View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Lancaster E-Prints

Journal of Forecasting
J. Forecast. **30**, 490–508 (2011)
Published online 7 June 2010 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/for.1184

Nonlinear Identification of Judgmental Forecasts Effects at SKU Level

JUAN R. TRAPERO,* ROBERT FILDES AND
ANDREY DAVYDENKO
Department of Management Science, Lancaster University, Lancaster, UK

ABSTRACT

Prediction of demand is a key component within supply chain management. Improved accuracy in forecasts directly affects all levels of the supply chain, reducing stock costs and increasing customer satisfaction. In many application areas, demand prediction relies on statistical software which provides an initial forecast subsequently modified by the expert's judgment. This paper outlines a new methodology based on state-dependent parameter (SDP) estimation techniques to identify the nonlinear behaviour of such managerial adjustments. This non-parametric SDP estimate is used as a guideline to propose a nonlinear model that corrects the bias introduced by the managerial adjustments. One-step-ahead forecasts of stock-keeping unit sales sampled monthly from a manufacturing company are utilized to test the proposed methodology. The results indicate that adjustments introduce a nonlinear pattern, undermining accuracy. This understanding can be used to enhance the design of the forecasting support system in order to help forecasters towards more efficient judgmental adjustments. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS forecast adjustment; supply chain; nonlinear system identification

INTRODUCTION

Companies working within supply chains use forecasts of demand to drive purchasing and supply chain management. Accurate forecasts can affect positively the operational management of companies, leading to significant monetary savings, greater competitiveness, enhanced channel relationships and customer satisfaction, lower inventory investment, reduced product obsolescence; they can improve distribution operations, schedule more efficient production and distribution, and enable more profitable financial decisions (Moon *et al*., 2003). For most of these companies, a particular type of a decision support system, known as a forecasting support system (FSS), is employed to prepare the forecasts (Fildes *et al*., 2006). These FSSs integrate a statistical forecasting approach with managerial judgment from forecasters in the organization.

* Correspondence to: Juan R. Trapero, Department of Management Science, Lancaster University, Lancaster LA1 4YX, UK. E-mail: j.traperoarenas@lancaster.ac.uk

Copyright © 2010 John Wiley & Sons, Ltd.

The manager's judgment is an important element within the forecasting process (Lawrence *et al.*, 2006). For instance, judgment influences a wide range of decisions which range from the selection of a more appropriate statistical method to direct modification of the quantity being forecast. In fact, managers may have access to information that is difficult to include in a statistical model, for example the effects of a promotion campaign. Thus judgmental adjustments can incorporate that information into the model in order to improve the forecast accuracy (Fildes *et al.*, 2006).

Franses and Legerstee (2009) presented a case study where the experts adjusted the statistical forecast in 89.5% of cases. Nonetheless, even though managers make frequent adjustments, the literature devoted to study its effect at the stock-keeping unit (SKU) level is scarce (see Mathews and Diamantopoulos, 1990; Fildes *et al.*, 2009; Syntetos *et al.*, 2009; Franses and Legerstee, 2009, 2010).

The recent literature suggests the existence of a bias towards making overly positive adjustments (Fildes *et al.*, 2009) or as a consequence of a non-symmetric loss function of the managers (Franses and Legerstee, 2009). Mello (2009) analyzes the biases introduced by means of forecast game playing, defined as the intentional manipulation of forecasting processes to gain personal, group, or corporate advantage. Eroglu and Croxton (2010) explore the effects of particular individual differences and suggest that a forecaster's personality and motivational orientation significantly influence the forecasting biases. Since the companies in the supply chain are interdependent, the bias introduced into sales forecasts by one company affects the rest of the companies along the chain. Therefore the reduction of biases in sales forecasts is of paramount importance.

In order to correct the presence of the bias several works have modelled the appropriate weight that statistical forecasting and judgmental forecasting should have. For instance, Blattberg and Hoch (1990) took the mean of each approach that proved effective. Fildes *et al.* (2009) propose an optimal adjust model based on linear regression, classifying the data depending on the adjustment sign. In contrast to Blattberg and Hoch (1990), it was found that negative adjustments were more precise than positive ones. This discontinuity between positive and negative adjustments may indicate the desirability of adopting nonlinear models to describe the judgmental process. In fact, such nonlinearities can be considered in the design of the FSS to mitigate the worst effects of such biases.

The present work reports the nonlinear effect of adjustments on the final forecast accuracy on the basis of a manufacturing company database containing one-step-ahead forecasts and the actual sales. Assuming the expert adjustment is predictable (Franses and Legerstee, 2009) or fixed, this nonlinear identification is employed to propose a model that can correct the aforementioned bias and improves overall forecasting accuracy.

A state-dependent estimation (SDP) approach is used to study the nonlinearities involved in the manager's adjustment. SDP nonlinear estimation belongs to a family of methods within data-based mechanistic modelling (DBM) developed by Young and co-workers (see Young *et al.*, 2001; Young, 2006; Young and Garnier, 2006; and references therein, among others). The SDP technique uses recursive methods like fixed interval smoothing (FIS) combined with special data reordering and 'backfitting' procedures which show in a non-parametric way, i.e. through a graph, the state dependency between the parameter under study and an associated state variable (Young *et al.*, 2001).

The outline of the paper is the following: the next section describes the problem formulation; the third section explains the SDP approach; the fourth section analyzes a case study to verify the model proposed; and finally the fifth section reports the main conclusions and their implications for achieving improvements in practice.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

PROBLEM FORMULATION

The optimal adjust model, proposed by Fildes *et al*. (2009), aimed at optimally combining two of the sources of information available to the forecaster: the system forecast and the forecaster's subjective adjustment in order to deliver a more accurate forecast. It is given by

$$y_{i,t} = \alpha_1 \mathrm{SF}_{i,t} + \alpha_2 \mathrm{Adj}_{i,t} + \nu_{i,t} \tag{1}$$

where $y_{i,t}$ is the actual value of sales for the ith product of the analysed company at time t. The regressors are $\mathrm{SF}_{i,t}$ and $\mathrm{Adj}_{i,t}$, which stand for the system forecast (statistical forecast) and the adjustment forecast one step ahead at time t, respectively. The adjustment forecast variable is computed as

$$\mathrm{Adj}_{i,t} = \mathrm{FF}_{i,t} - \mathrm{SF}_{i,t} \tag{2}$$

where $\mathrm{FF}_{i,t}$ is the final forecast employed by the managers. The error term is $\nu_{i,t}$.

In order to assess the influence of the judgmental adjustment on the accuracy of forecasts we propose a more flexible version of the optimal adjust model. Fildes *et al*. (2009) provided statistical tests which indicated that coefficients α_1 and α_2 are different depending on the adjustment sign. Also they claimed that, according to the data extracted from the four companies analysed, negative adjustments tended to improve the forecast accuracy and the size of the adjustment affected the accuracy. In turn, positive adjustments tended to decrease the forecast accuracy. In order to correct the aforementioned bias a nonlinear model is proposed:

$$y_{i,t} = \alpha_1(v_1(i,t))\mathrm{SF}_{i,t} + \alpha_2(v_2(i,t))\mathrm{Adj}_{i,t} + \nu_{i,t} \tag{3}$$

The aim is to determine the potential states $v_1(i, t)$ and $v_2(i, t)$, as well as to estimate the unknown SDP $\alpha_1(v_1(i, t))$ and $\alpha_2(v_2(i, t))$, which may offer a better explanation of the nonlinear process described by $y_{i,t}$. By understanding how the company's forecasters misweight the information available it may be possible to develop an FSS that overcomes some of the worst excesses (Fildes *et al*., 2009).

