive distribution, we follow the usual Bayesian approach summarized by the following predictive equation, in which $y_{1:t}$ denotes the sequence of observed values, and $\bar{y}$ denotes the set of values to be forecast.
\[ p(\tilde{y}|y_{1:t}) = \int p(\tilde{y}|\rho)p(\rho|y_{1:t})d\rho. \]

Thus, it is sufficient to sample from \( p(\tilde{y}|\rho^{(i)}) \), which can be achieved by iterating equations (3) and (2). With these predictive samples \( \tilde{y}^{(i)} \) we can compute statistics of interest regarding the predictive distribution \( p(\tilde{y}|y_{1:t}) \) such as the mean or variance (Monte Carlo estimates of \( E[\tilde{y}|y_{1:t}] \) and \( Var[\tilde{y}|y_{1:t}] \), respectively) or quantiles of interest.

### 2.5.1 Robustness

We can replace the assumption of Gaussian errors with student-\( t \) errors in the observation equation, thus leading to the model

\[ Y_t = F_t\theta_t + \epsilon_t, \quad \epsilon_t \sim T_\nu(0,\tau^2). \]

Typically, in these settings we can set the parameter \( \nu > 1 \) to make the variance finite, and this variance parameter can be estimated from the data using suitable priors. In this manner, we allow the model to predict occasional larger deviations, which is reasonable in the context of forecasting sales. For instance, a special event not taken into account through the predictor variables may lead to an increase in the sales for that week.

### 2.6 Decision support system

The model we propose can be used as a decision support system for the manager, helping her in adopting the investment strategy in advertising. The optimization problem for one-step-ahead forecasts can be formulated as the non linear problem

\[
\begin{align*}
\text{maximize} & \quad E[\tilde{y}_{t+1}|y_{1:t}, x_{t+1}, u_{t+1}] \\
\text{subject to} & \quad \sum_{i=1}^{k} u_{(t+1),i} \leq b_{t+1} \\
& \quad Var[\tilde{y}_{t+1}|y_{1:t}, x_{t+1}, u_{t+1}] \leq \sigma^2,
\end{align*}
\]

where \( b_t \) is the total advertising budget for week \( t \) and \( \sigma^2 \) is a parameter that controls the risk and must be elicited by the manager. We have made explicit the dependence on the regressor variables \( x_t \) and advertisement investments \( u_t \) in the mean and variance expressions.

Expected predicted sales and the above variance are straightforward to compute through Monte Carlo simulation, see Section 2.5. Note that due to the nature of the state-space model, it is straightforward to extend the previous optimization problem over several weeks.
timesteps. In Section 3.3.1 we discuss some possible approaches regarding the previous optimization problem.

3 Experimental setup and results

3.1 Data analysis

Figure 1: Total weekly sales. Jan-2011 to Jun-2015

The time series analyzed in this study contains the total weekly sales of a country-wide franchise of fast food restaurants, Figure 1. It covers the period starting in January 2011 up to June 2015, comprising a total of 234 observations. The total weekly sales is in fact the aggregated sum from the individual sales of the whole country network of 426 franchises, also allowing a fine-grained study down to the store level which however will not be carried out in this paper. Along with the sales figures, the series includes the investment levels in advertising during this period \{u_{it}\} for seven different channels \textit{viz.} OOH (Out-of-home, \textit{i.e.} billboards), Radio, TV, Online, Search, Press and Cinema, Figure 2.

A handful of other predictors $X_t$ which are also known to affect sales will be used in the model, all of them weekly sampled:

- Global economic indicators: unemployment rate (Unemp$\_IX$), price index (Price$\_IX$) and consumer confidence index (CC$\_IX$).
- Climate data: average weekly temperature along the country (AVG$\_Temp$) and weekly rainfall (AVG$\_Rain$).
Figure 2: Advertising expenditure. Jan-2011 through Jul-2015. Note the different scales on the y-axis.

- Special events: holidays (Hols) and sporting events (Sport_EV).

