

# On the identification of sales forecasting models in the presence of promotions

Juan R. Trapero<sup>a,\*</sup>, N. Kourentzes<sup>b</sup>, R. Fildes<sup>b</sup>

<sup>a</sup>*Universidad de Castilla-La Mancha*

*Departamento de Administracion de Empresas, Ciudad Real 13071, Spain*

<sup>b</sup>*Lancaster University Management School*

*Department of Management Science, Lancaster LA1 4YX, UK*

---

## Abstract

Shorter product life cycles and aggressive marketing, among other factors, have increased the complexity of sales forecasting. Forecasts are often produced using a Forecasting Support System that integrates univariate statistical forecasting with managerial judgment. Forecasting sales under promotional activity is one of the main reasons to use expert judgment. Alternatively, one can replace expert adjustments by regression models whose exogenous inputs are promotion features (price, display, etc.). However, these regression models may have large dimensionality as well as multicollinearity issues. We propose a novel promotional model that overcomes these limitations. It combines Principal Component Analysis to reduce the dimensionality of the problem and automatically identifies the demand dynamics. For items with limited history, the proposed model is capable of providing promotional forecasts by selectively pooling information across established products. The performance of the model is compared against forecasts provided by experts and statistical benchmarks, on weekly data; outperforming both substantially.

*Keywords:* Promotional modeling, demand forecasting, judgmental adjustments, principal components analysis

---

\*Corresponding author.

*Email addresses:* `juanramon.trapero@uclm.es` (Juan R. Trapero),  
`n.kourentzes@lancaster.ac.uk` (N. Kourentzes), `r.fildes@lancaster.ac.uk` (R. Fildes)

*Preprint submitted to Journal of the Operational Research Society*    *September 19, 2013*

## 1. Introduction

Manufacturing firms are fundamental in supporting most modern economies. These companies face an increasingly competitive environment, due to globalization. Improvements in supply chain management can lead to competitive advantage. In this context, Stock-Keeping Unit (SKU) demand forecasting is of paramount importance to reduce inventory investment, enhance customer satisfaction and improve distribution operations.

Demand forecasting at SKU level relies on a particular type of Decision Support System, known as a Forecasting Support System (FSS) (Fildes et al., 2006; Caniato et al., 2011). An FSS often integrates univariate statistical forecasting method with managerial judgment from forecasters in the organization. The first provides an automatic baseline statistical forecast, while the later are responsible for manually adjusting the baseline in order to include market forces, promotions and other influences.

Univariate forecasting methods are based on time series techniques that analyze past history sales in order to extract a demand pattern that is projected into the future (Ord and Fildes, 2012). An example of such techniques is the exponential smoothing family of methods (Gardner, 2006; Taylor, 2011). This technique is very well suited to companies that handle numerous SKUs and must produce forecasts semi-automatically. However, when a company has to face the impact of future promotions on sales, such univariate models cannot account for it. Therefore, experts have to modify the forecasts and include the expected impact on sales due to promotions. Fildes and Goodwin (2007) indicated promotional and advertising activity as one of the main drivers behind judgmental adjustments of statistical forecasts. In particular, Trapero et al. (2013) study the performance of judgmentally adjusted forecasts in the presence of promotions. They found evidence that experts might improve the baseline forecasts but not systematically. In addition, they suggest that past promotional information modeled adequately may bring important forecast error reductions with respect to judgmental forecasts.

An alternative approach to the problem of forecasting promotional sales is to use regression models, which use past promotional information to formulate causal models, Fildes et al. (2008). These models are usually based on multiple regression models whose exogenous inputs correspond to promotion features (price discounts, type of display, type of advertising, etc.), (Özden Gür Ali et al., 2009; Cooper et al., 1999). Considering technology advances

in computing and data acquisition, most of the recent Promotional Support Systems (PSS) have been based on forms of sophisticated regression, such as SCAN\*PRO by Leeflang et al. (2002), PromoCast<sup>TM</sup> by Cooper et al. (1999), and CHAN4CAST by Divakar et al. (2005) providing promising results. These PSS relate the sales of units with independent variables such as regular prices, price cuts, feature advertising, special displays, and other relevant inputs depending on the industry<sup>1</sup> under study. These are typically estimated at brand level and there is very limited work on SKU level promotional modelling, which is the focus of this paper.

Such PSSs require a high number of explanatory variables; for instance, Cooper et al. (1999) introduced 67 variables to their model. These demanding data requirements have important drawbacks: i) the cost of acquiring and maintaining the necessary information, if it is feasible to collect all data; ii) variable selection is required, due to the number of exogenous variables considered during model building; for example, CHAN4CAST utilizes a stepwise selection method; iii) it is possible that variables may be multicollinear; iv) some models require past promotions for each SKU and therefore are unable to provide promotional forecasts when there is limited or no past promotions in the SKU demand history; and v) such models do not deal with any potential serial correlation in the error term. Another critical feature is the need for in-company expertise to handle complex econometric promotional models. Hughes (2001) identified the lack of expertise and resources as two key limitations in the implementation of complex forecasting techniques in organizations, which may explain the lack of widespread diffusion of such models. Fildes and Goodwin (2007) and Lawrence et al. (2000) suggested that the reliance on expert interventions is more widespread.

