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

brought to you by CORE

provided by Lancaster E-Prints

Lancaster University Management School
Working Paper
2006/054

Producing more efficient demand forecasts

Robert Fildes, Paul Goodwin, Michael Lawrence and Konstantinos Nikolopoulos

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

© Robert Fildes, Paul Goodwin, Michael Lawrence and Konstantinos Nikolopoulos
All rights reserved. Short sections of text, not to exceed
two paragraphs, may be quoted without explicit permission,
provided that full acknowledgement is given.

The LUMS Working Papers series can be accessed at http://www.lums.lancs.ac.uk/publications/
LUMS home page: http://www.lums.lancs.ac.uk/

Producing more efficient demand forecasts

Robert Fildes*

Dept. Management Science, Lancaster University, LA1 4YX, UK

Email: **R.Fildes@lancaster.ac.uk**

Paul Goodwin,

School of Management, University of Bath,

Email: **P.Goodwin@bath.ac.uk**

Michael Lawrence

School of Information Systems, University of New South Wales, Australia

Email: **michael.lawrence@unsw.edu.au**

Konstantinos Nikolopoulos

Dept. Management Science, Lancaster University, LA1 4YX, UK

and Manchester Business School, Manchester M15 6PB

Email: **kostas.nikolopoulos@mbs.ac.uk**

Acknowledgements

This research was supported by Engineering and Physical Sciences Research Council (EPSRC) grants GR/60198/01 and GR/60181/01. Alastair Robertson contributed substantially to the statistical modelling.

* Corresponding author

21 November 2006

Producing more efficient demand forecasts

Abstract

Demand forecasting is a crucial aspect of the planning process in supply-chain companies. The most common approach to forecasting demand in these companies involves the use of a simple univariate statistical method to produce a forecast and the subsequent judgmental adjustment of this by the company's demand planners to take into account market intelligence relating to any exceptional circumstances expected over the planning horizon. Based on four company case studies, which included collecting more than 12,000 forecasts and outcomes, this paper examines: i) the extent to which the judgmental adjustments led to improvements in accuracy, ii) the extent to which the adjustments were biased and inefficient, iii) the circumstances where adjustments were detrimental or beneficial, and iv) methods that could lead to greater levels of accuracy. It was found that the judgmentally adjusted forecasts were both biased and inefficient. In particular, market intelligence that was expected to have a positive impact on demand was used far less effectively than intelligence suggesting a negative impact. The paper goes on to propose a set of improvements that could be applied to the forecasting processes in the companies and to the forecasting software that is used in these processes.

<u>Keywords</u>

Forecasting accuracy; judgment; heuristics and biases; supply chain; forecasting support systems; practice

1. Introduction

Supply chain planning is usually reliant on demand forecasts at stock keeping unit (SKU) level. The accuracy achieved for these forecasts has consequences for companies at all levels of the supply chain from retailer to raw materials supplier, and even for companies whose final product is 'make-to-order' (Yelland, 2006) . Errors at each stage are potentially amplified, resulting in poor service or excess inventory levels. The forecasting problem is difficult due to the inter-related nature of the data series with outliers, level and trend shifts (Fildes and Beard, 1992) and impacted by the complexities of the market and general economic environment. These data difficulties are compounded by the huge number of SKUs that often need to be forecast each period.

Because of the size and complexity of the forecasting task, it is generally impossible for all SKUs to be given individual attention by demand planners. The most common approach to forecasting demand in support of supply chain planning involves the use of a forecasting support system (FSS) which incorporates a simple univariate statistical method to produce an initial forecast. For key products, these initial statistical forecasts (hereafter called the 'system' forecasts) are reviewed and may be adjusted by the company's demand planners to take into account exceptional circumstances expected over the planning horizon (also referred to as 'market intelligence' or MI) or possibly to correct perceived inadequacies in the system forecast. This process is usually carried out in a committee setting where representatives from marketing, sales, production and logistics agree the 'final forecast': a combination of a statistical forecast and managerial judgement.

Improved demand forecasting accuracy can lead to significant monetary savings, greater competitiveness, enhanced channel relationships and customer satisfaction (Fildes and Beard, 1992, Moon et al., 2003) . Management clearly appreciates the importance of the forecasting function and allocates significant computing and management resources to the activity (Moon et al., 2003) . Despite the allocation of these resources, there is evidence that in many organisations it is carried out poorly (Lawrence et al., 2000, Moon et al., 2003) with forecast accuracy often not significantly better than from the naïve (no change) forecast. In addition, forecasters are often untrained in forecasting methods (Klassen and Flores, 2001) , ignorant of (or even denied) relevant market information and their performance is poorly measured (Moon et al., 2003) . In particular, there is apparently an over-reliance on the use of informal judgement at the expense of statistical methods (Fildes and Beard, 1992, Moon et al., 2003) . For example, judgement was the preferred method of forecasting in Sanders and Manrodt's (1994) survey of US forecasters . Even where quantitative forecasts were produced, they were usually adjusted. The respondents gave as their primary reason for adjustment, concern for accuracy and the need to incorporate knowledge of the environment. Although the statistical forecast has to be adjusted when abnormal 'one-off' events are known to be about to take place (e.g. a promotion) it appears that adjustments to the forecasts are much more frequent, reflecting the forecaster's concern for accuracy and the belief that the statistical model 'has not got it right'.

Blattberg and Hoch (1990) in their examination of five companies where forecasts were combined (rather than adjusted), identified this adjustment process as an important issue to study. Perhaps more tellingly, ad hoc surveys of practitioners and software companies regularly identify the adjustment process as a key element in attempting to ensure accurate SKU forecasts (Worthen, 2003) . However, the topic has generated little organisationally-based research and little is known about the effect of the types of adjustment (e.g. large or small, positive or negative) on accuracy or the extent to which the resulting forecasts are unbiased and efficient (i.e. make optimal use of available information).

This paper uses extensive data gathered from four supply-chain companies to address these questions and to consider how the process of adjustment can be made more effective. The next section considers the literature on adjustment and proposes hypotheses to extend our knowledge of the adjustment process. Section 3 describes the forecasting processes in the four companies and gives details of the frequency and nature of the judgmental interventions. Section 4 examines the detailed hypotheses developed in the literature review, while Section 5 evaluates some potential solutions to the problems that have been identified. The final section offers our recommendations for improvements in both forecasting software and in organisational forecasting processes.

2. Literature Review and Hypotheses

There is substantial evidence from the economic forecasting literature that statistical forecasts can be made more accurate when experts judgmentally adjust them to take into account the effects of special events and changes that were not incorporated into the statistical model (Donihue, 1993, McNees, 1990, Turner, 1990) . However, few studies have investigated judgmental adjustment in the context of company forecasts of the demand for SKUs. The exceptions were four studies all based in the same company by Mathews and Diamantopoulos (Mathews and Diamantopoulos, 1986, Mathews and Diamantopoulos, 1989, Mathews and Diamantopoulos, 1990, Mathews and Diamantopoulos, 1992) . These showed that judgmental 'revision' tends to improves accuracy even though sometimes only marginally, but also tends to introduce bias.

Experimental evidence generally suggests that forecasters often make unnecessary judgmental adjustments to statistical forecasts (Lawrence et al., 2006) . In particular, they make adjustments even when they do not possess extra information about special events. There is some evidence that this occurs because the forecasters see false patterns in the noise associated with time series (O'Connor et al., 1993) and, as a result, their interventions reduce accuracy. Lim and O'Connor (1995) found that this tendency persisted, despite a system display showing that the adjustments were reducing accuracy and costing them money. However, experimental evidence also suggests that, when an adjustment is made on the basis of events not reflected in the statistical forecast (e.g. a forthcoming sales promotion), it is likely to improve accuracy as long as the information about the event is reliable and its effect is not disguised by noise (Goodwin and Fildes, 1999, Lim and O'Connor, 1996) Forecast adjustments, by experts in a company environment with access to reliable market intelligence, are likely to yield greater benefits than those obtained in experiments by the student subjects. Thus, we hypothesize:

H_1: Judgemental forecast adjustments improve forecast accuracy.

H_{11}: Judgemental forecast adjustments improve forecast accuracy more under conditions of high reliability information (compared to low reliability information).

H_{12}: Judgemental forecast adjustments improve forecast accuracy more under low noise than high noise conditions.

