Information use in supply chain forecasting

Fildes, Robert and Goodwin, Paul and Onkal, Dilek

Lancaser University Dept. of Management Science, University of Bath Management School, Bilkent University, Turkey

May 2015

Lancaster University Management School

Department of Management Science Working Paper 2015:2

Information use in supply chain forecasting

Robert Fildes[a,1] , Paul Goodwin[b], Dilek Onkal[c],

a Lancaster Centre for Forecasting, Lancaster University, UK, LA1 4YX

b Management School, University of Bath, UK

c Faculty of Business Administration, Bilkent University, Turkey

[1] Corresponding author. Email addresses: r.fildes@lancaster.ac.uk, P.Goodwin@bath.ac.uk, onkal@bilkent.edu.tr.

Information use in supply chain forecasting

Robert Fildes, Lancaster University Centre for Forecasting, UK
Paul Goodwin, Management School, University of Bath, UK
Dilek Onkal, Bilkent University, Turkey

Abstract:

Demand forecasting to support supply chain planning is a critical activity, recognized as pivotal in manufacturing and retailing operations where information is shared across functional areas to produce final detailed forecasts. The approach generally encountered is that a baseline statistical forecast is examined in the light of shared information from sales, marketing and logistics and the statistical forecast may then be modified to take these various pieces of information into account. This experimental study explores forecasters' use of available information when they are faced with the task of adjusting a baseline forecast for a number of retail stock keeping units to take into account a forthcoming promotion. Forecasting demand in advance of promotions carries a particular significance given their intensive supply chain repercussions and financial impact. Both statistical and qualitative information was provided through a forecasting support system typical of those found in practice. Our results show participants responding to the quantity of information made available, though with decreasing scale effects. In addition, various statistical cues (which are themselves extraneous) were illustrated to be particularly important, including the size and timing of the last observed promotion. Overall, participants appeared to use a compensatory strategy when combining information that had either positive or negative implications for the success of the promotions. However, there was a consistent bias towards underestimating the effect of the promotions. These observed biases have important implications for the design of organizational sales and operations planning processes and the forecasting support systems that such processes rely on.

Key Words; Demand planning; Sales and Operations Planning; Behavioural operations; Forecasting support systems; Promotional planning; Information effects.

May 2015

Information use in supply chain forecasting

1. Introduction

Forecasts lie at the heart of supply chain operations: production and inventory planning and scheduling, logistics, marketing and finance all rely on short-term disaggregate forecasts at SKU level. Yet little research has been carried out into the way such forecasts are actually produced and the factors that influence their effectiveness (Seifert et al, 2015; Thomé et al., 2012; Tuomikangas and Kaipia, 2014). In contrast to the academic research literature, the practitioner literature is awash with descriptions and recommendations as to how 'Sales and Operations Planning (S&OP)' processes can be used to effectively integrate cross-functional information to produce forecasts, for example, Lapide (2007) and Stahl (2010). To address this research gap in part, the aim of this paper is to identify how forecasters respond to information that reflects the various countervailing events and trends that are expected to influence sales. The approach adopted involves a number of experiments in a realistic simulation of the supply chain forecaster's task environment.

The forecasts within S&OP are usually produced as a combination of a simple statistical forecast and judgment (Fildes and Goodwin 2007). The most common approach is where an initial statistical forecast is adjusted judgmentally. After the statistical forecast is produced (based on the SKU time series history for established products), the forecaster may receive information from other functional areas such as marketing and may adjust the statistical forecast to reflect this additional information. These judgmentally adjusted forecasts are then adopted as final forecasts across the supply chain. Moon et al. (2003) argue that the functional integration that is involved in the production of these forecasts is an important element of effective forecasting. However, in one of the few detailed case studies of forecasting practice, Goodwin et al. (2007) found that the overall benefits of the additional information were slight and often negative, when they observed a highly formalised procedure in a pharmaceutical company for adjusting the initial statistical forecast based on regularly scheduled meetings. Oliva and Watson (2011), in a case study of a consumer electronics manufacturer, also focused on processes for information sharing across functional areas and the harmful effects of what Moon (Moon, et al., 2003) have called functional silos, where there are disincentives for the different functions to be aligned. The careful design of the S&OP process, Oliva and Watson (2011) argue, can lead to effective forecasting. But this claim is based on limited evidence as to both the statistical features of the forecasting process they describe as well as how benefits accrue from better integration of information. A second route to developing insights into the benefits of such additional information can be gained through cross-sectional survey evidence, comparisons of statistical forecasts with the final forecasts after judgmental adjustment and the impact of information use. However, deriving conclusions from these surveys is not straightforward given the tricky interactions between use of different information sources and the forecasting performance metrics that are dependent on organizational processes (Danese and Kalchschmidt 2011a, 2011b).