The aim of this article is twofold: On the one hand, we propose a dynamic regression model that overcomes the technical issues discussed above, which limit the use of statistical models for forecasting promotional sales at SKU level. On the other hand, we investigate the relative performance of quantitative and judgmental approaches for forecasting promotional sales. We provide insight in this crucial question for business practice by comparing the proposed model with forecasts provided by human experts. Better understanding of the benefits achieved by quantitative modeling may encourage

---

<sup>1</sup>For instance, the temperature is a significant independent variable in the soft-drink industry as it is discussed by Divakar et al. (2005)

the use of such models by practitioners.

The proposed dynamic regression model has the following innovations: i) Principal Component Analysis is used to solve the problems of dimensionality reduction and multicollinearity, which are fairly common in practice due to simultaneous promotional activities, and; ii) if for some SKUs there is no prior promotional information, the model pools past information from other SKUs to produce a forecast. Our approach differs to previous implementations of pooled regression by Divakar et al. (2005), in the sense that our algorithm provides forecasts at SKU level and the pooling occurs only when limited or no promotional history is available for a SKU. iii) the dynamics of promotions are modeled through the inclusion of lagged explanatory variables; and finally iv) the noise term is automatically identified by minimizing the Schwartz Information Criterion, improving the forecasting accuracy and capturing demand dynamics that are not explained solely by promotional activity. Thus a strong time series element is introduced in the proposed model. The proposed model offers an automatic forecasting framework, able to produce both baseline and promotional forecasts.

The performance of the proposed model is illustrated using a case study with weekly data from a manufacturing company. Baseline statistical forecasts and expert judgmental forecasts are available to us from the company and are used as benchmarks. It should be noted that, to the best of our knowledge, this is the first case study that investigates the benefits of quantitative promotional models against judgmentally adjusted forecasts. While Trapero et al. (2013) provide a comparison between judgmental promotional forecasts and simple statistical forecasts, they do not explore promotional models capable of capturing different promotional activities that are needed in practice. Furthermore, we employ well established statistical benchmarks, such as the Naïve and single exponential smoothing. We also use the last like promotional model that is typically used as a benchmark in the marketing literature (Trusov et al., 2006). We find that the proposed model can reduce the forecast error of promotional sales up to 30% when compared to the experts' accuracy. Substantial accuracy improvements are observed against statistical benchmarks as well. In summary, this article makes two important contributions to the analysis of promotional forecasting behaviour at SKU level: Firstly, a theoretical model is proposed for overcoming the data complexity and capturing the dynamics inherent in promotions. Secondly, from a practical point of view, this model can automatically enhance SKU level forecasting, taking it beyond the current practice of relying on expert

adjustment of a univariate statistical model. We also rigorously demonstrate the improvements that may be achieved.

This article is organized as follows: Section 2 introduces the case study and carries out an exploratory data analysis. Section 3 introduces the models that will be used to forecast sales. Section 4 reports the experimental design and the results of the evaluation. Finally, main conclusions are drawn in Section 5.

## 2. Case study

Data from a manufacturing company specializing in household products has been collected. The data comprises: i) shipments; ii) one-step-ahead system forecasts; iii) one-step-ahead adjusted or final forecasts provided by experts in the company; and iv) promotional information. The latter includes the following variables: i) Price discount; ii) Feature advertising; iii) Shelf display; iv) Category, where 22 different categories are available; v) Customer (there are two main customers); and vi) Days promoted per week. This variable indicates how many days in a week are under promotion, for example, some weeks only have 3 days promoted.

This case study is focused on continuous demand items, since most of the intermittent demand SKUs are not subject to promotions. In particular, the selected SKUs have a percentage of zeros less than 10%, the sample size is greater than 20 observations and they have at least one week promoted. In total, 60 SKUs have been sampled, with history lengths ranging from 53 to 148 observations, resulting in a total sample of 7,790 observations.

Figure 1 presents four randomly selected time series from our case study. Promotional periods are highlighted with darker background. We can observe that the impact of promotions is very diverse and not always apparent.

### 2.1. Exploratory analysis

Sales, System Forecast (SF) and Final Forecast (FF) are normalized with respect to the sales standard deviation, since SKUs are very heterogeneous in terms of variability. Table 1 provides the Median, as a robust metric of the size, and the Median Absolute Deviation (MdAD) as a dispersion measure of Sales, SF and FF. Each statistic has been split into Promo and No Promo depending whether any promotion is active. The last column of the table shows the Median and the Median Absolute Deviation of the percentage Price Cut. As expected, both sales Median and MdAD are higher when