From such graphs we observe that:

- In Figure 1 the series has a characteristic seasonal pattern that shows peaks in sales coinciding with Christmas, Easter and summer vacations.

- In Figure 2 we observe that the investment levels at each channel vary largely in scale, with investments in TV, and OOH dominating the other channels.

- The investment strategies adopted by the firm at each channel are also qualitatively different, some of them showing spikes while others depicting a relatively even spread investment across time.
3.2 Experimental setup

Following the notation in Section 2.3, we consider three model variants for the particular dataset in increasing order of complexity:

- **Baseline model**, which makes no use of external variables
  \[ Y_t^B = Y_{NA,t} + Y_{T,t}. \]

- **Auto regression**: this model incorporates the external ambient and investment variables, so the equation of the model becomes
  \[ Y_t^{RA} = Y_{NA,t} + Y_{T,t} + Y_{R,t}. \]
  We select an expected model size of 5 in the spike and slab prior, letting all variables to be treated equally.

- **Regression (forcing)**: the model has same equation as before:
  \[ Y_t^{RF} = Y_{NA,t} + Y_{T,t} + Y_{R,t}. \]
  However, in the prior we force investment variables to be used by setting their corresponding \( \pi_i = 1 \), and imposing an expected model size of 5 for the rest of the variables.

In all cases, only the five principal advertising channels (i.e. TV, OOH, ONLINE, SEARCH and RADIO) will be used; the remaining two (CINEMA and PRESS) are sensibly lower both in magnitude and frequency than the others so we can safely disregard them in a first approximation. The models were implemented in R using the bsts package [15].

As customary in supervised learning setting with time series data, we perform the following split of our dataset: since it comprises four years of sales, we take the first two years of observations as training set, and the rest as a holdout set, in which we measure several predictive performance criteria. Before fitting the data, we scale the series to have zero mean and unit variance as thus increases MCMC stability. Reported sales forecasts are transformed back to the original scale for easy interpretation.

3.3 Discussion of the results

It is crucial that the models achieve good predictive performance. Otherwise, the proposed framework would be unusable in a real-world scenario such as the one described in the previous Section. For this reason, we test the predictive performance of our three models using the following two well-known metrics:
• Mean Percentage Error:

$$\text{MPE} = 100\% \frac{1}{T} \sum_{t=1}^{T} \frac{y_t - \hat{y}_t}{y_t},$$

where $y_t$ denotes the actual value, $\hat{y}_t$ the mean one-step-ahead prediction and $T$ is the length of the hold-out period.

• Mean Absolute Percentage Error:

$$\text{MAPE} = 100\% \frac{1}{T} \sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_t}{y_t} \right|,$$

These scores are reported in Table 1. Note that, unsurprisingly, the models which include external information (RA and RF) achieve better accuracy than the baseline. In addition, incorporating external information helps unbiasedness, as $\text{MPE}_{\text{RA}}$ and $\text{MPE}_{\text{RF}}$ are closer to zero than the baseline counterpart. Overall, we found the predictive performance of our models to be adequate to the manager’s requirements, as we achieve under 5% relative absolute error using the variants augmented with external information. This is clearly useful for the decision maker who may forecast their weekly sales one week ahead to within 5% error in the estimation.

| Model | MPE  | MAPE |
|-------|------|------|
| B     | -0.44% | 5.85% |
| RA    | -0.07% | 4.62% |
| RF    | 0.11%  | 4.59% |

Table 1: Accuracy measures for each model variant.

Figure 3 displays the predictive ability of model RF during the hold-out period. Note that the model is sufficiently flexible to adapt to fluctuations such as the peaks at Christmas. Predictive intervals also adjust their width with respect the time in order to reflect varying uncertainty, yet in the worst cases they are sufficiently narrow. In the Figure, 95% predictive intervals are depicted in light gray.