A STATE-DEPENDENT PARAMETER ESTIMATION APPROACH

Following the work of Fildes *et al*. (2009) and Syntetos *et al*. (2009), we have grouped all the observations as cross-sectional data, dealing with each observation as an individual case. Note that adjustments between adjacent time periods may be correlated. However, independence between consecutive adjustments is a reasonable assumption based on the decision maker's tendency to consider problems as unique (Kahneman and Lovallo, 1993).

Since the variance of each SKU can be different, data normalization is also required. For instance, it is possible to normalize with respect to the standard deviation of each SKU as proposed by Fildes *et al*. (2009). Furthermore, since the parameters are expected to vary depending on the adjustment size, the data are sorted with respect to the adjustments. Accordingly, the data can be reindexed by $k = 1, \ldots, N$, where N is the sample size. In this sense equation (3) is rewritten as

$$y_k = \alpha_1(v_1(k))SF_k + \alpha_2(v_2(k))Adj_k + \nu_k \tag{4}$$

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

The SDPs are expressed by $\alpha_1(v_1(k))$ and $\alpha_2(v_2(k))$, where $v_i(k)$, $i = 1, 2$ is the variable which drives the behaviour of the aforementioned SDP. The random noise ν_k is assumed Gaussian with zero mean and variance σ^2.

In order to determine $\alpha_1(v_1(k))$ and $\alpha_2(v_2(k))$ described in (4) several assumptions have been made to capture the adjustment process. Firstly, it is assumed that $v_1(k)$ remains constant. In other words, as a company usually needs to predict the demand of a vast number of products, an 'automatic' forecasting technique[1] implemented in a FSS is used for this purpose, providing the first regressor (SF_k) in (1). Therefore, it is expected that the weight of SF_k is approximately the same for a wide range of products, yielding a constant α_1. In the second place, since the effectiveness of the adjustments may differ depending on the adjustment sign, it is assumed the parameter α_2 does not remain constant. Indeed, we assume that α_2 is a function of the adjustments represented by Adj_k.

The SDP modelling procedure allows us to incorporate this form of nonlinearity. Therefore, taking into account the previous assumptions, the SDP-optimal adjust model is proposed as an extension of the optimal adjust model described in (1), such as

$$y_k = \alpha_1 \mathrm{SF}_k + \alpha_2(\mathrm{Adj}_k)\mathrm{Adj}_k + \nu_k \tag{5}$$

Note that it is possible to formulate more complicated models based on the SDP procedure; for example, we can assume α_1 is also state dependent. Nevertheless, we prefer expression (5) because even when it is a nonlinear model each term can be easily interpreted following the DBM philosophy (Young, 2006).

In order to let the parameter α_2 vary with adjustment, a first approach would be to define stochastically the parameter α_2 as a two-dimensional stochastic state vector, whose stochastic properties are defined by a generalized random walk (Jakeman and Young, 1984). There is a wide range of options (Pedregal and Young, 2002). Generally, the stochastic state vector is also called the time-varying parameter (TVP) because the data are ordered in a temporal fashion associated with time series problems. However, we are interested in looking for the variations of α_2 with respect to the adjustments instead of time. Since we are not working in an online fashion, the data can be sorted by adjustment size and then the TVP procedure can be carried out. After that, the data are 'unsorted' (an *unsort* operation, to reverse MATLAB's *sort*) to its original time order (Young *et al*., 2001).

Fortunately, the SDP technique allows us to sort and unsort the data and run the TVP procedure. Additionally, the algorithm includes back-fitting procedures employing recursive FIS algorithms to achieve estimations of any state-dependent parameter. The outcome of the SDP algorithm is a non-parametric estimate displayed as a graph of the state-dependent parameter ($\alpha_2(\mathrm{Adj}_k)$) against the variable which affects it in a nonlinear fashion (Adj_k).

The stochastic behaviour of α_2 chosen in this application is an *integrated random walk* which consists of

$$\begin{pmatrix} \alpha_2(k+1) \\ \alpha_2^*(k+1) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_2(k) \\ \alpha_2^*(k) \end{pmatrix} + \begin{pmatrix} 0 \\ w^*(k) \end{pmatrix} \tag{6}$$

where α_2 and α_2^* are associated with the changing level and slope of the SDP; the flexibility in the model is introduced by the random Gaussian noise $w^*(k)$ with mean zero and variance σ_α^2.

[1] For instance, an exponential smoothing method.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

The full model is formulated as a state space (SS) system by assembling the observation equation in (5) and the state equations in (6). The SS formulation is well suited for optimal recursive estimation accomplished by well-known recursive algorithms such as the Kalman filter (KF) in Kalman (1960) and the fixed interval smoothing (FIS) in Bryson and Ho (1969). However, in order to use these algorithms all the system matrices must be assumed known. In this case, the noise variances of the observation equation (σ^2) and the state equations (σ_α^2) are the unknown parameters (often called hyper-parameters to distinguish them from the main parameters or states in (6)).

Usually, the variances are normalized by the innovations variance (σ^2) reducing to one the number of unknown parameters. In this sense, the noise variance ratio (NVR) is defined as σ_α^2/σ^2. The optimization of the NVR can be done by maximum likelihood (ML) in the time domain obtained via 'prediction error decomposition' (see Harvey, 1989).

A complete description of the technique, with numerous examples, can be found in Young *et al*. (2001), and some applications to environmental systems are shown in Young and Garnier (2006) and Young (2006), among others. Additionally, SDP algorithms are available within the CAPTAIN[2] toolbox (Taylor *et al*., 2007), developed for use with MATLAB/Simulink™ software.

CASE STUDY

Data from a manufacturing company specialized in household products have been collected. The data have been split into three series which represent: (i) one-step-ahead systems forecasts; (ii) one-step-ahead final forecasts; and (iii) corresponding actual outcomes for 413 SKUs. The data comprises 7544 completed triplets that have been sampled monthly between 2004 and 2007. It is an extended dataset from that analysed as company A in Fildes *et al*. (2009).

Basically, the final forecast produced by the company is the result of two sources of information (Fildes *et al*., 2009). On the one hand, there is available computer software which provides the statistical system forecasts. On the other hand, various meetings, which involve the company forecasters meeting with personnel in sales, marketing, and production, occur to share pieces of information that cannot be included in the statistical model. The responsibility for the final forecast rests with the forecasters, however. Thus the previous system forecast is adjusted accordingly with the meeting group decisions, obtaining an agreed forecast (final forecast). For instance, approximately 65% of the 7544 complete triples were modified by the demand forecaster's adjustments.

Data selection

Since the characteristics of the data are heterogeneous a pre-treatment step is required in order to generate a homogeneous sample. We agree with Syntetos *et al*. (2009) concerning the fact that data selection has been overlooked in the past. Basically, this step eliminates those time series that are not useful for the experimental analysis (depending on the goals of the research). For instance, Fildes *et al*. (2009) remove SKUs without the required continuous forecast history or those with low-volume SKUs because these are the result of special circumstances, such as particular items having been withdrawn from the market. In contrast, Syntetos *et al*. (2009) focused on intermittent demand.