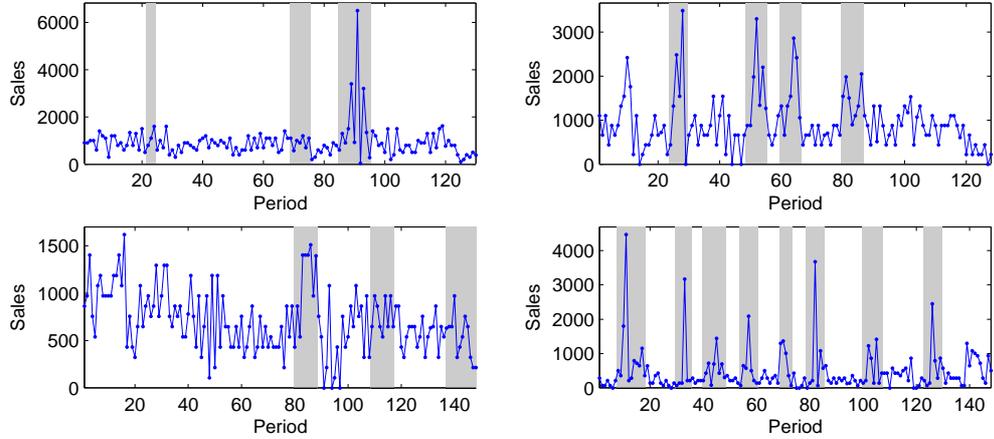


Figure 1: Four example time series. Promotional periods are highlighted by darker background.

		N	Sales	SF	FF	Price Cut
Median	Promo	1187	2.27	1.58	2.53	24%
	No Promo	6603	1.62	1.53	1.64	-
	Overall	7790	1.66	1.54	1.72	-
MdAD	Promo	1187	0.92	0.43	0.99	7%
	No Promo	6603	0.68	0.52	0.60	-
	Overall	7790	0.69	0.50	0.64	-

Table 1: Descriptive statistics

there are promotions. Given that the System Forecast cannot anticipate the presence of promotions, its forecast is biased to lower values. To correct that bias, experts modify the system forecasts providing the FF, whose median is higher than that of sales. According to this table, experts on average overforecast, thus producing an optimistic bias.

In the dataset, the number of promoted weeks is 1187 representing 15.2% of the whole dataset.

### 3. Models

#### 3.1. Naïve and Single Exponential Smoothing

Since the 1950s the use of Exponential Smoothing for forecasting has been the most popular forecasting method used in business and industry (Gardner,

2006; Hyndman et al., 2008). Basically, the Single Exponential Smoothing (SES) updates the previous forecast by weighting the last forecast error, i.e:

$$F_{t+1} = F_t + \alpha(y_t - F_t), \quad (1)$$

where  $\alpha$  is a constant between 0 and 1. This parameter may be set on *a priori* grounds and usually is between 0.05 and 0.3, (Makridakis et al., 1982). However, if an adequate fitting sample is available,  $\alpha$  should be estimated by minimizing the sum of the one-step-ahead in-sample squared forecast errors (Gardner, 1985). The Naïve is also used in this paper, which is essentially a SES with  $\alpha = 1$ , making the forecast always equal to the last known data point ( $y_t$ ); this is the well known Random Walk. Both SES and Naïve are univariate benchmarks that do not use any promotional information in their forecasts. As such, these are used to provide a baseline forecast benchmark.

### 3.2. Last like promotion

To overcome the limitations of the previous benchmarks we introduce a multivariate one. The forecasts change depending on whether there is a promotion or not. A common model that is frequently used to forecast promotional sales is known as the last like promotion. This benchmark model is implemented according to Özden Gür Ali et al. (2009), where an exponential smoothing is adjusted with the last observed promotional lift if there is a promotion in the next observation, i.e:

$$F_{(t+1)} = \begin{cases} M_{(t+1)} & \text{if } promo_t = 0, \\ M_{(t+1)} + L_{(t+1)} & \text{otherwise} \end{cases} \quad (2)$$

and

$$M_{(t+1)} = \begin{cases} (1 - \alpha)M_t + \alpha y_t & \text{if } promo_t = 0, \\ M_t & \text{otherwise,} \end{cases} \quad (3)$$

where  $promo_t$  is a dummy variable indicating whether any kind of promotion is occurring at time  $t$ .  $M_{t+1}$  refers to the smoothed number of units sold up to week  $t$ , based on non-promotion weeks. A constant smoothing value of 0.2 was used. The lift  $L_{(t+1)}$  is calculated as the difference of the actual sales at time  $t$  and the smoothed non-promotion sales at the time of the most recent promotion,  $L_{(t+1)} = y_t - M_t$ .

Note that the only information that we used besides sales, is whether there is a promotion or not. We do not specify any particular features of the promotional campaign that is active.

### 3.3. Dynamic Regression with Principal Components as regressors

In our case, information about promotional variables is available. In particular, price cut, shelf display, feature advertising, customer (2 main customers represented by a dummy variable), days promoted per week and categories (21 dummy variables account for 22 different product categories) are the variables that experts handle when they judgmentally adjust forecasts. In total there are 26 variables. These will be used to predict future promotional sales using quantitative models.

We will discuss the proposed dynamic regression model with principal components first for the promotional observations and then for the non-promotional observations. Consequently both parts will be joined in a final model.

#### 3.3.1. Principal components as regressors

The first option that we can try is to use a multiple linear regression (Leeflang et al., 2002; Divakar et al., 2005). The aim is to predict sales using the available promotional variables:

$$y_t^i = \beta_0^i + \sum_{j=1}^N \beta_j^i \cdot X_{j,t}^i + \epsilon_t^i \quad (4)$$

where  $y_t^i$  stands for sales of SKU  $i$  in week  $t$ ,  $X_j$ ,  $j = 1, \dots, N$  are the set of original promotional variables and  $\epsilon_t^i$  is the error term. In our case study,  $N = 26$  and the set of observations per SKU is  $(t = 1, 2, \dots, 148)$ . However, most promotional variables are multicollinear. When there is a promotion several promotional variables change at the same time, making it difficult to distinguish what is the influence of each causal variable on the sales level. In order to solve this problem we use Principal Component Analysis (PCA).