While accuracy is a most important property for a forecast, two further properties are also important: bias and efficiency. Bias, itself, can be decomposed into two components. Mean bias is a systematic tendency for the forecast to be less than or greater than the actual. Regression bias is the extent to which the forecasts systematically fail to track the actual observations. For example, forecasts may tend to be too high when outcomes are low and too low when outcomes are high (Theil, 1971) . Efficiency is the property that forecasts optimally incorporate relevant new information as it becomes available. While adjustments might improve accuracy they may not be unbiased or efficient. Lawrence et al. (2000) found that judgemental forecasts made by 13 large manufacturing organisations were generally neither unbiased nor efficient. They found many forecasters faced a situation of asymmetric management incentives (depending on the sign of the forecast error) which resulted in a biased forecast. In addition, over-stocking costs and under-stocking costs are not generally equal and this may lead to forecast bias. Sanders and Manrodt (1994) found 70.4% of US respondents to a survey on forecasting practices preferred to 'under-forecast' while Stewart in Fildes et al. (2003) citing evidence from a UK survey, found an equal predisposition for over and under forecasting. Fildes (1994) in a study of product managers across business units of a major multinational found that 92% of forecasters thought their forecasts were influenced by organisational politics, although the direction of influence was not explored. What seems to be clear is that political pressures may undermine accuracy (Deschamps, 2004) . Thus, although bias may be intentionally introduced, since inventory restocking processes assume that the forecast is unbiased, a biased forecast inevitably increases costs in supply chain management. Apart from Mathews and Diamantopoulos (1990) , who found evidence that adjustments introduced bias, few studies of unbiasedness and efficiency have been carried out on forecast *adjustments* to SKU data. However, based on the limited evidence available to date we hypothesise:

H_{21}: The adjusted forecasts are biased.

H_{22}: The adjusted forecasts are inefficient.

Evidence has also accumulated on additional factors beyond the strength of the external information and the noise that may influence the effectiveness of judgemental adjustments. Optimism bias has been one of the most researched biases (see: for example: Flyvbjerg et al. (2003)). In optimism bias, values viewed as positive are over-forecasted and values viewed as negative are under-forecasted. For example, in forecasting associated with a new investment, the cost of the project will be under-forecasted while its benefits will be over-forecasted. This bias has been shown to impact a wide variety of forecasts including security analysts forecasts (Helbok and Walker, 2004) , project time prediction (Buehler and Griffin, 2003) and capital budgeting (Flyvbjerg et al., 2003) . It is believed so prevalent that, for example, the UK Treasury has issued guidelines aimed at minimising its budgetary impact[1]. Based on this evidence we anticipate that product managers planning a major promotion would be likely to over forecast the sales of the promoted product while under forecasting any possible negative impacts on other products. Thus, both

[1] See: www.dh.gov.uk/.../ChangesTreasuryGreenBookArticle).

positive are negative adjustments are hypothesised to be biased upward (in the case of negative adjustments this implies that. the forecasts will be insufficiently reduced). This suggests the following hypotheses:

H_3: The improvement resulting from judgemental forecast adjustments will depend on the direction of the adjustments.

H_{31}: Positive adjustments will typically be too large (i.e. the forecast will be greater than the actual).

H_{32}: Negative adjustments will typically be too little (i.e. the forecast will be greater than the actual).

A consistent finding in judgemental extrapolation of time series, is that subjects tend to damp both up and down trends with down-trends damped more than up-trends (Eggleton, 1982, Lawrence and Makridakis, 1989, O'Connor et al., 1997) observed that subjects seemed less sure of down trends as they both widened their confidence bounds and damped their most likely estimates more than for up trends. O'Connor et al. (1997) confirmed the difficulty presented by down trends and that the forecasters' behaviour suggests an anticipation of a reversal in slope. Accordingly we hypothesise:

H_4: Adjustments to down trending series tend to be either damped or to reverse the trend more so than for upward trending series.

3. The Study

Data have been collected at SKU level from four companies, three in manufacturing with monthly data (pharmaceuticals (A), food (B), and household products (C)), and one retailer (D) forecasting weekly (where data on two separate product groups were available). For companies A-C all SKUs in the company were examined. For the retailer, data on two product groups each supplied by an individual manufacturer were made available (D1 and D2). The data included one-step ahead final forecasts, the corresponding actual outcomes, as well as the statistical system forecast. Those without the required continuous forecast history were excluded. Low volume SKUs, defined as those with actual outcomes or system forecasts less than 10 units, were eliminated from this analysis (and are examined in a separate paper). In addition, observations where the final or system forecast is zero have been eliminated as these were thought to result from special circumstances like the particular SKUs being withdrawn from the market.

In all four organisations the forecasting process was observed and discussions held with the principal forecasters. Each organisation used a broadly similar process to estimate their final forecasts. At the start of each forecasting period, the statistical ‘system’ forecasts are produced through the forecasting support system (FSS), based on variants of exponential smoothing or Focus forecasting (Gardner and Andersen , 1997) . A forecasting meeting generally comprising forecasting, marketing, production and sales personnel, examined the resulting system forecast in the light of various pieces of marketing and other information and agreed the final forecast. Also, particular aspects of recent forecasting performance such as a large error might be drawn to the group’s attention and highlighted in a summary screen in the FSS. The information used in these meetings included promotional plans, weather forecasts, stock information, customer information and company plans as well as evidence of recent forecasting performance. Such information might be available through the internet, printed report, phone or a scheduled meeting. All company forecasters interviewed affirmed that an objective of the forecasting process was producing accurate forecasts. and said that their final forecasts were not subsequently changed by more senior management (Fildes and Hastings, 1994) . For all companies, we only consider in this paper one period ahead forecasts.

Table 1 summarises the data base, showing the number of observations containing i) the statistical forecast, ii) the final forecast and iii) the actual value (the ‘triple’). Table 1 also shows the percentage of forecasts adjusted and the number of SKUs contained in the data set. Companies A-C, all manufacturers making monthly forecasts (identified in this study as the G1 organisations), adjust a substantially greater percentage of forecasts than Company D1 and D2, the retailer making weekly forecasts (identified as the G2 organisations). We have grouped the organisational data in this way as analysis shows that data values have common characteristics but any substantive individual company differences will also be noted. When modelling the data we have broken each company’s data into an estimation set used in model development and a test set for model validation.

Companies	Data	Available data	Total complete triples	% adjusted	No. of skus
A	Monthly	Years 2003-2005, Months 1-12	5165	63%	213
B		2004, Months 5-12, 2005, Months 1-12	2360	92%	296
C		2004, Months 3-11, 2005, Months 1-12	2803	61%	244
D1	Weekly	2004, Weeks 1-52, 2005, Weeks 1-52	12400	14%	191
D2		2004, Weeks 1-52, 2005, Weeks 1-52	43340	8%	592
Total			66068	19%	1536

Table 1 Data base of forecasts by company (certain periods are missing)

We take it as given that positive and negative adjustments are made when information is available which causes the forecaster to think that the system forecast will either under or overestimate demand respectively (when there are multiple sources of information giving indications in different directions we assume that the forecaster makes a judgment based on the relative strengths of the indications). We will therefore refer to information as being either positive or negative. Nevertheless, the organisational and psychological factors motivating a positive or a negative adjustment (or indeed, no adjustment) may differ markedly. In addition, negative adjustments are bounded below by zero. We have therefore split the sample into three sub-samples: no adjustment, positive adjustment and negative adjustment. Table 2 shows the mean and median sizes of the relative adjustments for each of these sub-samples where the relative adjustment is defined as 100 x (Final forecast –System forecast)/System forecast. In addition, we separate the results for the G1 and G2 organisations as the typical size of the adjustments is evidently different. The table shows that positive adjustments for both groups of organisations are very much larger than the negative adjustments and that the mean size of these adjustments ranges from 20 to 60 percent.

Org Group	Direction of Adjustment	Relative Adjustment	
		Mean	**Median**
G1	Positive adjustment	57.5%	20.2%
	Negative adjustment	21.7%	13.8%
G2	Positive adjustment	60.7%	32.6%
	Negative adjustment	40.3%	26.6%

Table 2 Mean and Median Relative Adjustment by Organisational Group and Direction of Adjustment.

Figure 1 presents the histogram of the relative adjustments for the G1 and G2 organisations. Both distributions are right skewed, however the distribution for the G2 organisations is far more platykurtic than the G1, that is the relative adjustments in the G2 data see more large adjustments whereas in the G1 data the vast majority fall in the small-size adjustments category.