Our aim here is to develop a model of the adjustment process for established SKUs that captures any nonlinear effects, so the pre-treatment stage in this work removes the time series which fulfils any of these conditions:

[2] see http://www.es.lancs.ac.uk/cres/captain/.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

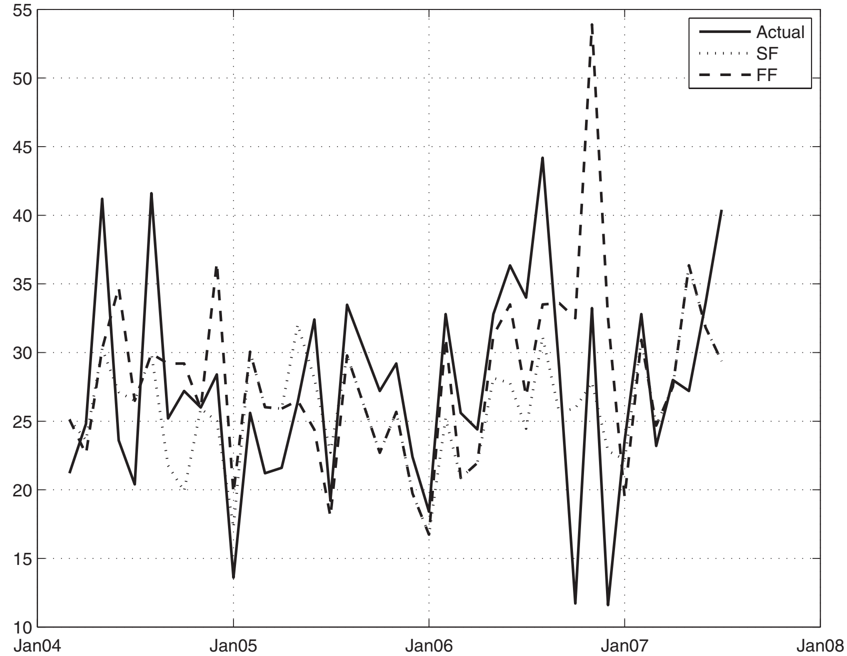

Figure 1. Example of time series considered per each SKU. Actual values are shown as a solid line, system forecasts as a dotted line and final forecasts as a dashed line

- time series with less than 12 months history available;
- time series with any actual observation equal to zero.

After this pre-treatment the number of SKUs is reduced to 91 with 2882 triplets. An example of the time series considered for each SKU can be seen in Figure 1, where actual values are shown as a solid line, system forecasts as a dotted line and final forecasts as a dashed line. Recall that $\mathrm{FF}_k = \mathrm{SF}_k + \mathrm{Adj}_k$.

Exploratory data analysis

In order to take advantage of the judgemental adjustments we have to check that adjustments improve the forecasting accuracy provided by the statistical forecast. Basically, if there is no evidence that the final forecasts based on adjustments beat the statistical forecast accuracy it would be unlikely to propose a model which uses as a regressor those adjustments capable of beating the statistical forecast. In this sense, Table I assesses the forecasting performance provided by the company under study, in which cases with lower error are shown in bold. In this table the mean absolute percentage error (MAPE) and the median absolute percentage error (MdAPE) were chosen as accuracy measures, i.e.

$$\begin{aligned} \mathrm{MAPE} &= \mathrm{mean}(|p_t|) \\ \mathrm{MdAPE} &= \mathrm{median}(|p_t|) \end{aligned} \tag{7}$$

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Table I. Mean of MAPE and MdAPE for SF and FF

Adjustment	No. of observations	Mean(MAPE)		Mean(MdAPE)	
		SF	FF	SF	FF
Positive	1249	**26.99**	32.72	23.37	**18.48**
Negative	601	71.42	**38.24**	34.13	**23.82**
None	1032	30.34	30.34	21.32	21.32
Overall adjusted	1850	40.39	**33.38**	24.04	**17.47**
Total	2882	38.05	**32.71**	21.49	**17.13**

where p_t is the percentage error given by $p_t = 100|Y_t - F_t|/Y_t$, $t = 1, \ldots, N$. In this expression Y_t stands for the actual value at time t and F_t is the forecast at that time. All forecasts considered are one step ahead and N is the sample size.

In particular, these error measures have been computed across time for each SKU, and then the mean of these values were calculated across SKUs. The last row in Table I shows the total mean of the MAPE and MdAPE, where the values in bold highlight the best performance method. In general terms, the FF is more accurate than the SF. Additionally, we can break down the errors according to the adjustment sign, obtaining three rows which analyse the forecasting performance of the positive and negative adjustments, as well as when there is no adjustment. The row denoted by 'Overall adjusted' comprises positive and negative adjustments excluding those observations that were not modified by judgmental adjustments. In addition, the second column also shows the sample size of each kind of adjustment, where it is possible to verify that positive adjustments are more frequent.

In relation to the negative adjustments, we can see from Table I that the FF is more accurate than the SF. Nevertheless, this same conclusion cannot be extrapolated to the positive adjustments case. In fact, there is no clear conclusion about whether the FF is more accurate than the SF because according to the mean(MAPE) the SF outperforms the FF but, conversely, assessing the mean(MdAPE) the FF beats the SF. A possible explanation of this discrepancy is that positive adjustments can frequently lead to over optimistic forecast errors, which occur when the actual value turns out to be below the forecast. According to Makridakis (1993), the absolute percentage error (APE) for such over-forecast errors is greater than for under-forecast errors (when the actual is above the forecast). However, while the MdAPE is robust to any distortion associated with over-forecast errors, the MAPE is quite sensitive to it. In order to solve this discrepancy we can normalize the data and then compute the mean absolute error (MAE = mean($|Y_t - F_t|$)) which is not a percentage error measure. In the next section such data normalization is carried out, where it will be shown that not only is it valuable for error comparison purposes but also it is necessary to identify a nonlinear pattern in the process of adjustments.

Data normalization

One of the main objectives of this work is to find out if there is a bias in the adjustments accomplished by the forecasters. If so, this information can be used to propose a model which improves forecasting accuracy. Nonetheless, we are mixing different SKUs with different statistical properties. Thus it is convenient to provide a framework where it is possible to compare them. This can be done by means of data normalization. In particular, each product can be normalized with respect to its sales standard deviation (Fildes *et al*., 2009). Note that other normalization alternatives are possible. In fact, the choice of normalization factor is still an open issue that from the authors' point of view

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Table II. Exploratory normalized data analysis

	Actual	System forecast	Final forecast
Mean	3.5	3.4	3.7
25th percentile	2.3	2.2	2.5
Median	3.4	3.3	3.6
75th percentile	4.6	4.5	4.7
SD	1.6	1.6	1.6
MAD	1.1	1.1	1.1

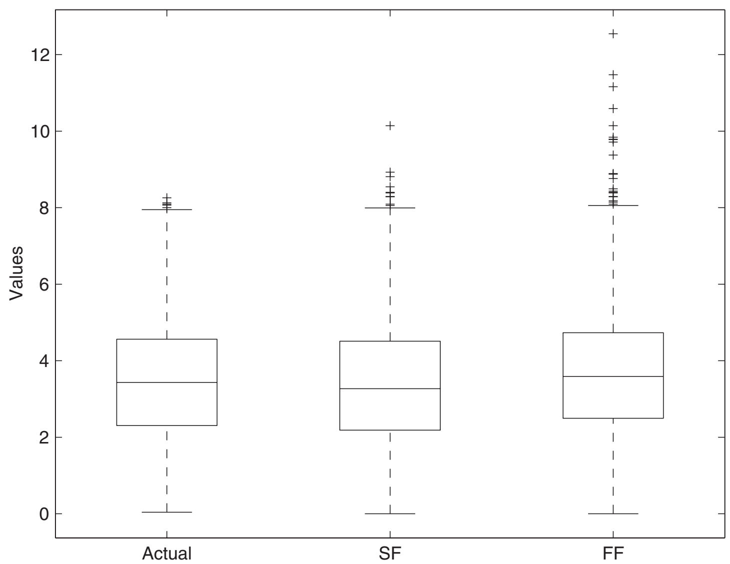

Figure 2. Box plots of the normalized actual values, system and final forecasts

deserves further work. Nonetheless, in this article the SKU sales standard deviation has been chosen as a normalization factor in order to be able to compare our results with previous published works.