Principal component analysis generates a new set of variables ( $\hat{X}_j$ ,  $j = 1, \dots, N$ ), called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information and the problem of multicollinearity is lifted (Jolliffe, 2002). Therefore, principal components can be used as regressors instead of the original variables. Here, the problem

is how to select the components to be included in the regression. Essentially, there are two extreme alternatives. One strategy is simply to delete all those components whose variances are less than some cut-off level. In contrast, the opposite strategy is to base selection only on values of t-statistics measuring the contribution of each principal component to the regression equation (Jolliffe, 1982). In our particular case, both alternatives were analyzed and better forecasting results were obtained by selecting components with a high correlation with the dependent variable through a stepwise procedure. In this sense, only a few principal components were included as regressors, such as  $\hat{X}_k$ ,  $k = 1, \dots, M$  with  $M < N$ , reducing the dimensionality of the problem.

$$y_t^i = \bar{\beta}_0^i + \bar{\beta}_1^i \cdot \hat{X}_{1,t}^i + \bar{\beta}_2^i \cdot \hat{X}_{2,t}^i + \dots + \bar{\beta}_M^i \cdot \hat{X}_{M,t}^i + \bar{\epsilon}_t^i \quad (5)$$

PCA is intended to capture the influence of promotional activities on the estimation sample and use that information to predict future promotional sales. This solution assumes that the multicollinearity is structural. The new estimated coefficients may be difficult to interpret. To overcome this, we can transform the regression coefficient estimates back to the original set of variables.

Note that using either the original variables or principal component regressors requires some promotions to have been observed in the estimation sample. If there are no historical promotions no such forecast can be produced, whereas if there is only a limited number of historical promotions the model estimates will be unreliable.

### 3.3.2. Pooled regression

Some SKUs may not have enough samples of sales under promotions in the estimation period. For those cases, an alternative approach is to estimate the effect of promotions on other SKUs by using a pooled regression (Sayrs, 1989), i.e:

$$y_t^i = \alpha_0 + \alpha_1 \cdot X_{1,t}^i + \alpha_2 \cdot X_{2,t}^i + \dots + \alpha_N \cdot X_{N,t}^i + e_t^i \quad (6)$$

Here, the main difference with respect to (4) is that super indexes  $i$  in parameter estimates have been removed because they are estimated across all SKUs. Since the mean and dispersion of different SKUs might be different, the data must be first normalised. Similarly to (5) a few Principal Components can be employed as independent variables in the pooled regression, thus reducing the dimensionality of the problem.

Using pooled PCA regression allows us to forecast promotional effects for SKUs with no or limited history of promotions. This is a substantial improvement over the previous model that is unable to produce reliable promotional forecasts when the SKU's history does not provide enough promotional observations.

If there is enough past promotional data available for a single SKU, the following question arises: should we use a regression for that SKU or a pooled regression considering all other SKUs? To answer that question, a simulation has been carried out to determine the number of past promotional observations required to choose only a single SKU information or the whole set of SKUs. Empirically, it has been determined that when the number of promoted weeks in the estimation sample is less than 5 the pooled regression yields better results. If there are more promotional weeks available for a certain SKU, then we should utilize the corresponding SKUs' past information.

### 3.3.3. Dynamics of regressors

Promotions are often associated with carry-over effects. For instance if a company in week 1 launches a promotion of the kind *buy one get one free* it might see effects on the level of sales in weeks 2 and/or 3. That means that the effect of promotions usually has some dynamics. In order to model this phenomenon, we include lagged variables of the previous 26 regressors up to order 3, summing up to 104 independent variables in the PCA.

### 3.3.4. Error term

In previous equations (4)-(6), the error term ( $e_t^i$ ) may be serially correlated. Another feature of the modeling methodology proposed here is the automatic identification of the error structure by minimizing the Schwarz Bayesian Criterion (SBC), (Schwarz, 1978). In that sense, the error can be expressed in the backshift notation (Box et al., 1994) as:

$$e_t^i = \frac{\theta^i(B)}{(1-B)^d \phi^i(B)} \eta_t^i \quad (7)$$

where  $\eta_t^i$  is white noise,  $B$  is the backshift operator ( $B\eta_t^i = \eta_{t-1}^i$ ),  $\theta^i(B) = (1 - \theta_1^i B - \theta_2^i B^2 - \dots - \theta_q^i B^q)$  is the Moving Average part of order  $q$  for the  $i^{th}$  SKU,  $\phi^i(B) = (1 - \phi_1^i B - \phi_2^i B^2 - \dots - \phi_p^i B^p)$  stands for the Autoregressive term of order  $p$  and  $d$  is the differencing operator order. In order to obtain a parsimonious model, the orders considered for  $p$  and  $q$  range from 0 to 3, and only up to single difference ( $d = 1$ ) has been analyzed.