Figure 1. Histogram of Relative Adjustments split by Organisational Group (12467 cases)

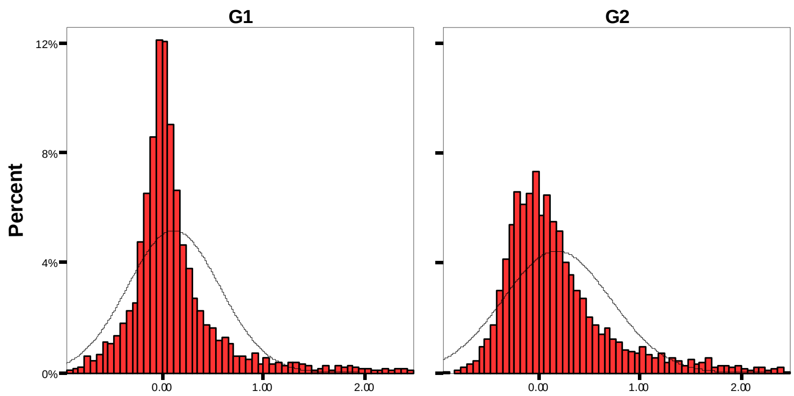

Figure 2 Histogram of Relative Adjustments by Company

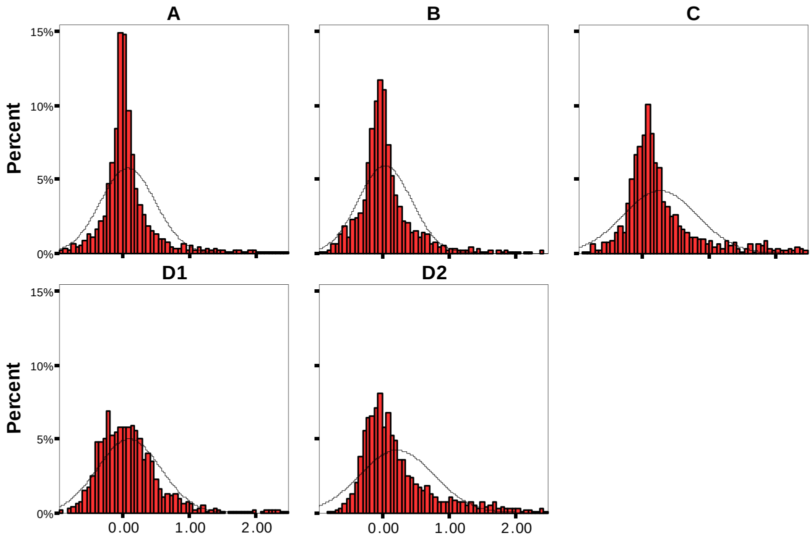

Figure 2 presents histograms of relative adjustment for the individual organisations. In group G1, Company B and C look quite similar while Company A differs as it is more leptokurtic (peaked). From the G2 data, company D2 is rather leptokurtic and skewed.

4. Accuracy, Unbiasedness and Efficiency in the Adjusted Forecasts

Accuracy

Although the measurement of forecasting accuracy is controversial (Armstrong and Fildes, 1995, Clements and Hendry, 1995) within company settings the use of absolute percentage error measures is now general (Fildes and Goodwin, 2006) . However, the disadvantages of such measures are well known. In particular, they suffer from a lack of robustness to extremes (Armstrong and Collopy, 1992) despite our large number of observations. In evaluating the accuracy and bias of the statistical forecast (SFC) and the final forecast (FFC) we therefore report two more robust variants, the trimmed mean absolute percentage error (MAPE) and the median absolute percentage error (MdAPE). We have trimmed outlying values using a 2% trim. Other accuracy measures such as the relative absolute error of FFC compared to SFC have been calculate but tell the same tale and are therefore not reported.

A summary of accuracy for the G1 and G2 organisations is given in Table 3 where the errors of the system forecast, the final forecasts and a naive (random walk) forecast are compared. It can be seen that for negative information adjustments the final forecast MAPE is significantly lower (more accurate) than the statistical forecast (paired t-test, for both G1 and G2, $t > 15$, $p<0.001$). However, for positive information the opposite is true (paired t-test, for both G1 and G2, $t < -11$, $p<0.001$). Thus there is strong support for H_1 only for negative adjustments. Various other accuracy measures, such as the MdAPE, provide support for H_1 for both positive and negative adjustments, but only for the G1 companies. Hence we conclude there is conditional support for H_1.

MEAN

		No. Observations, N (N naïve)	Naïve % error	System fcast % error	Final fcast % error	Naive Mape	System fcast Mape	Final fcast Mape
G1	No adjust.	3174 (3068)	-9.7%	-44.3%	-44.3%	37.5%	60.5%	60.5%
	Positive info	4013 (3735)	-9.0%	5.7%	-29.6%	39.8%	29.1%	39.6%
	Negative info	3392 (3136)	-15.5%	-38.1%	-4.5%	41.4%	46.9%	26.6%
G2	No adjust.	50427 (49794)	3.2%	-4.6%	-4.6%	17.5%	19.3%	19.3%
	Positive info	3049	10.1%	-7.4%	-57.5%	27.6%	32.1%	64.9%
	Negative info	2409	-1.8%	-35.5%	-3.5%	24.5%	40.9%	28.5%

Median

		No. Observations, N (N naïve)	Naïve % error	System fcast % error	Final fcast % error	Naive Mdape	System fcast Mdape	Final fcast Mdape
G1	No adjust.	3174 (3068)	2.9%	-1.4%	-1.4%	22.2%	13.6%	13.6%
	Positive info	4013 (3735)	2.5%	9.8%	-9.4%	25.8%	20.2%	17.6%
	Negative info	3392 (3136)	1.0%	-14.0%	2.2%	25.2%	20.7%	15.7%
G2	No adjust.	50427 (49794)	-0.1%	-1.4%	-1.4%	12.7%	13.5%	13.5%
	Positive info	3049	12.1%	0.0%	-39.2%	20.9%	21.2%	43.2%
	Negative info	2409	4.8%	-22.3%	3.6%	14.7%	25.0%	20.9%

Table 3 Forecast Error by Organisational Group and information class (no adjustment, positive and negative information)

The naïve forecast for G2 is more accurate for all sub samples than either the system forecast or the final forecast as Lawrence, et al. (2000) also found. The poor showing of the system forecast compared to the naïve could suggest that forecast adjustment may be a response to perceived inadequacies of the system forecast rather than a response to market information. For G1 data, where the forecasts have been left unadjusted, the MAPE (but not the MdAPE) of the Naive forecast is considerably better than that of the system forecasts. This suggests that in a limited number of cases the forecasters may be failing to adjust when there is considerable information in the most recent actual so that adjustment *is* needed (i.e. the naïve forecast).

We now investigate how the accuracy of judgemental adjustments is affected by the reliability of the market information (H_{11}) and the noise in the data (H_{12}). It seems reasonable to assume that a forecast adjustment is made when a forecaster has market information that suggests the actual will be either greater or smaller than the system forecast value. A simple test of the worth of this market information is to investigate how frequently the forecaster has successfully picked the right direction to adjust the forecast. That is, for example, when the forecaster makes an upward adjustment, how often does the actual indeed exceed the system forecast? As hypothesised, the reliability of the market information will significantly impact the success of picking the right direction for the adjustment. We take as a surrogate variable for reliability, the relative size of the adjustment defined as the ratio of the adjustment size divided by the system forecast. The justification of this choice is that a large adjustment will only be undertaken when it is believed that the market information is highly reliable. When information is perceived to be less reliable the adjustment may be expected to be less because forecasters may 'hedge their bets' (O'Connor et al., 2001) .

Table 4 shows the percentage of times that the chosen direction for the adjustment is correct (a 'right sided adjustment'). The table also displays results for the low reliability and high reliability groups. The low (high) reliability group is comprised of those forecast adjustments where the relative adjustment is less than (greater than) the median value for each adjustment direction within each organisation. The results show a great difference in the percent of right sided adjustments with a range from a low of 36% to a high of 82%. For both G1 and G2 organisations the percentage of adjustments downward that prove to be correct is higher than upward adjustments. This suggests the forecasters are better at interpreting negative market information. As anticipated, the high reliability group has uniformly higher values of right sided adjustments suggesting that with more reliable market information the adjustment direction is more accurate. Overall, the values suggest the reason for the failure to improve accuracy when adjusting forecasts upward is due to misreading the market information about half the time when the information is of low reliability and about a third to a quarter the time when the information is high reliability. Further, the results suggest a very considerable scope for improving forecast adjustments by (i) not adjusting the forecast when the market information is of low reliability and (ii) enhancing the quality of interpretation of what forecasters see to be high reliability market information.

Adjust Direction	Reliability	Percent right sided G1 Companies	G2 Companies
Information	Low	55%	36%
positive	High	74%	61%
Information	Low	63%	77%
negative	High	79%	82%

Table 4 Percentage of Correct Direction Adjustments by Adjustment Direction and Reliability.