A statistical description of the normalized data can be found in Table II. As a consequence of the normalization each statistic can be interpreted as a ratio. For example, the first row and column mean that the actual sales level is, on average, 3.5 higher than its standard deviation. Additionally, it is also interesting to note that central measures like the mean and median corresponding to the 'Actual' column are higher than their system forecast counterpart. This difference between the real and SF has been detected and managerial adjustments were imposed to compensate this SF bias, as can be seen in the statistics for the FF. However, this compensation was too optimistic, achieving an FF mean and median higher than the actual ones. In the last two rows of this table we can also find two dispersion measures: the standard deviation (SD) and the median absolute deviation (MAD). These measures show a similar dispersion between the SF, FF and actual values.

Figure 2 depicts the box plot of the normalized actual values, as well as the system and final forecast provided by the company. Note that there are a higher number of extreme values in the final forecast which do not correspond to any actual value. This implies that some large positive adjustments have been made incorrectly.

Copyright © 2010 John Wiley & Sons, Ltd.
DOI: 10.1002/for

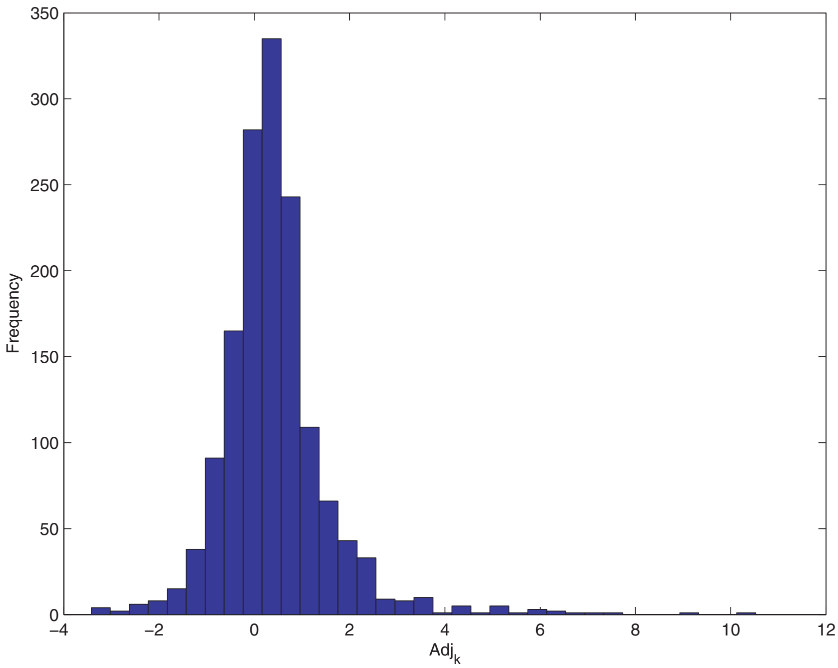

Figure 3. Histogram of normalized adjustments

Table III. Evidence of optimism bias in adjustments

Adjustments	% of times adjustment is too large	% of times adjustment is in wrong direction	Total % of adjustments that are overoptimistic
Positive	34.3	27.7	62.0
Negative	25.3	23.0	51.7

Figure 3 shows the histogram of the normalized adjustments, where it is possible to see that the adjustments are right skewed. In this sense, Table III analyses the overoptimism in adjustments. For instance, the first row shows that positive adjustments tend to overestimate. In particular, 62% of them are positively biased because they were too large or they were in the wrong direction. In contrast, negative adjustments are less biased since only 51.7% of them were overoptimistic.

Previously, it was mentioned that the mean(MAPE) and the mean(MdAPE) might not be good error measures to compare the accuracy of the system and final forecast for positive adjustments because of the greater APE associated with over-forecast errors. Therefore we can take advantage of the suggested normalization to solve this problem by computing the MAE on this normalized dataset. In this way, Table IV shows the MAE accomplished by the SF and FF. Unlike Table I, Table IV shows that the FF is more accurate than SF even for positive adjustments. The explanation of the difference between the results for the normalized versus the percentage error measures lies in the different weight given to larger errors associated with volatile SKUs.

Another advantage of the normalization is that we can analyse the relationship of the MAE with respect to the size of the adjustment. Indeed, Figures 4 and 5 depict the aforementioned evolution

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Table IV. MAE on the basis of normalized data

Adjustment	System forecast	Final forecast
Positive	0.798	**0.719**
Negative	0.779	**0.513**
None	0.552	0.552
Overall adjusted	0.789	**0.652**
Total	0.704	**0.616**

with regard to positive and negative adjustments, respectively. We can observe that: (i) FF is more accurate than SF for larger adjustments; (ii) there are no large differences between SF and FF methods for small adjustments; (iii) the improvement achieved by the FF in negative adjustments is larger than for the positive counterparts. It should be pointed out that these findings agree with those suggested in Fildes *et al*. (2009).

Non-parametric SDP estimation

Under the assumption that some parameters may be state dependent, SDP algorithms have been employed to obtain non-parametric estimates. The model estimated assumed that the weight of the adjustments in (5) is potentially dependent on the adjustments as a state. In order to perform this estimation an integrated random walk described in (6) is used to model variations in the SDP, where the observations are ordered by adjustment size. As a result of this stage, a graph is provided giving us an indication of the possible nonlinearity shape. This stage has been accomplished with the *MATLAB*™ toolbox called CAPTAIN (Taylor *et al*., 2007).

Figure 6 depicts the estimation of $\alpha_2(\mathrm{Adj}_k)$ as a solid line, and the standard errors of its estimation in dashed lines. According to this figure, there is a variation of the parameter α_2 depending on the adjustment sign. It is interesting to note that the weight of the adjustments represented by α_2^* is greater for negative adjustments. This means that positive adjustments tend to be optimistic and the SDP is tuned to damp this optimism. In other words, negative adjustments are more accurate than positive adjustments. It is interesting to note the model residuals show no significant correlation with the adjustments. Additionally, confidence intervals show that there is a high uncertainty for adjustment values close to zero. One explanation is that forecasters may make small adjustments when they mistake noise for patterns in the signal. Furthermore, confidence intervals are tighter for positive small adjustments, indicating that a large quantity of data is concentrated in this range (see Figure 3).

Identification and estimation of nonlinearities regarding SDP

With the non-parametric estimate computed in the previous step, we now examine the graph obtained to propose a nonlinear model capable of capturing the source of such nonlinearities. Typically, this task can be done via nonlinear parametric models which may range, for instance, from a radial basis function to a sigmoidal law (see Young, 2006; Young *et al*., 2001). Then, the parametric model is efficiently estimated by nonlinear least squares, prediction error minimization or maximum likelihood optimization.