### 3.3.5. Non promotional observations

The dynamic regression explained so far is used to forecast promotional sales. Nonetheless, a SKU will have observations that are not promoted. In this case, the dynamic regression is simplified to the following equation:

$$y_t^i = \frac{1 + \gamma^i \cdot B}{1 - B} \eta_t^i \quad (8)$$

The transfer function shown in (8) is equivalent to the well-known single exponential smoothing, where  $\gamma = 1 - \alpha$ . Here  $\alpha$  is the exponential smoothing parameter defined in (1). The noise term  $\eta_t^i$  is assumed normally distributed with mean zero and variance  $\sigma^2$ ; both of them constant. It should be noted that the estimation of the parameter  $\gamma^i$  has excluded the observations under promotions.

### 3.3.6. Final model

In summary, the dynamic regression employed to forecasting sales that can be subject to promotional campaigns can be expressed as:

$$F_{(t+1)} = \begin{cases} \beta_0^i + \sum_{j=1}^N \beta_j^i \cdot X_{j,t+1}^i + \frac{\theta^i(B)}{(1-B)^{d\phi^i(B)}} \eta_t^i & \text{if } promo_{t+1} = 1 \text{ and } PromoIn \geq 5, \\ \beta_0 + \sum_{j=1}^N \beta_j \cdot X_{j,t+1}^i + \frac{\theta^i(B)}{(1-B)^{d\phi^i(B)}} \eta_t^i & \text{if } promo_{t+1} = 1 \text{ and } PromoIn < 5, \\ \frac{1 + \gamma^i \cdot B}{1 - B} \eta_t^i & \text{if } promo_{t+1} = 0 \end{cases} \quad (9)$$

where  $PromoIn$  is the number of weeks under promotions in the estimation sample.

It is important to explain why the third equation in (9) is required. We are introducing a different noise model for promotional and no promotional observations. The reason behind this is that the error term identified for promotional observations may not be adequate for non-promotional periods. In fact, in simulations on our dataset we have obtained more accurate results by including a separate error term for “normal” sales rather than keeping the noise term identified during promotion periods. Additionally, note that the inclusion of explanatory variable  $X_{j,t+1}^i$  assumes that the future promotion should be known in advance. This is a valid assumption, as the duration and terms of promotions are most often part of an accepted promotional schedule.

#### 4. Experimental results

In order to test the aforementioned models, a predictive empirical experiment is carried out. The last 30 weeks of each SKU time series (20% of the whole dataset) are reserved as a hold-out sample and are used for evaluating the different forecasting models. In this hold-out sample, there are 249 weeks under promotion, i.e., 21% of the promoted weeks in the dataset are in the hold-out; thus, evaluating the forecasting performance on a representative mix of promoted and non-promoted weeks.

A rolling origin evaluation experiment is designed as follows. Once the forecast is made, the origin is moved one week ahead until the complete 30 weeks of the hold-out sample is exhausted. The forecasting horizon is one week ahead, which is relevant to this case study. All models are estimated using the ECOTOOL Matlab toolbox (Pedregal et al., 2010).

To compare the forecasting techniques we use the Mean Absolute Error (MAE), which is defined as  $MAE = mean(|E_t|)$  and  $E_t$  is the normalized error given by:

$$E_t = (\bar{y}_t - \bar{F}_t), \quad t = 1, \dots, T \quad (10)$$

Here,  $\bar{y}_t$  and  $\bar{F}_t$  stand for the normalized actual value and the forecast, respectively, at time  $t$ . The sales standard deviation in the estimation sample has been used as a normalization factor. The MAE is computed across time and then across SKUs providing an overall summary error metric. This is possible as we have first normalized the values.

Before comparing the final dynamic regression with other methods, it is important to assess the performance improvements achieved by each of the different model components introduced in section 3.3. Table 2 shows the MAE of promotional weeks obtained from the simple linear regression with the original variables to the final dynamic regression shown in (9). The first innovation introduced was the substitution of the original variables by the principal components (section 3.3.1). The results of the regression based on principal components is shown in the 5th. row of the table and its performance was better than the simple linear regression with the original variables (4th. row) and the system and expert forecasts (FF). Note that those results only use a few Principal Components instead of the original variables. The models in the sixth row carries out a pooled regression for SKUs without enough promotions in the estimation sample (section 3.3.2). It should be noted that the pooled regression may utilize more Principal Components than the regression per SKU, since variables such as Category and Customer

Table 2: MAE associated to the different regression types in promotional periods

Method	MAE
System Forecast (SF)	1.069
Final Forecast (FF)	1.327
Regression with original variables	1.116
Regression with Principal Components	1.019
Pooled Regression with PC	1.008
Dynamic Pooled Regression with PC	0.969
Dynamic Pooled Regression with PC and modeled error (DR)	<b>0.898</b>

play an important role. These variables were found not to be significant when regression with principal components was computed for each single SKU. Here, apart from the error improvement, the pooled regression allow us to forecast some SKU weeks that the previous regression were not able to. Dynamics of the promotional effects (section 3.3.3) are incorporated in the model (7th. row) decreasing again the forecasting error. Finally, the last row reports the importance of modeling the error term (section 3.3.4) by the considerable improvement reached by the final Dynamic Regression model described in (9), hereafter the DR model. Summarizing, Table 2 exhibits the improvement achieved by each regression model refinement.