Further information can be tracked in Figure 4, where mean standardized residuals are plotted for each model variant. Notice that the residuals for models RA and RF are roughly comparable, being both sensibly smaller than that of the baseline counterpart. This means that the simpler Nerlove-Arrow model is biased and benefits from the addition of the ambient variables $X_i$, as suggested in our findings from Table 1.
Having built good predictive models, we shall turn to inspect them more closely, with the aim of performing valuable inference for the business setting. The average estimated parameter and expected standard deviations of the $q_i$ coefficients for the different advertising channels are displayed in Table 2. We consider coefficient $c$ to be statistically significant if $0 \not\in [\mu_c - 2\sigma_c, \mu_c + 2\sigma_c]$. We also show the weights of the ambient variables $X_i$ for the augmented models RA and RF in Table 3, as well as the probability of a variable being selected in the MCMC simulation for a given model, which is shown in Figure 5. The color code shows variables positively (white) or negatively (black) correlated with sales.
| Channel | Model B | Model RA | Model RF |
|---------|---------|----------|----------|
|         | mean    | sd       | mean     | sd       | mean     | sd       |
| y_AR    | 8.00e-01 | 5.10e-02 | 5.17e-01 | 4.50e-02 | 5.07e-01 | 4.55e-02 |
| OOH     | 9.30e-03 | 3.11e-02 | 1.15e-03 | 9.54e-03 | 6.10e-02 | 3.54e-02 |
| ONLINE  | 1.10e-03 | 9.25e-03 | 1.95e-03 | 1.23e-02 | 9.04e-02 | 4.01e-02 |
| RADIO   | -5.34e-04 | 6.80e-03 | -2.61e-05 | 2.30e-03 | -2.57e-02 | 3.42e-02 |
| TV      | -2.85e-04 | 5.12e-03 | -5.53e-05 | 3.71e-03 | -6.14e-02 | 4.21e-02 |
| SEARCH  | 1.51e-05 | 2.74e-03 | 8.58e-05 | 2.49e-03 | 1.38e-02 | 3.50e-02 |

Table 2: Expected value and standard error of the coefficients $q_i$. Statistically significant coefficients appear in bold.

| Channel       | Model RA | Model RF |
|---------------|----------|----------|
|               | mean     | sd       | mean     | sd       |
| Sport_EV      | -2.06e-01 | 3.80e-02 | -2.00e-01 | 3.71e-02 |
| AVG_Temp      | 2.57e-01 | 4.43e-02 | 2.43e-01 | 4.65e-02 |
| Hols          | 3.10e-01 | 3.70e-02 | 3.15e-01 | 3.72e-02 |
| AVG_Rain      | -2.86e-02 | 5.01e-02 | -1.96e-02 | 4.14e-02 |
| Price IX      | 1.38e-03 | 1.15e-02 | 3.60e-03 | 1.92e-02 |
| Unemp IX      | 6.34e-05 | 3.32e-03 | 2.36e-04 | 7.79e-03 |
| CC IX         | 6.27e-07 | 2.94e-03 | 2.85e-04 | 5.50e-03 |

Table 3: Expected value and standard error of coefficients $\beta_i$. Statistically significant coefficients are in bold.
Looking at the ambient variables, the following comments seem in order:

- From Table 3 we see that the socio-economic indicators (unemployment rate, inflation and consumer confidence) do not seem statistically relevant for the problem at hand.

- Looking at the sign of the coefficients of the most significant regressors $X_i$ (Hols, Sport_EV, AVG_Temp and AVG_Rain) we see that they are as we would naturally expect. Moreover, their absolute value is well above the error in both models RA and RF, a strong indicator of their influence in the expected weekly sales (cf. Table 3, Figure 5).

- We see, for instance, that sporting events are negatively correlated with sales. The restaurant chain in this study has no TV broadcasting in their facilities, so customers probably choose alternative places to spend their time on a day when Sport_EV = 1.

- The sign of Hols and AVG_Temp is positive, showing strong evidence for the fact that sales increase in periods of the year where potential customers have greater amounts of leisure time, i.e., during national holidays or the summer vacation.

- AVG_Rain is not statistically significant. A possible explanation is that AVG_Rain records average rainfall over a large country.