Considering Figure 6 the following nonlinear model is proposed:

$$y_k = \beta_1 \cdot \mathrm{SF}_k + \beta_2 \cdot \mathrm{Adj}_k + \left(\beta_3 + \beta_4 \cdot e^{-\beta_5 \cdot \mathrm{Adj}_k}\right)\mathrm{Adj}_k \cdot X_d + \nu_k \quad (8)$$

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

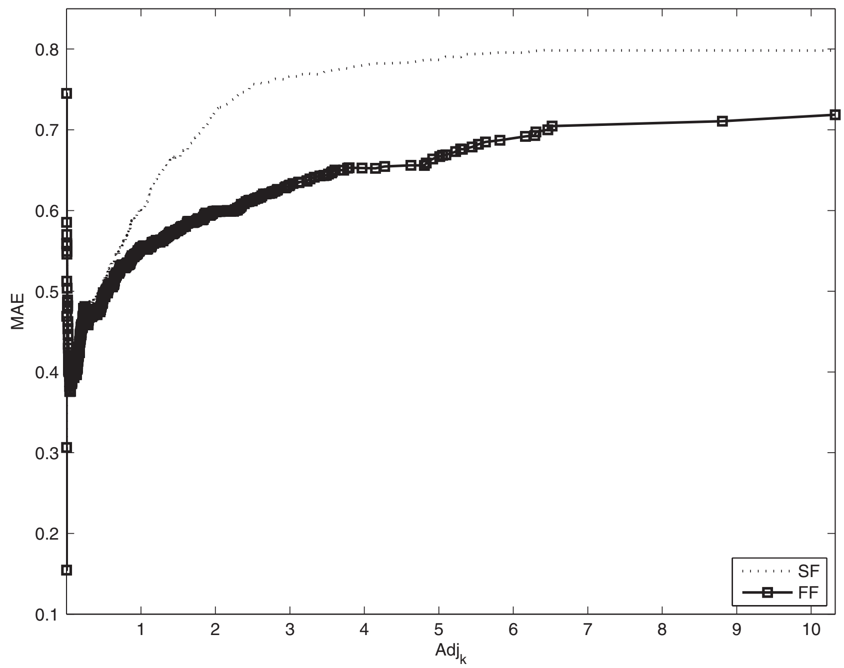

Figure 4. MAE for normalized positive adjustments

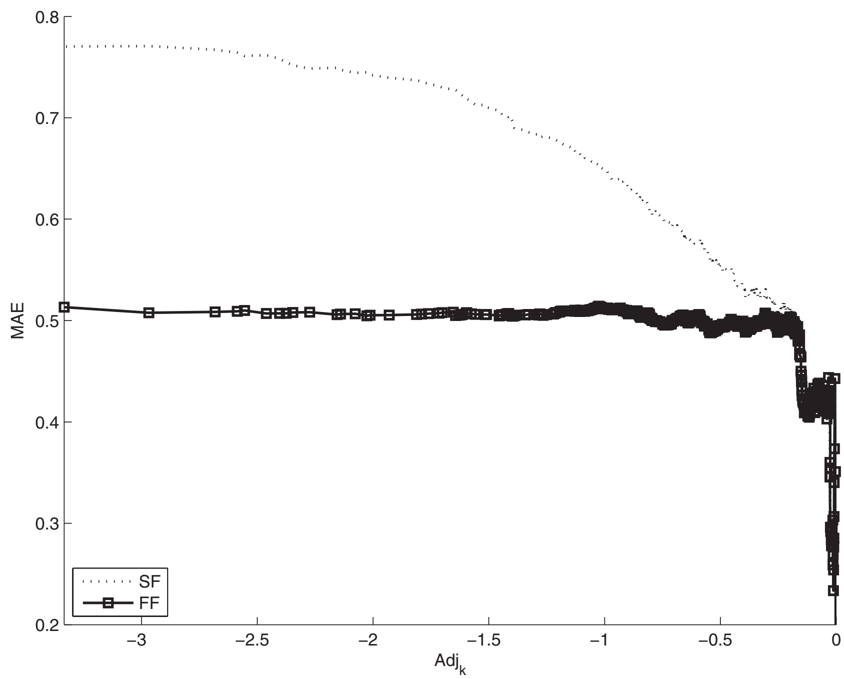

Figure 5. MAE for normalized negative adjustments

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

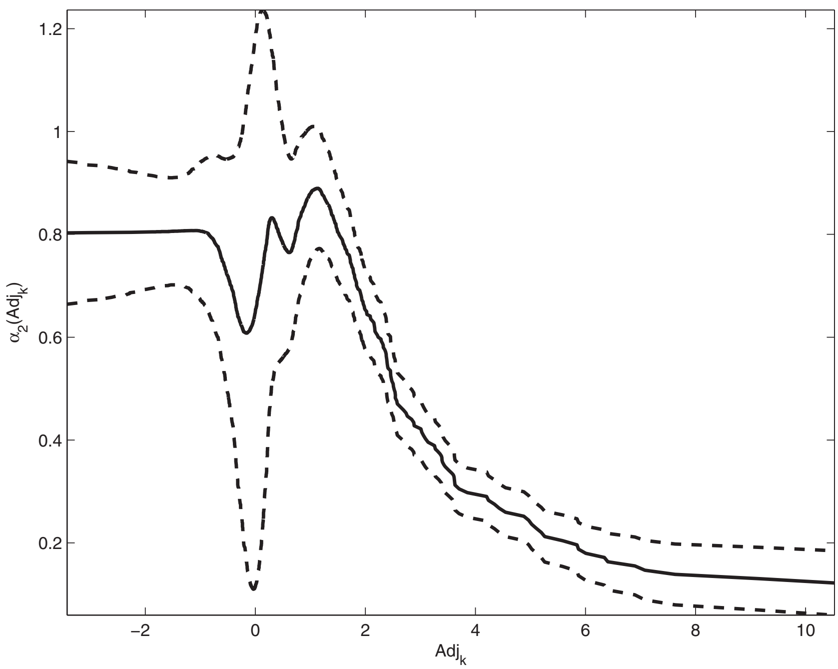

Figure 6. Non-parametric estimation of $\alpha_2(\mathrm{Adj}_k)$ against Adj_k

where X_d is a dummy variable such as

$$X_d = \begin{cases} 0 & \text{if} \quad \mathrm{Adj}_k < 0 \\ 1 & \text{if} \quad \mathrm{Adj}_k > 0 \end{cases} \tag{9}$$

The estimates of the model parameters are given below (the respective estimated standard deviation is in parentheses):

$$\begin{array}{ll} \hat{\beta}_1 = 0.94919 \ (7.36 \cdot 10^{-5}) & \hat{\beta}_2 = 0.824 \ (2.74 \cdot 10^{-3}) \\ \hat{\beta}_3 = -0.94 \ (1.36 \cdot 10^{-2}) & \hat{\beta}_4 = 1.323 \ (6.70 \cdot 10^{-3}) \\ \hat{\beta}_5 = 0.247 \ (2.52 \cdot 10^{-3}) & \end{array}$$

Hereafter the nonlinear expression in (8) is referenced as (NL).