An interesting fact shown in Table 2 is that FF accuracy is worse than SF. This indicates that experts do not improve the baseline forecast accuracy in promotional periods. Nevertheless, even when FF on average seems to perform worse, it can be the consequence of outliers. In order to have a broader picture of the experts accuracy, we can plot the MAE vs the normalized adjustments sorted with respect to their size. In that sense, the evolution of the error against the size and sign of adjustments can be analyzed. Figures 2-4 depicts the MAE versus normalized adjustments for different situations.

Figure 2 shows the performance of the forecasting techniques when observations are under promotions. In this figure, there is an interval where FF is more accurate than SF (between 1 and 3 scaled adjustments). Here, FF performs better because experts knowing that there is a promotion look back to previous promotions and modify system forecast to achieve similar promotional sales patterns to the past. Those modifications up to certain adjustments size are fine, however, when adjustments are too big, experts produce bigger errors than the SF.

In the same figure, we can see that the Dynamic Regression (DR) also captures the past promotional sales pattern and projects it for future promotions yielding the best results. In fact, unlike the rest of the methods, it provides a good performance for large adjustments. It should be noted that most of the adjustments under promotions are positive as it can be seen in the lower panel histogram in Figure 2. Furthermore, negative adjustments are smaller and less frequent.

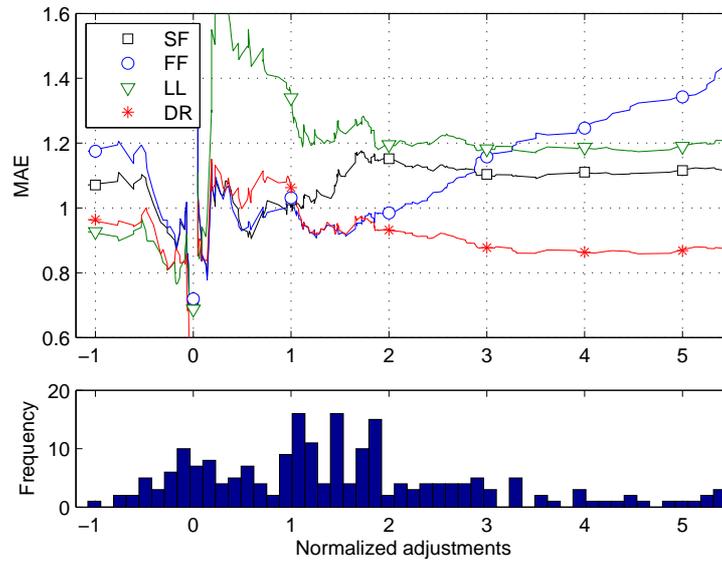


Figure 2: MAE of considered models under promotions by normalized adjustment size. The lower part of the figure provides a histogram of the adjustments. SF: System Forecast; FF: Final Forecast, LL: Last like, DR: Dynamic Regression

Figure 3 only depicts non-promotional observations. Here, it is interesting to point out that all methods that somehow include information about promotions (FF, LL, DR) performs similarly well for negative adjustments. After a promotion has occurred, it is reasonable that sales would decrease. However, the System Forecast does not handle additional information to indicate when a promotion has ended; hence, their forecast remains high. In order to correct the higher forecasts given by the SF, experts modify them by judgmentally reducing the value of forecasts, i.e., making negative adjustments. This effect is also automatically identified by LL and DR. This

offers an informative and good explanation of the good performance of negative adjustments on forecasting accuracy. Previous references found that experts' negative adjustments were more accurate (Fildes et al., 2009; Trapero et al., 2011); here, we provide a feasible explanation why this is so, at least regarding adjustments due to promotional activity.

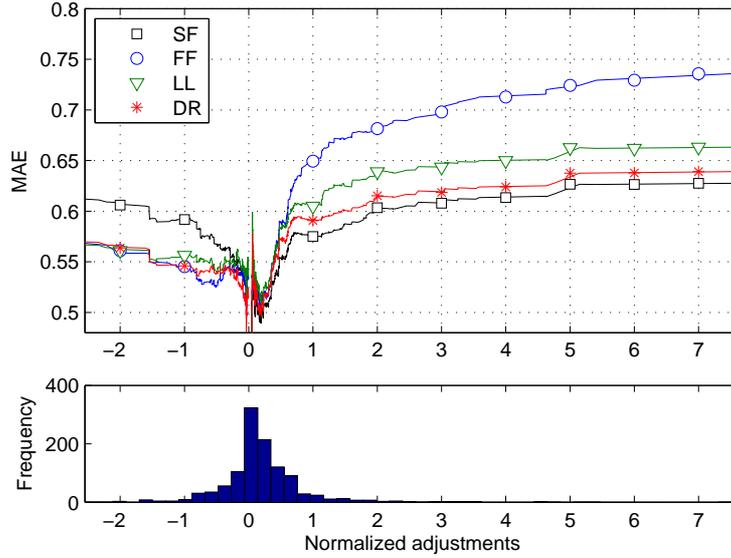


Figure 3: MAE of proposed models when there are no promotions. The lower part of the figure provides a histogram of the adjustments. SF: System Forecast; FF: Final Forecast, LL: Last like, DR: Dynamic Regression

Figure 4 depicts the overall results of the forecasting methods under study considering all the observations irrespective of whether they were promoted or not. Here, the forecasting technique DR provides very competitive results for both positive and negative adjustments.