Next, we turn our attention to the investment variables across different advertising channels.
• The advertising channels $u_i$ are almost never selected in model RA, and their $q_i$ coefficients are not significantly higher than their errors to be considered influential in the model. In model RF, however, there is strong evidence that their effect is more than a random fluctuation (cf. Table 2, Figure 5).

• The negative sign in both RADIO and TV in all three models suggests that (at least locally) part of the expenditures in these two channels should be diverted towards other channels with positive sign on their $q_i$ coefficients, specially to the channel with the strongest positive coefficient (ONLINE and OOH).

• It is quite interesting that the trend that shows the year-to-year advertising budget of this firm (cf. Figure 6) has a significant reduction in TV expenditures and a big increase in OOH. RADIO however is not reduced accordingly but increased, and ONLINE — which our model considers the best local inversion alternative — follows the inverse path.

• The autoregressive term is close to 0.5, which means that the immediate effect of advertising is roughly half to the long run accumulated effects.

3.3.1 Budget allocation. Model-based solutions

Regarding the optimization problem described in Section 2.6 a possible alternative would be to rewrite the objective function as

$$E[\hat{y}_{t+1}|y_{1:t}, x_{t+1}, u_{t+1}] - \lambda \sqrt{Var[\hat{y}_{t+1}|y_{1:t}, x_{t+1}, u_{t+1}]}$$
and assume that the predictive samples $\bar{y}_{t+1}^{(i)}$ are normally distributed so that the previous expression may be regarded as a lower quantile.

The above optimization problem may not need to be solved by exhaustively searching over the space of possible channel investments. In a typical business setting, the manager considers a discrete set of $S$ investment strategies, each one defined by values for the coefficients $q_i$, so she may perform $S$ simulations of the predictive distribution and choose the strategy that optimizes the previous criterion. For instance, in Figure 7 two strategies for a one-week ahead period are compared. The corresponding two predictive distributions are plotted. Whereas the New plan may achieve higher mean expected sales than the Base plan, the former one may incur in higher risk, c.f. the dotted vertical lines depicting lower quantiles, as predicted by our model. Thus, by eliciting the $\lambda$ parameter the manager may discard strategies that are predicted as too risky.

![Figure 7: Predictive distributions under two strategies for a given one-step ahead forecast.](image)

4 Summary and future work

We have developed a data-driven approach for the management of advertising investments of a firm. First, using the firm’s investment levels in advertising, we propose a formulation of the Nerlove-Arrow model via a Bayesian structural time series to predict an economic variable (global sales) which also incorporates information from the external environment (climate, economical situation and special events). The model thus defined offers low predictive errors while maintaining interpretability and can be built in a modular fashion, which offers great flexibility to adapt it to other business scenarios. The model performs variable selection and allows to incorporate priori information via the spike-and-slab prior. It can
handle non-gaussian deviations and also provides hints to which of the advertising channels are having positive effects upon sales. The model can be used as a decision support system by the manager of the firm, helping with the task of allocating ad investments.

Possible extensions of this model include:

1. Use a different model to explain the influence of advertising upon sales. This model could take into account interactions between the channels or allow for different long-term effects for each of them.

2. Develop a model-based strategy for long-run temporal and cross-sectional budget allocation.

3. Model each restaurant individually instead of using total aggregated sales. For this approach to be sound, it would entail to use both local values of the ambient variables $X_i$ and some of the investment channels $u_i$ (e.g. OOH).

4. Include in the model the effect of special discounts, promotions and coupons, since these are probably highly correlated with some of the channels, specially ONLINE and SEARCH.

Acknowledgements

The authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, through the “Severo Ochoa Programme for Centres of Excellence in R&D” (SEV-2015-0554). V.G. acknowledges support for grant FPU16/05034. The research of D.G-U is supported in part by Spanish MINECO-FEDER Grant MTM2015-65888-C4-3. The authors also thank the members of the SPOR-Datalab group at ICMAT for their suggestions and support.

References

[1] Gert Assmus, John U. Farley, and Donald R. Lehmann. How advertising affects sales: Meta-analysis of econometric results. *Journal of Marketing Research*, pages 65–74, 1984.