Comparison with previous methodologies

Once we have estimated the SDP parameter in (5) and the nonlinear function in (8), we will compare those results with two approaches. Firs, we will use the Blattberg–Hoch (B-H) '50% model, 50 % manager' as a benchmark (Blattberg and Hoch, 1990), where

Copyright © 2010 John Wiley & Sons, Ltd.

$$\begin{aligned} y_k &= 0.5 \cdot \mathrm{SF}_k + 0.5 \cdot (\mathrm{SF}_k + \mathrm{Adj}_k) + \nu_k \\ &= \mathrm{SF}_k + 0.5 \cdot \mathrm{Adj}_k + \nu_k \end{aligned} \tag{10}$$

We also analyze the optimal adjust (OA) model proposed by Fildes *et al*. (2009) that can be expressed using the dummy variable X_d defined in (9) such as

$$y_k = \gamma_1 \cdot \mathrm{SF}_k + \gamma_2 \cdot \mathrm{Adj}_k + \gamma_3 \cdot \mathrm{SF}_k \cdot X_d + \gamma_4 \cdot \mathrm{Adj}_k \cdot X_d + \nu_k \tag{11}$$

The estimates of the model described in (11) are given below:

$$\begin{aligned} &\hat{\gamma}_1 = 0.96\ \ (0.014) \quad \hat{\gamma}_2 = 0.81\ \ (0.073) \\ &\hat{\gamma}_3 = 0.07\ \ (0.018) \quad \hat{\gamma}_4 = -0.42\ \ (0.076) \end{aligned}$$

In order to compare models (8), (10) and (11) we related them to the general equation in (5). Assuming that the system forecast weight of the aforementioned models is approximately 1, it is possible to plot in the same graph the adjustments weight computed by the different models. For instance, α_1 and α_2 are defined as the SF and adjustments weights, respectively in (5). The equivalent of α_1 in the nonlinear model (8) is β_1 and the equivalent of $\alpha_2(\mathrm{Adj}_k)$ is given by $\alpha_2(\mathrm{Adj}_k) = \beta_2 + (\beta_3 + \beta_4 \cdot e^{-\beta_5 \cdot \mathrm{Adj}k})X_d$. Furthermore, we can see that $\hat{\beta}_1 = 0.949 \approx 1$.

Figure 7 depicts the estimation of the adjustment weight accomplished by (8) by a solid line; the non-parametric SDP estimation is depicted by a dashed line and the OA model in (11) by a dotted line. The dash-dot line shows the Blattberg and Hoch model described in (10).

According to the non-parametric SDP estimation shown in Figure 7, the explanatory weight of the managerial adjustments depends on its sign. Basically, without the SDP guidance one might consider as a starting point the adjustments average (Blattberg and Hoch, 1990) to ponder the influence of adjustments on forecasting accuracy. Nevertheless, the analysis carried out by Fildes *et al*. (2009) over different companies shows that adjustments accuracy was asymmetric with respect to its sign; i.e., negative adjustments were shown to be more precise than positive ones. Effectively, the OA model proposed in that reference is a better approximation of the nonlinear nature of the adjustment process. Nonetheless, the OA model only allows the variation of α_2 between constant values. This restriction is valid for negative adjustments (see Figure 7) but it is apparently not the best method to describe positive adjustments in relation to the non-parametric SDP. In order to resolve this limitation, the nonlinear function given by (8) is proposed, which models the negative adjustments with a constant (as the OA model) but it uses an exponential function to describe the positive adjustments. Note this nonlinear function is derived from the non-parametric SDP estimate.

Model validation

In this section predictive validation is used to compare models, where we expect that if a better description of the adjustment process is offered by the nonlinear model(s) these models should contribute to reducing the forecasting error compared to the simpler linear models. For this purpose, 20% of the data (582 triplets) constituted by the last months of each SKU, which were not used for the parameter estimation of the models, were employed as the hold-out sample to compare the performance of the proposed models. This hold-out sample design results in a more demanding experiment than selecting 20% of the data randomly (Fildes *et al*., 2009).

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

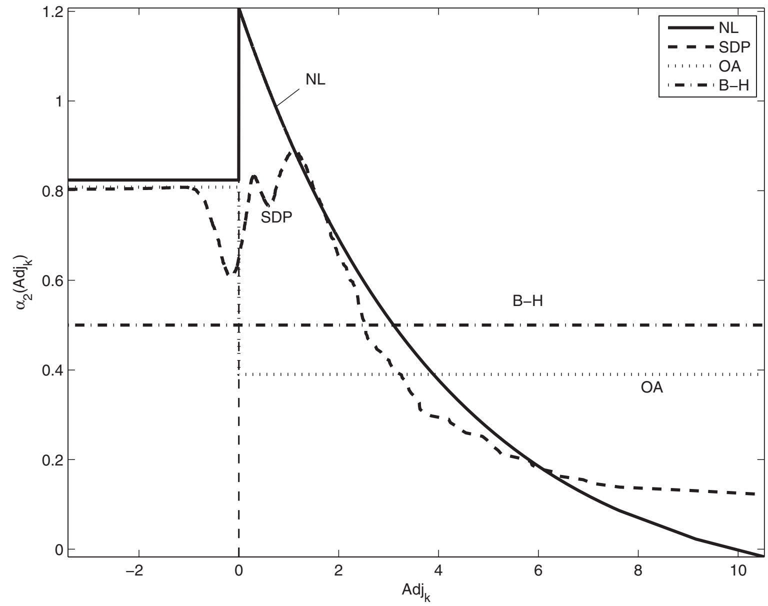

Figure 7. Adjustment weight estimation accomplished by the nonlinear (NL) model (8), the state-dependent parameter (SDP) model, the optimal adjust (OA) model (11) and the Blattberg and Hoch (B-H) model on normalized data (10)

Table V. Mean of MAPE and MdAPE for the validation dataset

Adjustment	No. of observations	Error (mean)	System forecast	Final forecast	SDP	NL	Optimal adjust	Blattberg–Hoch
Positive	235	MAPE	**27.04**	33.97	28.13	27.79	27.91	27.71
		MdAPE	25.09	26.78	24.70	24.22	23.71	**23.16**
Negative	125	MAPE	70.72	36.58	39.90	**36.14**	37.10	50.69
		MdAPE	68.06	35.49	39.30	**35.15**	36.07	49.15
None	222	MAPE	35.28	35.28	35.92	**33.06**	33.44	35.28
		MdAPE	24.34	24.34	24.77	**22.82**	23.07	24.34
Overall adjusted	360	MAPE	39.00	30.66	28.75	**27.46**	28.06	31.95
		MdAPE	26.38	21.30	20.40	**20.15**	20.21	21.34
Total	582	MAPE	38.52	31.08	30.12	**28.53**	29.34	32.61
		MdAPE	23.73	19.09	18.26	**18.23**	18.78	19.46

Table V shows the mean(MAPE) and mean(MdAPE) on the validation dataset. In the lower part of the table we can see the overall performance of the methods analysed. In order to get a deeper insight into the adjustment sign influence, the results have been separated according to the adjustment sign. Essentially, the final forecast beats the system forecast (quite substantially for negative adjustments). The NL method outperforms the non-parametric 'state-dependent model' and in particular the linear models. The number of observations taken into account are shown in the second column. Again, the NL method delivers very promising results except for positive adjustments.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Table VI. MAE for the normalized validation dataset