In summary, Table 3 shows the MAE achieved for each method. The values highlighted in bold show the method with the most accurate results. The last row in Table 3 shows the overall performance, where the best performance is obtained by the Dynamic Regression (DR). Second and first rows break down the Overall results for promotional or non-promotional periods. Again, DR outperforms the rest of the models. Notice that DR achieves substantial improvements on promotional periods, being 16% and 32% more

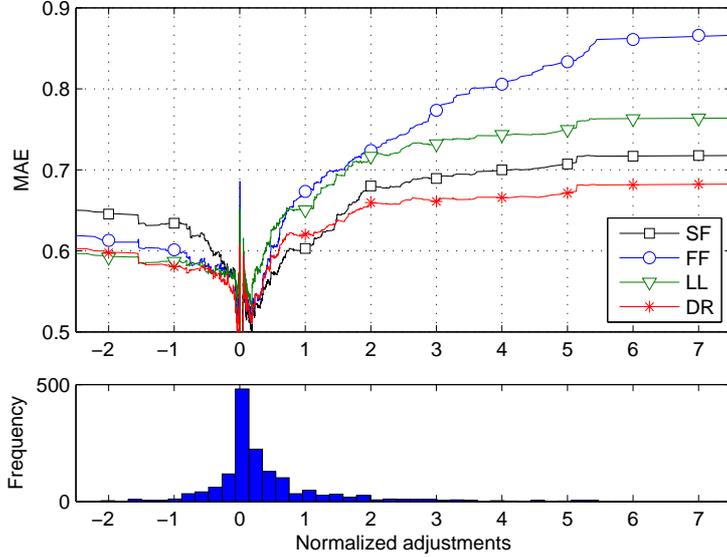


Figure 4: Overall models performance. The lower part of the figure provides a histogram of the adjustments. SF: System Forecast; FF: Final Forecast, LL: Last like, DR: Dynamic Regression

accurate than SF and FF respectively, which is the current practice in our case study company.

Table 3: Hold-out sample MAE of methods

	SF	FF	Naïve	SES	LL	DR
No promo	0.607	0.652	0.762	0.629	0.612	<b>0.601</b>
Promo	1.069	1.327	1.092	0.904	1.122	<b>0.898</b>
Overall	0.671	0.745	0.808	0.667	0.682	<b>0.642</b>

Mean of Absolute Percentage Error (MAPE) is an easy to communicate error measure and the most commonly used in practice (Hyndman and Koehler, 2006). This is calculated as the mean of  $APE_t = |y_t - F_t|/y_t$ . Absolute percentage errors have been criticized to provide biased results, and can be misleading when extremely high percentages, arising from relatively low actual demand values, are present. In order to mitigate these problems, Table 4 summarizes our results using the median. DR considerably outper-

forms FF, SF and the statistical benchmarks in the presence of promotions. These results indicate the validity of DR as a sales forecasting tool under promotional activity.

Table 4: Hold out Median(MAPE) of methods

	SF	FF	Naïve	SES	LL	DR
No promo	<b>29.33</b>	30.00	35.71	30.97	28.75	29.53
Promo	36.36	41.00	33.33	30.42	36.94	<b>25.82</b>
Overall	32.85	33.53	30.81	29.98	<b>29.16</b>	30.60

Although our case study company requires one-week ahead forecasts of future demand, we evaluate the performance of the different methods for longer forecast horizons to explore whether the superior performance of DR is consistent. Table 5 provides the MAE for promotional periods for up to four-weeks ahead forecasts. The results are consistent with the previous findings. Note that in Table 5 we do not provide the performance of SF and FF as these are only available from our case study company for one-week ahead forecasts.

Table 5: Hold out MAE for different forecasting horizons

Horizon (weeks)	Naïve	SES	LL	DR
1	1.092	0.904	1.122	<b>0.898</b>
2	1.081	0.927	1.051	<b>0.911</b>
3	1.118	0.960	1.127	<b>0.940</b>
4	1.189	1.004	1.137	<b>0.953</b>

## 5. Conclusions

Recent literature has proposed multivariate regression models in contrast to judgmental forecasting to predict sales in the presence of promotions. This paper investigates the strengths and weaknesses of this kind of models and suggests several innovations in the context of promotional modeling at SKU level: i) Principal Component Analysis is used to reduce the dimension of the explanatory variables and avoid multicollinearity; ii) Pooled regression

is applied when the estimation sample does not contain enough promoted weeks; iii) The error term is modeled as a transfer function, automatically identified by minimizing the SBC; iv) These features can be expressed as a dynamic regression that outperforms the experts forecasts of a case study company, as well as other well-known benchmarks.

The results shows that the Dynamic Regression achieves the lowest forecast error. This outcome has several managerial implications. On the demand planners behalf, they can benefit from reduced effort to judgmentally adjust forecast, while forecasting accuracy is improved. On the Forecasting Support System developers behalf, multivariate models with the innovations proposed in this paper should be considered in their software to help their potential clients that usually works with promotional policies. Established commercial software does not offer this facility, leaving the inclusion of promotional factors to expert judgment. Last but not least, the increased accuracy of the proposed model has direct implications for inventory management. The effects of promotions are accurately captured and reflected in the expected demand. Moreover, the forecast uncertainty is reduced, thus enabling the calculation of more appropriate safety stock.