[2] Gerard J. Tellis, GJ Tellis, and T. Ambler. Advertising effectiveness in contemporary markets. *The SAGE Handbook of Advertising*, page 264, 2007.

[3] Xueming Luo and Pieter J. de Jong. Does advertising spending really work? the intermediate role of analysts in the impact of advertising on firm value. *Journal of the Academy of Marketing Science*, 40(4):605–624, 2012.
[4] Thorsten Wiesel, Koen Pauwels, and Joep Arts. Practice prize paper—marketing’s
profit impact: quantifying online and off-line funnel progression. *Marketing Science*,
30(4):604–611, 2011.

[5] Gerard J Tellis. Generalizations about advertising effectiveness in markets. *Journal of
Advertising Research*, 49(2):240–245, 2009.

[6] Marc Nerlove and Kenneth J Arrow. Optimal advertising policy under dynamic condi-
tions. *Econometrica*, pages 129–142, 1962.

[7] John DC Little. Aggregate advertising models: The state of the art. *Operations
Research*, 27(4):629–667, 1979.

[8] ML Vidale and HB Wolfe. An operations-research study of sales response to advertising.
*Operations research*, 5(3):370–381, 1957.

[9] John DC Little. Brandaid: A marketing-mix model, part 1: Structure. *Operations
Research*, 23(4):628–655, 1975.

[10] Steven L Scott and Hal R Varian. Predicting the present with bayesian structural time
series. *International Journal of Mathematical Modelling and Numerical Optimisation*,
5(1-2):4–23, 2014.

[11] Norbert Wiener, Norbert Wiener, Cyberneticist Mathematician, Norbert Wiener, Nor-
bert Wiener, and Cybernéticien Mathématicien. Extrapolation, interpolation, and
smoothing of stationary time series: with engineering applications. 1949.

[12] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems.
*Journal of Basic Engineering*, 82(1):35–45, 1960.

[13] Simo Särkkä. *Bayesian Filtering and Smoothing*, volume 3. Cambridge University
Press, 2013.

[14] Giovanni Petris and R An. An r package for dynamic linear models. *Journal of
Statistical Software*, 36(12):1–16, 2010.

[15] Steven L Scott. bsts: Bayesian structural time series, 2016.

[16] Kay H Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, Steven L Scott, et al.
Inferring causal impact using bayesian structural time-series models. *The Annals of
Applied Statistics*, 9(1):247–274, 2015.

[17] Steven L Scott and Hal R Varian. Bayesian variable selection for nowcasting economic
time series. In *Economic Analysis of the Digital Economy*, pages 119–135. University
of Chicago Press, 2015.
[18] Prasad A Naik, Murali K Mantrala, and Alan G Sawyer. Planning media schedules in the presence of dynamic advertising quality. *Marketing science*, 17(3):214–235, 1998.

[19] Frank M Bass, Norris Bruce, Sumit Majumdar, and BPS Murthi. Wearout effects of different advertising themes: A dynamic bayesian model of the advertising-sales relationship. *Marketing Science*, 26(2):179–195, 2007.

[20] James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods*, volume 38. OUP Oxford, 2012.

[21] Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. Dynamic linear models. In *Dynamic Linear Models with R*, pages 31–84. Springer, 2009.

[22] Mike West and Jeff Harrison. Bayesian forecasting and dynamic models (2nd edn). *Journal of the Operational Research Society*, 49(2):179–179, 1998.

[23] Darral G Clarke. Econometric measurement of the duration of advertising effect on sales. *Journal of Marketing Research*, pages 345–357, 1976.

[24] Leendert Marinus Koyck. *Distributed Lags and Investment Analysis*, volume 4. North-Holland Publishing Company, 1954.

[25] Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. Model specification. In *Dynamic Linear Models with R*, pages 85–84. Springer, 2009.

[26] Toby J Mitchell and John J Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

[27] Edward I George and Robert E McCulloch. Approaches for bayesian variable selection. *Statistica Sinica*, pages 339–373, 1997.