It is clearly not possible to determine the true level of noise in company time series (studies that have investigated the effect of noise on judgmental forecasts have usually relied on artificially generated series). We have therefore used the coefficient of variation of a series as a proxy measure for its noise level Each series has been categorised as belonging to either a high or low noise group depending on whether the value of its coefficient of variation is less than or greater than the median value for that organisation. We define our measure of forecast improvement FCIMP as:

FCIMP = 100*(⎮FFC - Actual⎮- ⎮SFC- actual⎮)/ actual

This variable is positive when the final forecast is less accurate than the system forecast (i.e. the adjustment has degraded the system forecast) and negative when the final forecast is more accurate (i.e. the adjustment has improved the forecast accuracy). The scaling by the actual makes the measure comparable to the percentage error measure. An ANOVA with noise and reliability as explanatory variables and firm as covariate run for the G1 organisations showed company ($F_{1,7064} = 1.5$, $p \approx 0.22$) not significant, noise ($F_{1,7064} = 5.3$, $p < 0.02$) as significant and reliability as highly significant ($F_{1,7064} = 67.9$, $p < 0.001$). For the G2 organisations the ANOVA results showed firm not significant ($F_{1,55733} = 1.9$, $p < 0.17$) while noise ($F_{1,55733} = 24.6$, $p < 0.001$) and reliability ($F_{1,55733} = 269.8$, $p < 0.001$) were both highly significant. Thus H_{11} and H_{12} are both supported for all organisations.

Unbiasedness

We use two alternative measures of bias. The percentage error (100 x (forecast – actual)/ actual) measures mean bias and is presented in Table 3. If the forecasts do not suffer from mean bias then the median (or mean) percentage error is close to zero. The median percentage errors for G1 and G2 for the no adjustment sub-sample at -1.4% demonstrates, in effect, no bias. For the G1 group, the system forecast is biased for both the positive and negative information sub-samples with the bias in the direction that we would expect given the fact that these sub-samples are conditioned on the expectation that the actual will be higher or lower than the system forecast. Thus these biases show that on average the managers have correctly identified the two groups for adjustment. However, after adjustment the median percentage error shows that the bias has been reversed in the case of positive information while for negative information the bias is almost zero. It is clear that the adjustment is going too far in the case of positive information. For G2 the identification of the series for positive adjustment appears to have been poorly performed as the system forecast bias is effectively zero, resulting after adjustment in a highly biased final forecast. The negative information group for G2 is well selected: the large negative system forecast bias is changed, after adjustment, to an almost unbiased final forecast. Thus in summary for both G1 and G2 the no adjustment and negative information groups have very low final forecast bias while the positive information group demonstrates a highly biased final forecast. Hence there is mixed support for H_{21}.

The second measure of unbiasedness is based on regression analysis and provides a joint test of mean and regression bias. Initially we test hypothesis H_{21}, that the adjusted forecasts are in aggregate unbiased, by estimating the following regression model:

$$Y_{ij,t} - F_{ij,t-1}(1) = \alpha_i + \beta_i F_{ij,t-1}(1) + \nu_{ijt} \qquad (M1)$$

where $Y_{ij,t}$ represents the actual sales in the ith company for the jth SKU in period t, and $F_{ij,t}(1)$ is the one period ahead final forecast made at period t. If the company forecasts are unbiased in aggregate, $\alpha_i=\beta_i=0$ for each company *i*. The normal regression assumptions cannot be expected to hold in that the data have very different levels and the errors can therefore be assumed to depend on the level of $Y_{ij,t}$ as well as the noise (as a starting assumption to be tested). The variables have therefore been normalised using the standard deviation of sales for each SKU. In addition, the data have been grouped by size of adjustment (in percentage terms) and models have been estimated for each sub-group.

The sub-sample of forecasts that have been adjusted contain many extreme observations. For example, 3.1% of observations have final forecast errors greater than 250%. We have initially removed all observations with such large errors from the analysis. In addition, when estimating model (*M1*) we have removed outliers (with absolute studentized residuals greater than 2.5) and high leverage points (using

DFIT statistics greater than $2\sqrt{p/n}$, the recommended cut-off value)[2]. Sensitivity testing was carried out on these data filtering decisions and the results we present appeared robust. After the adjustment procedure had been applied the residuals proved well-behaved (with normality and homoscedasticity accepted for most models), demonstrating the effectiveness of the normalisation process.

Initially the model was estimated jointly with the parameters (and error distribution) assumed constant across companies. Using a general linear model, with the companies as factors, and $F_{ij,t}$ as covariate, together with an interaction between them, leads to a rejection of the hypothesis that the biases are independent of company. The two separate sets of forecasts available for Company D were also tested for equality, while the other three companies were examined in a pairwise fashion, again leading to a formal rejection of statistical equivalence despite face similarities. From Table 3 the direction of the adjustment influences the bias of the adjustments. We have therefore estimated models separately.

Table 5 shows the results of estimating the separate equations (t-statistics are shown in parentheses). The common feature is the consistent overestimate of the actual outcomes for all companies. However the R^2 and β values are so small for companies A and B that the effect size is almost negligible. For all companies whatever the direction of the adjustment, unbiasedness is rejected. The table shows that in general the constant term is positive but all the β coefficients of the final forecasts are negative. At the mean level of the final forecasts this leads to an upwards bias. Thus on average, the adjusted forecasts are typically over-estimates for all the companies, sometimes very substantial (as shown in Table 3). The weekly retail (D1 and D2) forecasts are particularly biased while company A shows the least bias. Bias is more marked where the forecaster perceives there to be positive information about the market. In addition, if we analyse the biases by the size of the adjustment made (as a percentage of the actual) the biases tend to increase with the size of the adjustment, that is to say the forecaster increasingly over-weights the market information, the more important it is perceived to be.

[2] Leverage points are those that exert an 'undue' influence on the regression coefficients. An extreme leverage point would effectively determine the regression equation at the expense of describing the hundreds of other data points. The exact cut off value proved unimportant in reaching the final model specification.

Company	N (no. dropped)	Constant	β Coefficient	R^2
A +ve info	1616 (79)	-0.120 (-3.09)	-0.063 (-7.33)	2.5%
-ve info	1534 (85)	0.242 (7.46)	-0.041 (-4.93)	1.2%
B +ve info	880 (56)	0.174 (3.54)	-0.132 (-8.86)	6.7%
-ve info	1095 (63)	0.239 (7.37)	-0.058 (-5.74)	2.4%
C +ve info	1063 (58)	0.524 (8.46)	-0.246 (-14.76)	16.5%
-ve info	486 (30)	0.446 (6.33)	-0.137 (-7.63)	9.0%
D1 +ve info	880 (62)	0.145 (4.55)	-0.287 (-22.15)	29.6%
-ve info	775 (58)	0.747 (17.67)	-0.191 (-10.87)	10.0%
D2 +ve info	1880 (148)	0.325 (9.42)	-0.399 (-39.15)	44.9%
-ve info	1474 (102)	0.328 (10.57)	-0.166 (-12.43)	9.0%

Table 5 Forecast Bias by Company and Information Direction

Efficiency

With bias established the next question to examine is that of efficiency. The most immediate data the forecaster can bring to bear in making the adjustment is the time series history and the recent forecast errors. For all the companies, the latest observation is only known provisionally at the time of making the forecast. A suitable test of efficiency with this information set is the following model:

$$
\begin{aligned}
\textit{Define } e_{ij,t} &= Y_{ij,t} - F_{ij,t-1}(1) \\
&= \alpha_i SFC_{ij,t-1}(1) + \beta_{i,1} Y_{ij,t-1} + \beta_{i,,2} Y_{ij,t-2} + \gamma_1 e_{ij,t-1} + \gamma_2 e_{ij,t-2} + \nu_{ij,t} \qquad (M2)
\end{aligned}
$$

The constant term has been suppressed as the model's objective is to re-weight and combine the available information to explain the observed error. In the model, a significant coefficient for an independent variable is indicative that the forecaster has not used the information represented by that variable efficiently. Before the final estimated equation can be established the outliers must again be removed, and the errors rendered homoscedastic via normalisation (where as before the standard deviation of the actuals has been used). We also attempted to pool estimates using a general linear model but this again leads to rejection of the hypothesis that the error model is constant across companies. There is some limited evidence of seasonality in the errors but its removal through the inclusion of dummy variables affects the results only slightly.

All companies show signs of inefficiency in that their forecast error can be improved by better weighting the available information on the past observations and forecast errors (see table 6). Typically the current forecast assigns too little weight to the latest observed error: possibly due to the fact that it is not always known exactly when the forecast is made. There is no evidence that the most recent observation or the error are mis-weighted, but over (under) estimates from previous periods are not successfully taken into account. In addition, the system forecast is over-weighted. Interestingly, this result differs from that of Goodwin and Fildes (1999) who found in a laboratory study that forecasters ignored the statistical forecast when making judgmental interventions to take into account promotion effects. Overall, the results confirm those derived from examining bias – there is more inefficiency shown by the forecasters where positive market information has been perceived.