In addition, our modeling approach overcomes some of the limiting factors of the existing literature on promotional modeling; in particular the need for extensive data and secondly, the limited modeling experience in many companies that makes the use of complex econometric promotional models difficult, if not impossible.

We evaluated the performance of the proposed model for a manufacturer case study. Extending this study to other companies will allow us to better understand the performance of such promotional models. Additionally, further research should include alternative error metrics. Particularly, recent studies (Davydenko and Fildes, 2013; Snyder et al., 2012) propose the Average Relative MAE and distribution based scores respectively, to enhance the forecasting accuracy measure in supply chain forecasting applications.

## References

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., 1994. Time Series Analysis: Forecasting and Control. Prentice-Hall, Englewood Cliffs.
- Caniato, F., Kalchschmidt, M., Ronchi, S., 2011. Integrating quantitative and qualitative forecasting approaches: organizational learning in an ac-

- tion research case. *Journal of the Operational Research Society* 62, 413–424.
- Cooper, L. G., Baron, P., Levy, W., Swisher, M., Gogos, P., 1999. Promocast trademark: A new forecasting method for promotion planning. *Marketing Science* 18 (3), 301.
- Davydenko, A., Fildes, R., 2013. Measuring forecasting accuracy: The case of judgmental adjustments to sku-level demand forecasts. *International Journal of Forecasting* 29 (3), 510 – 522.
- Divakar, S., Ratchford, B. T., Shankar, V., 2005. CHAN4CAST: A multi-channel, multiregion sales forecasting model and decision support system for consumer packaged goods. *Marketing Science* 24, 334–350.
- Fildes, R., Goodwin, P., 2007. Against your better judgment? how organizations can improve their use of management judgment in forecasting. *Interfaces* 37, 70–576.
- Fildes, R., Goodwin, P., Lawrence, M., 2006. The design features of forecasting support systems and their effectiveness. *Decision Support Systems* 42 (1), 351 – 361.
- Fildes, R., Goodwin, P., Lawrence, M., Nikolopoulos, K., 2009. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting* 25, 3–23.
- Fildes, R., Nikolopoulos, K., Crone, S. F., Syntetos, A. A., 2008. Forecasting and operational research: a review. *Journal of the Operational Research Society* 59 (9), 1150–1172.
- Gardner, E. S., 1985. Exponential smoothing: The state of the art. *Journal of Forecasting* 4 (1), 1–28.
- Gardner, E. S., 2006. Exponential smoothing: The state of the art, Part II. *International Journal of Forecasting* 22, 637–666.
- Hughes, M. C., 2001. Forecasting practice: organisational issues. *Journal of the Operational Research Society* 52 (2), 143–149.

- Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 679–688.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer-Verlag, Berlin.
- Jolliffe, I. T., 1982. A note of the use of Principal Component in regression. *Applied Statistics* 31, 300–303.
- Jolliffe, I. T., 2002. *Principal Component Analysis.*, 2nd Edition. Springer series in statistics. Springer, New York.
- Lawrence, M., O’Connor, M., Edmundson, B., 2000. A field study of sales forecasting accuracy and processes. *European Journal of Operational Research* 122 (1), 151 – 160.
- Leeftang, P. S., van Heerde, H. J., Wittink, D., 2002. How promotions work: SCAN\*PRO-based evolutionay model building. *Schmalenbach Business Review* 54, 198–220.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R., Winkler, R., 1982. The accuracy of extrapolation (time series) methods; results of a forecasting competition. *Journal of Forecasting* 1, 111–153.
- Ord, J. K., Fildes, R., 2012. *Principles of Business Forecasting*. South-Western Cengage Learning, Mason, Ohio.
- Özden Gür Ali, Sayin, S., van Woensel, T., Fransoo, J., 2009. SKU demand forecasting in the presence of promotions. *Expert Systems with Applications* 36 (10), 12340 – 12348.
- Pedregal, D. J., Contreras, J., Sanchez, A., 2010. *Handbook of Networks in Power Systems*. Springer Verlag, Ch. ECOTOOL: A general MATLAB forecasting toolbox with applications to electricity markets, pp. 69–104.
- Sayrs, L., 1989. *Pooled Time Series Analysis*. Sage Publications, Newbury Park.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Stat.* 6, 461–464.

- Snyder, R. D., Ord, J. K., Beaumont, A., 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting* 28 (2), 485 – 496.
- Taylor, J. W., 2011. Multi-item sales forecasting with total and split exponential smoothing. *Journal of the Operational Research Society* 62, 555–563.
- Trapero, J. R., Fildes, R., Davydenko, A., 2011. Nonlinear identification of judgmental forecasts effects at SKU level. *Journal of Forecasting* 30, 490–508.
- Trapero, J. R., Pedregal, D. J., Fildes, R., Kourentzes, N., 2013. Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting* 29 (2), 234 – 243.
- Trusov, M., Bodapati, A. V., Cooper, L. G., 2006. Retailer promotion planning: Improving forecast accuracy and interpretability. *Journal of Interactive Marketing* 20, 71–81.