Adjustment	System forecast	Final forecast	SDP	NL	Optimal adjust	Blattberg–Hoch
Positive	0.929	0.725	0.699	**0.698**	0.737	0.710
Negative	0.747	0.459	0.457	**0.455**	0.456	0.560
None	0.579	0.579	**0.563**	0.564	0.564	0.579
Overall adjusted	0.866	0.633	0.615	**0.613**	0.640	0.658
Total	0.757	0.612	0.595	**0.594**	0.611	0.628

Table VII. Standard deviation of the absolute error for the normalized validation dataset

Adjustment	System forecast	Final forecast	SDP	NL	Optimal adjust	Blattberg–Hoch
Positive	0.912	0.823	0.693	0.686	0.672	**0.651**
Negative	0.611	0.446	0.433	0.428	**0.426**	0.461
None	0.561	0.561	**0.54**	0.549	0.548	0.561
Overall adjusted	0.825	0.726	0.626	0.620	0.613	**0.597**
Total	0.749	0.669	0.595	0.594	0.590	**0.585**

However, the mean(MAPE) and mean(MdAPE) are not well suited to measure the positive adjustments performance for the reasons given above ('Exploratory data analysis'). As previously, the normalized data were employed to compute the MAE on the validation dataset, shown in Table VI. From this table we can corroborate the good performance of the NL model for positive adjustments as well. Additionally, Table VII shows the standard deviation of the absolute error in order to analyse the forecasting error dispersion. Assessing Tables VI and VII we can conclude that the nonlinear model(s) proposed achieve a lower forecasting error and also reduce the variance of such errors compared to the final forecast.

Figures 8 and 9 depict the MAE against the size of positive and negative adjustments for the validation dataset, respectively. From these figures we can observe that the FF is more accurate than the SF. Furthermore, the margin of improvement is more visible for larger adjustments. Note that these findings are consistent with those reached in the exploratory data analysis assessing the SF and FF, above ('Exploratory data analysis').

Regarding the proposed models for positive adjustments in Figure 8, the NL approach outperforms the other models. In addition, the OA model works slightly worse than the FF and the B-H models. Since neither the OA model nor the B-H model is sufficiently flexible to describe the nonlinear process associated with positive adjustments (see Figure 7), the differences found between them may not be systematic. This means that this result may not be consistent for another sample.

In relation to the negative adjustments shown in Figure 9, the analysed methods, except for the B-H technique, achieve a similar performance, where the NL method outperforms them slightly. The B-H method performs rather worse than the other methods. This poor performance can be explained by analysing negative adjustments in Figure 7. From this figure we can see that the weight adjustment $\alpha_2(\mathrm{Adj}_k)$ suggested by the B-H for negative adjustments is lower than the one computed by the other methods based on parameter estimation from the data. Therefore, this discrepancy results in a larger forecasting error for the B-H method.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

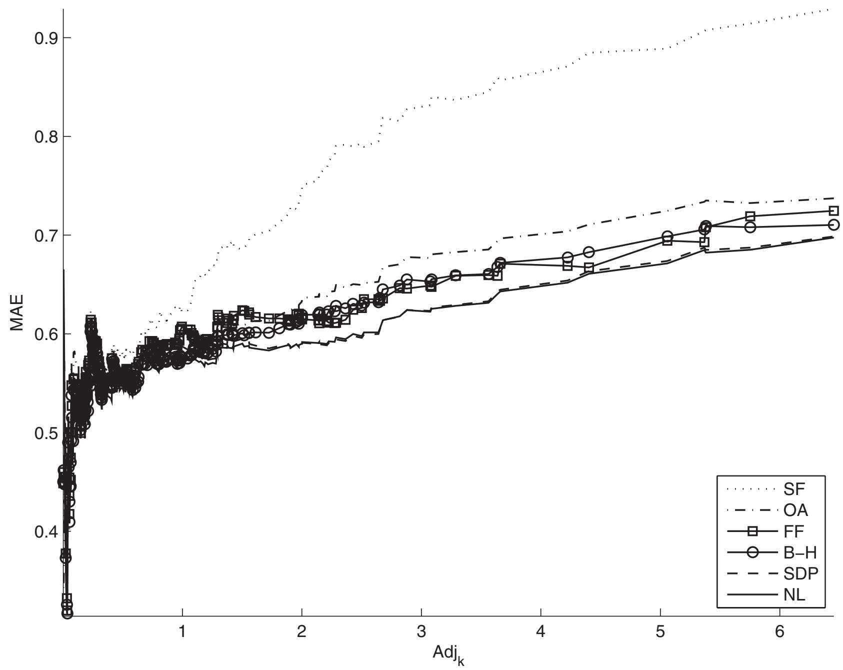

Figure 8. MAE for normalized positive adjustments on the validation dataset

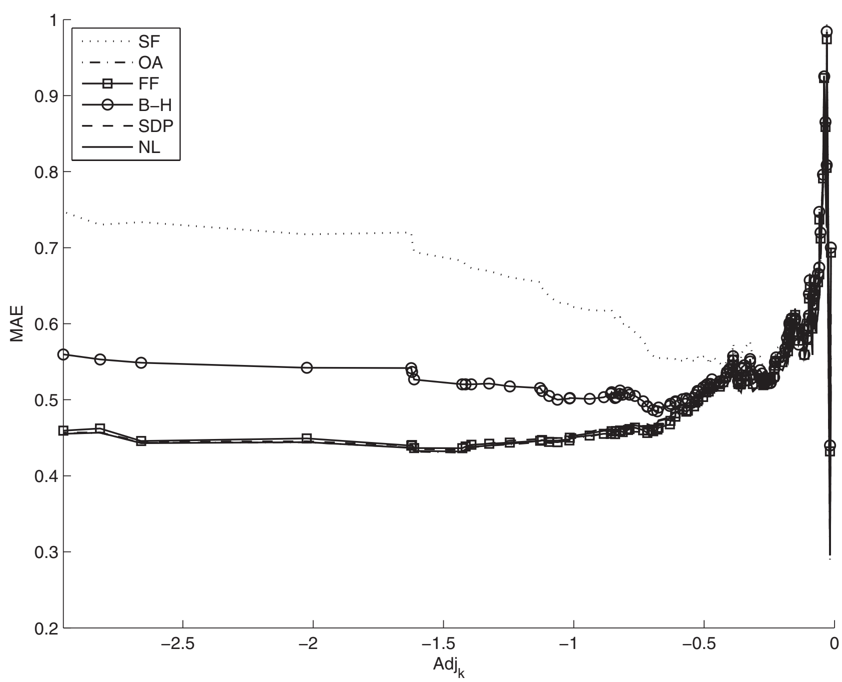

Figure 9. MAE for normalized negative adjustments on the validation dataset

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Considering the case data where no managerial adjustment is made, the NL method also improves forecasting accuracy. This indicates that there is room to improve the SF design.

Finally, it is interesting to note that NL beats SDP, but only slightly since NL is estimated in a more efficient way than the SDP.

Practical considerations

The previous results show that judgmental forecast is nonlinearly dependent on the adjustments size. In fact, such nonlinearities can be explained by means of both parametric and non-parametric approaches. However, how can we use these models to improve the forecasting accuracy in organizations?

Mechanical integration is an alternative to correct automatically the judgmental forecast as well as the statistical forecast (Goodwin, 2000, 2005). Nevertheless, such an alternative may face several problems. On behalf of the forecasters, they may find it less motivating to adjust the forecast, putting less effort into performing the task (Belton and Goodwin, 1996); or they may attempt to pre-empt the corrections by modifying their adjustments. Furthermore, the origin of the biases can also be time-varying (Fildes *et al*., 2009).