While the detailed case studies provide insight into the processes by which information is combined, they have told us little about the statistical characteristics of supply chain forecasts and whether there is any potential for significant improvement. This aspect of the use of information has recently been examined by Fildes et al. (2009) and Franses and Legerstree (2010, 2011, 2013) who between them have analysed over 100,000 forecasts from a number of companies. By contrasting the revealed sales, the initial statistical forecast and the final adjusted forecast, various hypotheses have been explored including whether the forecasts are biased and what determines forecast error. Overall, the results suggest that the information that is brought to bear in the production of the final forecast is not used rationally or efficiently (Fildes et al. 2009), that the experts place excessive weight on their own judgment (Franses and Legerstee 2010, 2013), and that the resulting adjusted forecasts are biased (Fildes et al, 2009). Also, negative information leading to downward adjustments of the statistical forecast is weighted differently from positive information, and this leads to downward adjustments having greater value (Fildes et al. 2009). Furthermore, lower accuracy of expert-adjusted forecasts (as compared to model-based forecasts) appears to persist across various forecast horizons (Franses and Legerstee 2011).

But what is the information that is being used when the judgmental adjustments are being made? In a survey of primarily supply chain forecasters (Fildes and Goodwin 2007), the key pieces of information concerned the effects of marketing activities in particular promotions, together with the effects of advertising and price changes. Supply limitations were seen as important whilst holiday and weather effects were also listed by around 20% of respondents. Changes in government policy and regulation also influenced the adjustments forecasters made. Trapero et al. (2013) have gone some way to understanding how promotional information is used. Using data on promotions as well as forecasts and sales from a household product manufacturer, they showed that important promotional information leading to large-scale adjustments had extremely deleterious effects on forecast accuracy. In essence, the forecasters misinterpreted the positive information in their possession.

From the case studies of the S&OP process and the statistical modelling we have just described, various aspects of the task of forecast production have proved important; in particular, the integration of different types of information including both the time series history and statistical forecasts supplemented by descriptive information, along with the context in which the information is used including the forecaster's characteristics and motivation. The consequences of how the information is processed are observed biases and inefficiencies in forecasting accuracy. In this paper we take an experimental approach, effectively completing the methodological triangulation. The aim is to understand in a controlled setting the differential effects of the time series cues, and the descriptive information, both 'hard' and 'soft', that an S&OP process typically delivers. In particular, we will focus on how forecasters handle information of different types from a range of different sources.

The paper is in five further sections. The next section discusses what the literature can tell us about how forecasters use and combine time series and contextual information to produce judgmental estimates and any effects of their different motivations and understanding when engaging with this process. This leads us to a set of hypotheses. Section 3 considers our methodology, including the contribution experimental research makes to the study of supply chain forecasting, the participants and the experimental setting. The fourth section contains the results and Section 5 provides a discussion of our findings. Finally, Section 6 summarizes our conclusions as well as providing suggestions for further work.

2. The use of information in supply chain forecasting

Accurate forecasts are critical for effective performance of supply chain systems. Although previous work has acknowledged the importance of an expanded information set (Fildes and Hastings 1994), information sharing (Lee, So and Tang 2014, Önkal and Aktas, 2011, Yu, Yan and Cheng, 2001) and collaborative forecasting across organisations (Aviv 2001, Eksoz et al. 2014) in order to enhance supply chain performance, there appears to be surprisingly little empirical work on the actual use of information in supply chain forecasting. In attempting to fill this research gap, the current study examines four factors that may constitute barriers to the effective use of information in supply chain forecasting: (i) biases associated with the use of time series information, (ii) the cognitive challenges of making an assessment of the aggregate effects of *multiple items* of contextual information that may be presented in a qualitative or narrative form, (iii) the challenge of combining time series information, a statistical forecast and contextual information, and (iv) the motivations and prior understanding of the forecaster.