No uniform pattern of response is observed across the companies and this was confirmed through a generalised linear model. For example, company B fails to take into account both most recent sales and its previous forecast error. Company C shows the least signs of significant inefficiencies. The retailer D shows major inefficiencies particularly with regard to the most recent errors. The final column comments on how much extra explanatory power is gained by including the two period lags in the model. Due to autocorrelation in only one of the estimated models is the gain substantial. In summary, the final adjusted forecasts have been shown to be biased and inefficient overall: H_{21} and H_{21} are supported.

The direction of forecast adjustment

Market intelligence whether positive or negative, has been shown to improve forecasts for G1 organisations but not for G2. However, as indicated earlier, these benefits were affected by the direction of the intelligence. Table 7 provides an explanation of this.

Company		N (no. dropped)	Model Coefficients					R^2
			System forecast	Lag 1 Actual	Lag2 Actual	Lag 1 Error	Lag 2 Error	
A	Positive info	1314 (238)	-.083 (-4.25)	n.s.	.052 (2.65)	.123 (5.24)	n.s.	17.8
	Negative info	1240 (255)	-.047 (-2.69)	n.s.	n.s.	n.s.	.085 (4.05)	10.4
B	Positive info	706 (89)	-.220 (-6.98)	.153 (4.91)	n.s.	.186 (4.74)	.113 (2.75)	24.5
	Negative info	812 (128)	-.191 (-10.2)	.206 (8.97)	n.s.	n.s.	n.s.	16.7
C	Positive info	799 (141)	-.084 (-2.75)	n.s.	n.s.	n.s.	.118 (4.22)	16.3
	Negative info	379 (64)	-.074 (-3.11)	.124 (4.10)	-.067 (-2.05)	-.132 (-3.62)	n.s.	7.6
D1	Positive info	791 (151)	n.s.	n.s.	.-.257 (-5.77)	.571 (14.6)	.106 (3.12)	74.3
	Negative info	699 (134)	.168 (4.94)	n.s.	n.s.	.708 (18.2)	-.099 (-3.46)	59.6
D2	Positive info	1747 (281)	n.s.	-.071 (-2.02)	-0.72 (-2.38)	.573 (27.0)	-.050 (-2.54)	76.9
	Negative info	1386 (190)	-.125 (-6.50)	n.s.	.156 (8.03)	.262 (11.92)	.052 (3.72)	40.6

Table 6 The Efficiency of the Final Forecasts by Company and Information Direction (t statistics are in parentheses).

Table 7 a

Company	Information direction	Reaction to Information Percentage of forecasts which:- Under-adjust	Over-adjust	Adjust in wrong direction	% of over optimistic forecasts
G1	Positive (4013)	33.6	32.0	34.4	66.4
	Negative (3392)	46.5	25.2	28.3	45.8
	Total (7405)	39.5	28.9	31.6	57.0
G2	Positive (3049)	17.0	32.4	50.6	82.9
	Negative (2409)	46.3	33.2	20.5	54.4
	Total (5458)	30.0	32.7	37.3	66.4

Table 7b

Company	Information direction	Measure	Mean and Median change in accuracy (FFC error –SFC error: a negative value represents an improvement) Under-adjust	Over adjust	Adjust in wrong direction	Over optimism
G1	Positive (4013)	Mean	-16%	8%	44%	26%
		Median	-15%	-1%	22%	7%
	Negative (3392)	Mean	-47%	-15%	11%	-46%
		Median	-21%	-3%	11%	-21%
	Total (7405)	Mean	-30%	-1%	29%	-14%
		Median	-17%	-3%	14%	-2%
G2	Positive (2409)	Mean	-17%	29%	54%	44%
		Median	-17%	16%	43%	34%
	Negative (3049)	Mean	-36%	-1%	18%	-36%
		Median	-29%	0%	19%	-29%
	Total (5458)	Mean	-29%	15%	44%	-7%
		Median	-24%	7%	32%	15%

Table 7 Over-adjustment, Under-adjustment and Forecaster Optimism (Nos. of observations are in brackets)

Table 7a shows the percentage of forecasts which are under and over-adjustments, but in the correct direction, the percentage in the wrong direction and the percentage suffering from over-optimism. Table 7b shows the improvements deriving from the adjustment, measured as average (mean and median) final forecast error (FFC) compared to the system forecast error (SFC). The percentage of forecasts in the correct direction is lower for positive info (because the company forecasters are optimistic and wrongly expecting increases in demand). Moreover positive adjustments are more likely to be too large, worsening the system forecasts. Under-adjustment is consistently successful. The table highlights the damaging adjustments that occur (i) when they are in the wrong direction (particularly if positive), since some 35% of adjustments are in the wrong direction, (ii) for positive over-adjustments (but in the right direction) when again there are damaging large adjustments made all too frequently. Thus, for positive adjustments there is evidence of an optimism bias and support for H_{31}. Similarly, where the information is perceived as negative, the forecasts (consistent across companies except for D1), also tend to be too optimistic. Thus H_{32} is also supported

With such a large percentage of forecast adjustments turning out to have been in the wrong direction, the table also illustrates the forecasters' propensity to misunderstand the impact of perceived MI. One consequential question is whether a wrong-sided adjustment is simply a timing error in that the forecaster anticipates an effect which actually occurs a period later than expected? If this is the case there should be very few wrong sided adjustments in the following period. Our data suggests that this is not the case. For example, for the G1 companies 33.0% of adjustments made in the period following a wrong-side adjustment were also wrong sided. This pattern was particularly marked for upwards adjustments that were made following a wrong side adjustment: 43.3% of these were also wrong sided.

The effect of trend

Was there any evidence that the company forecasters made adjustments in order to damp trends particularly for downward trended series, as suggested by H_4? Evidence from observing forecasters in the companies suggested that most attention was paid to the most recent movement in sales and often the change between the two most recent periods was regarded as the trend. We therefore examined the pattern of adjustments in relation to the slope of the last segment in the sales graphs. While the forecasters did tend to apply damping slightly more often when the last segment slope was down, this only represents weak support for H_4 and could derive from optimism bias.. Interestingly, adjustments which were in the same direction as the last segment slope tended to lead to the greatest gains in accuracy. The adjustments which 'went with the flow' reduced the MdAPE by 6.1 and 12.1 percentage points following up and down trends, respectively. Adjustments, which went 'against the flow', (i.e. which represented damping) made little difference to accuracy in either case. We conjecture that this was due to the special event, for which the adjustment was being made, already impacting the slope and continuing into the next period.

5. Reaping the benefits from expert adjustment

The previous sections have shown that, despite the generally improved accuracy deriving from forecast adjustments, the final forecasts are inefficient and biased with over-optimism prevalent in response to positive market intelligence. In this section we evaluate four possible methods for improving the accuracy of the adjustments. The first two methods use a statistical procedure to correct the adjustments. The others would require either training for forecasters, better market information or restrictiveness within the forecasting software.

i) The Blattberg-Hoch approach

The '50% model, 50% manager heuristic' proposed by Blattberg and Hoch (1990) involves taking the mean of independent management judgmental and statistical forecasts. In our case, application of the heuristic will involve taking a mean of the system and final forecasts. However, because the managers saw the system forecasts (indeed they typically over-weight them) before making their adjustments the system and final forecasts are not independent. In this case the heuristic will simply act as a damper on the adjustments, restricting them to 50% of the change indicated by the managers. If there is a propensity to over-adjust forecasts (e.g. as a result of an optimism bias) then this damping might improve accuracy. However, if the forecasters are anchoring on the statistical forecast and conforming to the anchor and adjustment heuristic (Tversky and Kahneman, 1974) they will be under-adjusting from the statistical forecast and the damping is likely to be too severe.

ii) Bootstrap rules

When consistent mis-weightings of information are observed, it is possible to devise bootstrap rules which model the relationship between the judgemental forecast error and the available cue variables. Such rules have typically outperformed the unadjusted raw judgements in many studies (Dawes et al., 1989) though the evidence is primarily based on cross-sectional forecasting problems where the cues are not autocorrelated as in time series data. The time series evidence is much more limited (Lawrence et al., 2000, Lawrence and O'Connor, 1996) though Fildes (1991) showed how the mis-weighting of both a causal driver (GDP) and past actuals could be used to improve accuracy. The lack of time series evidence, despite inefficiencies being common, seems to arise from the lack of stability (over time) in the data which in turn implies that while ex post there are weights that would lead to improved accuracy, such weights cannot be estimated ex ante.