Another option is to incorporate the proposed model(s) into FSSs. In fact, the nonlinear weights derived from the model(s) we have proposed might have implications for the design of FSSs. If such systems are to be effective in supporting judgmental interventions, they need to help users distinguish between the various sources of information, guiding them in weighting reliable and major pieces of information much more effectively. Nonetheless, some practical challenges may come up in order to convince FSS users to modify their adjustments based on a relatively complex technique.

CONCLUSIONS

The use of SDP estimation was exploited in a new application in order to understand the nonlinear complexity involved in judgmental adjustments. These adjustments are of paramount importance in numerous companies since they have a direct and substantial influence on the forecasting accuracy of supply chain demand. Actual data sampled monthly were collected from a manufacturing company to verify the approach. In fact, an SDP estimate was the baseline to formulate a nonlinear model which was employed to reduce the forecasting errors on the basis of a better description of the nonlinearity observed in managerial adjustments. In order to compare the performance of the methods, several well-known error measures were considered, including MAPE and MdAPE. Nonetheless, it was shown that normalization of the data can be very helpful in gaining a better understanding of the influence of adjustment size. This normalization allows us to use another error measure (MAE) that avoids the heavy penalization which the percentage errors apply to over-forecast errors. Therefore, this MAE gives a better description of the methods' relative performance when positive adjustments are considered.

Putting these together, several conclusions can be drawn: (i) there were no big differences between the methods analysed for small adjustments; (ii) FF forecasts outperform SF when adjustments are larger; (iii) the NL model proposed on the basis of a non-parametric SDP estimation was shown to provide a description of the nonlinear behaviour involved in the adjustment process of the company analysed by means of an efficient estimation. This ability was translated into a reduction of the forecasting error on the hold-out sample data.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Since there is considerable potential in SDP models in this field, further research is needed to analyse a wider range of datasets from more companies with different features, as Fildes *et al*. (2009) have already shown that companies differ in their responses to information when making adjustments.

ACKNOWLEDGEMENTS

J. R. Trapero was partly supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme and La Consejería de Educación y Ciencia de la Junta de Comunidades de Castilla-La Mancha.

REFERENCES

Belton V, Goodwin P. 1996. On the application of the analytic hierarchy process to judgmental forecasting. *International Journal of Forecasting* **12**: 155–161.

Blattberg RC, Hoch SJ. 1990. Database models and managerial intuition: 50% model + 50% manager. *Management Science* **36**: 887–899.

Bryson A, Ho Y. 1969. *Applied Optimal Control, Optimization, Estimation and Control*. Blaisdell: New York.

Eroglu C, Croxton KL. 2010. Biases in judgmental adjustments of statistical forecasts: the role of individual differences. *International Journal of Forecasting* **26**: 116–133.

Fildes R, Goodwin P, Lawrence M. 2006. The design features of forecasting support systems and their effectiveness. *Decision Support Systems* **42**: 351–361.

Fildes R, Goodwin P, Lawrence M, Nikolopoulos K. 2009. Effective forecasting and jugdmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting* **25**: 3–23.

Franses PH, Legerstee R. 2009. Properties of expert adjustments on model-based SKU-level forecasts. *International Journal of Forecasting* **25**: 35–47.

Franses PH, Legerstee R. 2010. Do Experts' Adjustments on Model-Based SKU-Level Forecasts Improve Forecast Quality? *Journal of Forecasting* **29**: 331–340.

Goodwin P. 2000. Correct or combine? mechanically integrating judgmental forecasts with statistical methods. *International Journal of Forecasting* **16**: 261–275.

Goodwin P. 2005. How to integrate management judgment with statistical forecasts. *Foresight* **1**: 8–12.

Harvey A. 1989. *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge, UK.

Jakeman AJ, Young PC. 1984. Recursive filtering and the inversion of ill-posed causal problems. *Utilitas Mathematica* **35**: 351–376.

Kahneman D, Lovallo D. 1993. Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Science* **39**: 17–31.

Kalman RE. 1960. A new approach to linear filtering and prediction problems. *ASME Transactions: Journal of Basic Engineering* **83-D**: 95–108.

Lawrence M, Goodwin P, OConnor M, Önkal D. 2006. Judgmental forecasting: a review of progress over the last 25 years. *International Journal of Forecasting* **22**: 493–518.

Makridakis S. 1993. Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting* **9**: 527–529.

Mathews B, Diamantopoulos A. 1990. Judgmental revision of sales forecasts: effectiveness of forecast selection. *Journal of Forecasting* **9**: 407–415.

Mello J. 2009. The impact of sales forecast game playing on supply chains. *Foresight* **13**: 13–22.

Moon MA, Mentzer JT, Smith CD. 2003. Conducting a sales forecasting audit. *International Journal of Forecasting* **19**: 5–25.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Pedregal DJ, Young PC. 2002. Statistical approaches to modelling and forecasting time series. In *A Companion to Economic Forecasting*, Clements MP, Hendry DF (eds). Blackwell: Oxford; 69–104.

Syntetos AA, Nikolopoulos K, Boylan JE, Fildes R, Goodwin P. 2009. The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics* **118**: 72–81.

Taylor CJ, Pedregal DJ, Young PC, Tych W. 2007. Environmental time series analysis and forecasting with the captain toolbox. *Environmental Modelling and Software* **22**(6): 797–814.

Young PC. 2006. The data-based mechanistic approach to the modelling, forecasting and control of environmental systems. *Annual Reviews in Control* **30**: 169–182.

Young PC, Garnier H. 2006. Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems. *Enviromental Modelling and Software* **21**: 1055–1072.

Young PC, McKenna P, Bruun J. 2001. Identification of non-linear stochastic systems by state dependent parameter estimation. *International Journal of Control* **74**: 1837–1857.

Authors' biographies:

Juan R. Trapero is a Postdoctoral researcher in the School of Management, Lancaster University (UK), funded by a Marie-Curie Intra-European Fellowship. He obtained the Ingeniero Industrial degree in 2003 from Universidad de Castilla-La Mancha, (UCLM, Spain); his M.B.A. in 2004 from the UCLM; and his Ph.D. in February 2008 from the UCLM. His research interests include algebraic identification and estimation in continuous-time of dynamical linear and non-linear systems and State Space methods applied to time series, forecasting and control.

Robert Fildes is Professor of Management Science in the School of Management, Lancaster University, and Director of the Lancaster Centre for Forecasting. He has a mathematics degree from Oxford and a Ph.D. in statistics from the University of California. He was co-founder of the Journal of Forecasting in 1981 and of the International Journal of Forecasting in l985. For ten years from 1988 he was Editor-in-Chief of the IJF. He was president of the International Institute of Forecasters between 2000 and 2004. His current research interests are concerned with the comparative evaluation of different forecasting methods, the implementation of improved forecasting procedures in organizations and the design of forecasting systems.

Andrey Davydenko is a Ph.D. student in the Department of Management Science at Lancaster University. He holds a Candidate of Science degree in mathematical methods in economics. He has worked in the area of the development and software implementation of statistical techniques for business forecasting. His current research focuses on the composite use of judgmental and statistical information in forecasting support systems.

Authors' addresses:

Juan R. Trapero, Robert Fildes and **Andrey Davydenko**, Department of Management Science, Lancaster University, Lancaster LA1 4YX, UK.

Copyright © 2010 John Wiley & Sons, Ltd.