We have developed various models in order to examine whether the observed inefficiencies can be used to improve accuracy. Although lags of two periods were shown to be significant in the earlier efficiency analysis shown in table 6, the longer lags typically have a low impact on the models' standard errors, once the most recent data is forced into the model, and the additional complexity of such rules seemed likely to limit their operational impact. The model estimated here combines the various raw information sources using only one period lags - the most recent data available to the forecasters.

$$Model \;\; Y_{ij,t}$$
$$= \lambda_1 SFC_{ij,t-1}(1) + \lambda_2 MI_{ij,t} + \beta_{i,1} Y_{ij,t-1} + \gamma_1 e_{ij,t-1} + \nu_{ij,t} \qquad (M3)$$

where $SFC_{ij,t-1}(1)$ is the one-period ahead system forecast for the ith product, the jth company, made in period t-1, *MI* is the market intelligence available to the forecaster, and *e* the last period's error. Once *M3* is estimated it can be used to produce forecasts of demand.

To ensure a rigorous evaluation of the proposed models, the database was split into an estimation set of approximately 80% of the total data set for each company and a test set of the remainder. For example, for company A the last 7 months constituted the test data. This design of a hold-out sample is more demanding of the model than the alternative of selecting 80% of the data at random as hold-out. Model performance has been evaluated again using a trimmed mean (trimmed removing 1% of those extreme final forecast errors). The test of equality for positive information and negative information was highly significant and therefore the two classes of data have been modelled separately.

Various models were estimated based on the size of the adjustments made since preliminary analysis had shown that the size of adjustment affected the model coefficients (e.g., set MI=0 for small values). However, the default model where all data (apart from outliers) were used in model estimation proved most successful in terms of improved accuracy. The two models we analyse here are (1) the *full model*, incorporating the system forecasts, market intelligence and past cues, and (2) the *optimal adjust* model using just the system forecast and market intelligence . Two typical bootstrap models (outlier adjusted) are shown in Appendix 1 for company A when positive and negative information was available. Full details of all the models are available from the corresponding author. They show that the companies differ, often quite substantially, in their ability to use information in their environment to achieve optimally accurate forecasts. However, a number of traits were common to all companies, in particular for the G1 group of companies the past actuals and errors (whilst significant) were unimportant in contrast to the G2 group. Second, MI was almost optimally incorporated for negative information for G1

iii) Avoiding small adjustments

The results displayed in table 4 showed that small adjustments had a higher probability of being made in the wrong direction thus damaging forecast accuracy. We argued that this was because small adjustments tended to be the product of relatively unreliable market intelligence leading forecasters to be hesitant to commit themselves to bigger changes. This suggests that a strategy which stops forecasters from making these smaller adjustments would improve accuracy. There are a number of ways in which such a strategy could be implemented including training of forecasters and the use of software that prohibits adjustments below a pre-set percentage. Alternatively, efforts to improve the reliability of market intelligence would reduce the need for such adjustments. We examined the potential accuracy gains that could be achieved through a strategy of avoiding the smallest 50% of adjustments to see if it would be worth implementing.

iv) Avoiding wrong-sided adjustments

Our earlier analysis showed that wrong-sided adjustments are particularly damaging to forecast accuracy. We examined the potential gains that could be achieved if improvements in market intelligence could be used to eliminate *half* of these wrong sided adjustments. We did this by randomly allocating each wrong sided adjustment to either a *no change* group or a *change* group. Each member of the *change* group had its adjustment set to zero so that its forecast became the system forecast.

A comparison of the methods

We compared the accuracy of methods on both the in-sample and hold-out data using both the MAPE and MdAPE. We also ranked the results for each company (1 being most accurate) and summed the ranks. Tables 8a and 8b summarise the in-sample and hold-out sample results for the G1 and G2 companies. The in-sample results are included to demonstrate the robustness of the methods.

Companies	Information	Accuracy measures	System forecast	Final forecast	Blattberg-Hoch	Full model	Optimal MI model
G1	Positive	MAPE	30.6%	39.2%	30.3%	28.9%	29.5%
	(n=3096)	MdAPE	20.4%	15.2%	16.9%	15.9%	15.7%
		Sum of ranks	25	24	16	8	16
	Negative	MAPE	21.1%	20.0%	17.6%	18.5%	18.2%
	(n=2416)	MdAPE	32.4%	26.9%	33.0%	27.7%	27.5%
		Sum of ranks	27	17	17	14	14
G2	Positive	MAPE	36.1%	79.1%	52.4%	23.3%	31.9%
	(n=2422)	MdAPE	20.5%	44.2%	28.3%	15.7%	20.6%
		Sum of ranks	11	20	16	4	9
	Negative	MAPE	36.8%	33.7%	35.5%	23.9%	31.3%
	(n=1797)	MdAPE	23.9%	23.5%	19.9%	15.1%	18.6%
		Sum of ranks	17	16	14	4	9

n = number of observations

Table 8a In-sample accuracy of models and system and final forecasts

Companies	Information	Accuracy measures	System forecast	Final forecast	Blattberg-Hoch	Optimal adjust model	Avoid 50% smallest adjustments	Remove 50% of wrong side adjustments
G1	Positive	MAPE	32.3%	42.5%	38.1%	32.9%	45.1%	35.0%
	(n=639)	MdAPE	21.0%	21.7%	18.1%	18.9%	20.0%	18.0%
		Sum of ranks	20	30	19	12	31	14
	Negative	MAPE	52.6%	29.8%	49.3%	28.6%	36.0%	28.6%
	(n=720)	MdAPE	19.2%	18.3%	16.9%	18.4%	18.4%	17.0%
		Sum of ranks	34	16	27	15	25	8
G2	Positive	MAPE	30.3%	47.7%	38.5%	27.4%	47.3%	40.0%
	(n=627)	MdAPE	26.1%	41.8%	31.1%	21.1%	35.6%	32.9%
		Sum of ranks	8	23	14	4	21	14
	Negative	MAPE	35.2%	23.6%	27.8%	18.9%	27.0%	21.7%
	(n=612)	MdAPE	28.0%	19.4%	23.2%	11.8%	21.3%	18.0%
		Sum of ranks	18	16	14	5	20	11

n = number of observations

Table 8b Hold-out sample accuracy

It can be seen from the hold-out results that there are clear differences in the accuracy of the methods depending on whether the information is positive or negative. When information is positive the MAPEs show that, for all companies, the final forecasts tend to be less accurate than the forecasts obtained from the system so the judgmental interventions are, on average, only serving to reduce accuracy. However, comparisons with the MdAPEs indicate that much of this reduced accuracy is caused by a few extremely damaging adjustments (particularly for the G1 companies). These will be either over-large adjustments in the correct direction or wrong-sided adjustments, which as table 7 shows, are relatively more frequent for positive information.

How do the possible improvement methods fare when information is positive? Preventing forecasters from making the 50% smallest adjustments is ineffective. Because large inaccurate adjustments are a common problem when positive information is available, preventing small adjustments offers little improvement.

However, the other methods *are* effective when information is positive. Rather than removing the small adjustments the Blattberg-Hoch method, as applied here, acts by damping *all* adjustments by 50%. This method therefore improves accuracy by reduces the damaging impact of the large adjustments. Wrong-sided adjustments are also frequently associated with positive information (especially for the G2 companies –see Table 7). Not surprisingly, developing measures to prevent half of these adjustments also leads to substantial improvements over the final forecasts. However, the final improvement method, the optimal adjust model, performs best, reflecting the serious mis-weighting of market intelligence that occurs when this information is positive. For the G2 companies the improvements it yields over the final forecasts are substantial.

When information is negative a contrasting set of results is obtained. Table 8 shows that, the judgmental adjustments tend, on average, to improve accuracy –the direct opposite of the situation with positive information. In this case, comparison between the MAPEs and MdAPEs suggests that the judgmental adjustments are particularly effective because they are reducing many of the extreme errors in the system forecasts. Table 7 shows that reactions to negative information frequently take the form of under-adjustments in the correct direction. When the system forecast errors are extreme, these under-adjustments improve accuracy, though they do not, of course, completely eliminate the errors.

How effective are the improvement methods when information is negative? Table 8 shows that two of the methods are at best ineffective. Indeed, if measures are taken to remove the 50% smallest adjustments, accuracy will actually be reduced (especially for the G2 companies). This is because the measure would eliminate the small under-adjustments which are nevertheless in the correct direction.

Clearly, this can have no value unless the effort that these adjustments entail is not warranted by the improvements in accuracy that they yield. Similarly, by damping the size of the adjustments, the Blattberg-Hoch method only exacerbates the tendency to under-adjust.[3]

Preventing half of the wrong-sided adjustments is bound, by definition, to lead to improved accuracy. However, when information is negative there tend to be fewer wrong-sided adjustments (see table 7) so the benefits derived from measures designed to eliminate half of them are less than those obtained with positive information. The optimal adjust model does not yield improvements for the G1 companies, reflecting the fact that forecasters in these companies attached virtually optimum weights to market intelligence when this information was negative, but it does lead to improved accuracy for the G2 companies where market intelligence was mis-weighted.

In summary, preventing small adjustments is not likely to lead to improvements in accuracy, whatever the information direction. The other methods are effective when information is positive. When it is negative they lead to smaller improvements in accuracy or none at all. As a strategy, the optimal adjustment procedure works well overall; it generally leads to improved accuracy, sometimes substantial, and with no damaging effects for either positive or negative information or for both groups of companies.

6. Discussion

Our analysis has revealed major differences in the forecasting accuracy obtained by the companies. It also showed the potential for improvement that could be achieved by focusing on the more effective use of market intelligence and the removal of consistent biases. However, the four improvement methods we examined would all face possible obstacles if an attempt was made to implement them in many organisations. Because the Blattberg-Hoch and bootstrap approaches are automatic correction procedures their use may lead to the demotivation of forecasters, with less effort being applied to the original judgments (Goodwin, 1996) . Alternatively, the nature of the biases may change over time, possibly as a

[3] For the G1 companies the Blattberg-Hoch methods yields a slightly lower MdAPE (16.9%) than the final forecast (18.3%), but the result derives from the better performance one just one of the three companies as the rank comparison demonstrates.

result of training, or the company forecasters may seek to pre-empt the corrections by distorting their judgmental inputs into the process. Improvements in forecasting through training and better use of market intelligence, which are required by the other two improvement methods we examined, are also not straightforward to achieve and there are many barriers to adopting new forecasting procedures (see for example Schultz (1984) In this section we use our experience of extensive meetings with the companies concerned and observation of their forecasting processes to provide explanations for the results and to identify how improvements might be achieved.

The Forecasters

Fildes and Hastings (1994) and Moon et al. (2003) identified motivation and training as potentially important in attaining accurate forecasts. With the wrong incentives forecasters may be motivated to produce forecast that are biased. For example, in their survey of US sales forecasting practice Sanders and Manrodt (1994) found a preference for under forecasting. Similar results were found in a recent survey by Fildes and Goodwin (2006) . In our research, despite all company forecasters explicitly stating that accuracy was an important goal (for them and the company) company A preferred to over-forecast while the retailer was found, in many periods, to be confusing the demand forecasts with decisions on inventory levels. Thus the forecasts were not a genuine expectation of the next period's demand, but a decision on how much of the product to stock. This probably accounts for much of the bias associated with these 'forecasts'.

In none of the companies were the forecasters expert in the statistical aspects of forecasting such as error measurement or alternative forecasting methods. Nor were they aware of the many biases associated with judgmental interventions in forecasting. All the senior forecasters were, however, immersed in process and management issues relating to forecasting with one acting as a regular presenter at professional forecasting events. It therefore appears that training and the use of appropriately targeted incentives would be likely to lead to improvements in accuracy.

The Forecasting Support System

All four company systems were professionally developed, but they had inflexible interfaces and poor (or in one case non-existent) graphics. Yet the format of the interface can be important in improving accuracy (Tashman and Hoover, 2001) . The system forecasts were produced using models far removed from best practice (see, for example,Gardner and Andersen (1997)), nor had the chosen methods been tuned to produce the best possible results from the software. Standard exponential smoothing models are now known to require just such tuning in the choice of smoothing parameters so this might explain the inadequacies (Gardner, 2006) . This was underlined in the strong performance of the naïve forecast compared to the system forecast for companies B and D. Perceived inadequacies in the system forecasts lead to a complex adjustment process (see Goodwin et al. (2006) for a case study of company A) and, typically, worse forecasts (Goodwin et al. 2007) . Analysis of screen displays show that the forecasters do not have clear guidance on the previous actuals and previous errors. Summary error measures are not easily available and those that are provided are subject to outlier and intermittent demand effects. Thus a reliable assessment of the gains or losses in accuracy resulting from judgmental adjustments could not be made and there was therefore little opportunity to learn from experience about the appropriateness of judgmental intervention in different circumstances. Improved statistical forecasting systems to provide better base line forecasts and accuracy monitoring systems using well-designed error measures are therefore needed.

Although three of the systems had a ‘notes’ facility, whereby the forecaster could explain the reasons for the adjustments they make, they had none of the features that might make it easy to use and effective. Use was spasmodic and incoherent in that forecasters could not explain the past adjustments they had made to us by referring to their ‘notes’ system. Clearly, requiring forecasters to record reasons for their adjustments in a standard format (e.g. by selecting a reason from a list) might serve to reduce the number of relatively small, but damaging adjustments that may be based on unreliable market intelligence or other untrustworthy information (Goodwin, 2000) . This would also allow forecasters to understand why and how market intelligence is so often misinterpreted. In addition, it would permit the decomposition of market intelligence into key drivers, thereby lessening the likelihood of double counting.

While enhancements in forecasting support systems in recent years have reflected advances in statistical forecasting methods, the development of facilities to support managers' judgmental inputs has been almost non-existent (Fildes et al., 2006) . Experimental evidence suggests that the incorporation of guidance systems, such as those which allow the formal use of analogies (e.g., past promotions and their effects (Lee et al., 2006) would improve the quality of judgments based on market intelligence. In addition, the evidence from the retailer (company D) suggests the need for a de-coupled system which clearly distinguishes between forecasting and the associated supply chain decisions. A system which allows the two to be compounded is difficult to improve since there is not clear view of forecast error or optimal inventories. Such a system might help to mitigate the pressures towards bias, both personal and organisational, that exist in many companies.

Market Intelligence

The company forecasters identified promotions as the most important driver of their judgemental adjustment (though not for company A, the pharmaceutical company). Other important aspects include price changes (Companies C & D), weather (B, D), inventories (Company A, D). Where market intelligence is strong and its direction clear, there are major improvements in accuracy as seen in Table 4. However, the process by which such intelligence is gained is, as Moon et al. (2003) point out in an examination of 16 case studies, often very flawed, primarily through the lack of coordination and communication between the different organisational units involved in supply chain operations, sales and marketing. Here, while all companies apparently consult widely on important drivers, the evidence that is collected is not compiled in a manner, which through, for example, a notes system, leads to learning by analogy from earlier exemplars. Each event is instead treated as unique.

In addition, different sources of intelligence are not identified and quantified separately using a recommended decomposition approach. This can lead to double counting or omission (MacGregor, 2001) . None of the companies attempted to review the reasons for wrongly interpreting the direction of an adjustment, a cause of major forecast error.

7. Conclusions

Judgemental adjustments to statistical forecasts is very common with up to 80% of forecasts adjusted in. some companies. In the context of forecasting SKU data to support supply chain operations and planning, the accuracy of such forecasts and the judgemental adjustments is important, impacting on profit and service levels. However, the effectiveness of these judgements is moot with very limited empirical evidence available. In this paper we have shown conclusively that the value of these adjustments depends on the company context, but where the forecasters' principal motivation is towards improved accuracy, there is substantial added value in the adjustments made.

However, the company forecasts we observed proved to be biased and inefficient. Optimism bias was prevalent. The forecasters tended to over-weight the statistical systems forecast, which for some of the companies was itself flawed in comparison to a naïve forecast. We therefore developed various models for capitalising on the biases and inefficiencies, showing that the most appropriate model depends on the circumstances, in particular the nature of perceived intelligence, positive or negative. For example, the effectiveness of the Blattberg-Hoch heuristic of 50% model+50% man, as we applied it, is itself limited to positive adjustments. Based on over 12000 judgementally adjusted forecasts and 1536 SKUs we can therefore conclude that, at least for the companies analysed here, they could improve their forecasts substantially by more effective incorporation of market intelligence. In particular, they should put into place processes to avoid the 'optimism' bias, either as part of the FSS or in the motivation and monitoring aspects of the organisational process. Avoiding small adjustments, perhaps through a constraining FSS, would free time to focus on what is undoubtedly the most important issue, identifying the direction of market intelligence.

Forecast adjustment is the only practical way for most organisations to improve their incorporation of key drivers into their disaggregated sales forecasts. While the evidence we present shows the benefits of adjustment, there remains plenty of opportunity for both companies and researchers to understand how such factors are best included. The result should be major improvements in accuracy.

Appendix 1

Table A1 Incorporating Market Intelligence: Predictive Models of Sales and their forecasting Accuracy (by company group).

			System forecast	Past Actual	Past error	MI adjust	R sq.	Insample MAPE (MdAPE)	Holdout MAPE (MdAPE)
G1	Positive info 1367; estimation set holdout 639	M3: all info	.842 (62.5)	.169 (13.5)	n.s.	.424 (22.6)	.97	.368 (.155)	.391 (.176)
		Optimal Adjust	.987 (211)	*	*	.450 (25.2)	.97	.373 (.165)	.383 (.175)
		50-50	1	*	*	.5	*	.384 (.160)	.393 (.174)
	Negative info 1087 (446)	M3: all info	.827 (59.30)	.181 (13.2)	n.s.	.804 32.0)	.97	.420 (.152)	.337 (.159)
		Optimal Adjust	.995 (197)	*	*	.989 (43.1)	.96	.415 (.152)	.321 (.156)
		50-50	1	*	*	.5	*	.564 (.173)	.448 (.162)
G2	Positive info 2063 Holdout 627	M3: all info	.392 (17.2)	.590 (24.1)	n.s.	.203 (8.90)	.97	.278 (.164)	.357 (.195)
		Optimal Adjust	.864 (122)	*	*	.272 (17.9)	.94	.381 (.121)1	.303 (.203)
		50-50	1	*	*	.5	*	.565 (.294)	.373 (.256)
	Negative info 1797	M3: all info	.465 (17.4)	.573 (18.7)	.089 (4.14)	.371 (13.8)	.97	.256 (.147)	.216 (.118)
	Holdout 612	Optimal Adjust	.913 (104)	*	*	.622 (21.4)	.94	.320 (.195)	.244 (.204)
		50-50	1	*	*	.5	*	.371 (.202)	.292 (.206)

The table shows the estimated cue models based on M3 which use the system forecast, the past cues and market intelligence to estimate actual demand. If the final forecast were the best achievable the coefficients of SFC (the system forecast) and MI (market intelligence) would both be 1, with zero weight assigned to past actuals and past errors. The data have been grouped into the manufacturing companies (forecasting monthly) and the retailer (forecasting weekly). For the monthly data the drivers of the past actual and past error have limited or no predictive power. However the forecasters over-respond to market intelligence, particularly in the case where information is positive, and therefore damping the adjustment improves accuracy.

References

Armstrong, J. S., F.Collopy.1992. Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting* **8** 69-80.

Armstrong, J. S., R.Fildes.1995. On the selection of error measures for comparisons among forecasting methods. *Journal of Forecasting* **14** 67-71.

Blattberg, R. C., S.J.Hoch.1990. Database models and managerial intuition: 50% model + 50% manager. *Management Science* **36** 887-899.

Buehler, R., D.Griffin.2003. Planning, personality, and prediction: The role of future focus in optimistic time predictions. *Organizational Behavior and Human Decision Processes* **92** 80-90.

Clements, M. P., D.F.Hendry.1995. On the selection of error measures for comparisons among forecasting methods -reply. *Journal of Forecasting* **14** 73-75.

Dawes, R. M., D.Faust, P.E.Meehl.1989. Clinical versus actuarial judgment. *Science* **243** 1668-1673.

Deschamps, E.2004. The impact of institutional change on forecast accuracy: A case study of budget forecasting in Washington State. *International Journal of Forecasting* **20** 647-657.

Donihue, M. R.1993. Evaluating the role judgment plays in forecast accuracy. *Journal of Forecasting* **12** 81-92.

Eggleton, I. R. C.1982. Intuitive time-series extrapolation. *Journal of Accounting Research* **20** 68-102.

Fildes, R.1991. Efficient use of information in the formation of subjective industry forecasts. *Journal of Forecasting* **10** 597-617.

Fildes, R., C.Beard.1992. Forecasting systems for production and inventory control. *International Journal of Operations and Production Management* **12** 4-27.

Fildes, R., S.Bretschneider, F.Collopy, M.Lawrence, D.Stewart, H.Winklhofer, J.T.Mentzer, M.A.Moon.2003. Researching sales forecasting practice - Commentaries and authors' response on "conducting a sales forecasting audit" by M.A. Moon, J.T. Mentzer & C.D. Smith. *International Journal of Forecasting* **19** 27-42.

Fildes, R., P.Goodwin.2006. Fine judgement: do organizations follow best practice when applying management judgement to forecasting? *Lancaster University Working Paper*

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

Fildes, R., R.Hastings.1994. The organisation and improvement of market forecasting. *Journal of the Operational Research Society* **45** 1-16.

Flyvbjerg,B., N.Bruzelius, W.Rothengatter. *Megaprojects and risk: An anatomy of ambition.* Cambridge University Press, Cambridge.

Gardner, E. S., E.A.Andersen .1997. Focus forecasting reconsidered. *International Journal of Forecasting* **13** 501-508.

Gardner, J. E. S.2006. Exponential smoothing: The state of the art -Part II. *International Journal of Forecasting* **22** 637-666.

Goodwin, P.1996. Statistical correction of judgmental point forecasts and decisions. *Omega: International Journal of Management Science* **24** 551-559.

Goodwin, P.2000. Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting* **16** 85-99.

Goodwin, P., R.Fildes.1999. Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making.* **12** 37-53.

Goodwin, P.,R.Fildes,M.Lawrence,K.Nikolopoulos.2007. The process of using a forecast support system

Goodwin, P.,W.Y.Lee,R.Fildes,K.Nikolopoulos,M.Lawrence.2006. Understanding the Use of Forecasting Systems: An Interpretive study in a Supply-Chain Company

Helbok, G., M.Walker .2004. On the nature and rationality of analysts' forecasts under earnings conservatism 36 45-77.

Klassen, R. D., B.E.Flores.2001. Forecasting practices of Canadian firms: Survey results and comparisons. *International Journal of Production Economics* **70** 163-174.

Lawrence, M., P.Goodwin, M.O'Connor, D.Onkal.2006. Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting* **22** 493-518.

Lawrence, M., M.O'Connor, R.Edmundson.2000. A field study of sales forecasting accuracy and processes. *European Journal of Operations Research* **122** 151-160.

Lawrence, M., M.J.O'Connor.1996. Judgement or models: The importance of task differences. *Omega: International Journal of Management Science* **24** 245-254.

Lawrence, M. J., S.Makridakis.1989. Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes* **42** 172-187.

Lee, W. Y.,P.Goodwin,R.Fildes,K.Nikolopoulos,M.Lawrence.2006. Providing support for the use of analogies in demand forecasting tasks

Lim, J. S., M.O'Connor.1995. Judgemental Adjustment of Initial Forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making* **8** 149-168.

Lim, J. S., M.O'Connor.1996. Judgmental forecasting with interactive forecasting support systems. *Decision Support Systems* **16** 339-357.

MacGregor,D.G.2001. Decomposition for judgmental forecasting and estimation. In J.S.Armstrong (Ed.) *Principles of Forecasting.* Kluwer Academic Publishers, Boston.

Mathews, B. P., A.Diamantopoulos.1986. Managerial intervention in forecasting. An empirical investigation of forecast manipulation. *International Journal of Research in Marketing* **3** 3-10.

Mathews, B. P., A.Diamantopoulos.1989. Judgemental revision of sales forecasts: A longitudinal extension. *Journal of Forecasting* **8** 129-140.

Mathews, B. P., A.Diamantopoulos.1990. Judgemental revision of sales forecasts: Effectiveness of forecast selection. *Journal of Forecasting* **9** 407-415.

Mathews, B. P., A.Diamantopoulos.1992. Judgmental revision of sales forecasts -the relative performance of judgementally revised versus nonrevsied forecasts. *Journal of Forecasting* **11** 569-576.

McNees, S. K.1990. The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting* **6** 287-299.

Moon, M. A., J.T.Mentzer, C.D.Smith.2003. Conducting a sales forecasting audit. *International Journal of Forecasting* **19** 5-25.

O'Connor, M., W.Remus, K.Griggs.1993. Judgemental forecasting in times of change. *International Journal of Forecasting* **9** 163-172.

O'Connor, M., W.Remus, K.Griggs.1997. Going up-going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting* **16** 165-176.

O'Connor, M., W.Remus, K.Griggs.2001. On the asymmetry of judgmental confidence intervals. *International Journal of Forecasting* **17** 623-633.

Sanders, N. R., K.B.Manrodt.1994. Forecasting practices in United-States corporations -survey results. *Interfaces* **24** 92-100.

Schultz, R. L.1984. The implementation of forecasting models. *Journal of Forecasting* **3** 43-55.

Tashman,L.J. , J.Hoover.2001. Diffusion of forecasting principles through software. In J.S.Armstrong (Ed.) *Principles of Forecasting: A handbook for researchers and practitioners.* Kluwer Academic Publishers, Boston.

Theil,H. *Applied Economic Forecasting.* North-Holland Publishing Company, Amsterdam.

Turner, D. S.1990. The role of judgment in macroeconomic forecasting. *Journal of Forecasting* **9** 315-345.

Tversky, A., D.Kahneman.1974. Judgment under uncertainty: Heuristics and biases. *Science* **185** 1124-1131.

Worthen, D.2003. Future results not guaranteed: contrary to what vendors tell you, computer systems alone are incapable of producing accurate forecasts. *CIO* **16** 1-1.

Yelland, P. M.2006. Stable seasonal pattern models for forecast revision: A comparative study. *International Journal of Forecasting* **22** 799-818.