2.1 Heuristics and biases associated with the use of time series information

Forecasts can be systematically biased even when forecasters have the relatively simple task of basing their estimates only on time series information (Lawrence et al. 2006). When time series are subject to noise, forecasters have a tendency to implicitly perceive autocorrelation in series which have independent observations and to add noise to their forecasts so that they appear to be representative of past observations (Reimers & Harvey 2011). These effects may be intensified the higher the level of noise. Related to this is a tendency to pay too much attention to the latest observation in a series. As a result, for untrended series, or series with high serial dependence, the forecast tends to be a weighted average of the latest observation and the long term mean of the series (Harvey et al. 1994). When a trend is present, and serial correlation (after removal of the trend) is low, people tend to forecast by adding a proportion of the last difference in the series to the last observation (Bolger and Harvey 1993).

When the effects of past special events are contained in the time series and a forecast needs to be made for a period when such an event is expected, Goodwin and Fildes (1999) found evidence that forecasters used a pattern matching strategy (Hoch and Schkade 1996). This involved searching for the past special event that was most similar to the forthcoming event and using the actual sales for this past event as the basis for the forecast (as a match with actual sales rather than the uplift from a baseline forecast). This is analogous to the standard approach used in industry for forecasting sales in promotion periods, though here the most recent promotion is the one that is usually judged to be most similar to the forthcoming promotion (Cooper et al. 1999). Given the prevalence of a recency bias in judgmental forecasting it seems likely that the proximity of a past event (such as the latest promotion) to the current period will increase the salience of the effects of that event and hence its influence on the forecast. A recency bias is often associated with a reduced ability to recall older information from memory, but the over-attention that is paid to the latest observation when a time series is presented suggests that it may also apply when all available information is simultaneously presented.

The Goodwin and Fildes (1999) study found that pattern matching was not evident under conditions of high noise where the effects of past special events were submerged in the random movements of series. Nevertheless, even if pattern matching is not used the greater volatility of high noise series may suggest that a wider range of outcomes is possible so that the extreme values associated with special events may be seen as more plausible.

2.2 The effect of providing a statistical time series forecast

In companies it is common for statistical time-series methods, such as exponential smoothing, to be used to provide baseline forecasts of sales (i.e. forecast of sales before the effects of special event are considered). These are typically provided through a Forecasting Support System (Fildes et al. 2006). But how are statistical forecasts used when it is known that a special event will produce effects that the statistical methods have not taken into account? In the study by Goodwin and Fildes (1999) when sales were subject to promotion effects the forecasters appeared to ignore the statistical forecast completely, possibly because they perceived that the regular underlying time series pattern was suspended in promotion periods. However, much may depend on the way the judgmental inputs are elicited. In the Goodwin and Fildes study, participants were simply told of the statistical baseline forecast and asked to enter their own forecast. An invitation to *adjust* the statistical forecast seems more likely to focus attention on it. Thus the interplay between the influence of the statistical forecast, the latest observation and the most recent promotion is likely to be crucial. There appear to be two main possibilities:

1. The statistical forecast will be judgmentally adjusted to match the sales achieved in the previous promotion and the last observation will be ignored. If this is the case larger statistical forecasts will be associated with smaller adjustments, as they will be closer to the sales achieved in the previous promotion, and the last observation will not be associated with the size of the adjustment.

2. The judgmental adjustment will be performed as in (1), but an additional upwards adjustment will be made to reflect the last observation as it might be seen as reflecting a recent change in the baseline level of sales (e.g. a recent increase in the popularity of a product). This suggests that the forecaster would attempt to match the previous promotion and then add or subtract from this depending on whether the previous observation is relatively high or low. If this is the case, larger statistical forecasts will be associated with smaller adjustments, as before, but the higher the last observation the higher will be the upwards adjustment.

Given the attention that forecasters tend to pay to the latest observation, (2) would appear to be more likely with both the latest observation and the previous promotion's sales potentially used as double anchors.

In summary, it appears that where a forecaster is asked to adjust a statistical baseline forecast of sales to take into account an imminent sales promotion, the resulting final forecast is likely to fall between the statistical forecast and the sales achieved in the previous promotion. This means that the closer the baseline forecast is to the previous promotion's sales, the lower will be the adjustment that is perceived to be necessary. However, three other factors seem likely to lead to larger upwards adjustments. First, the influence of the previous promotion is likely to be greater if it is more recent as it may then be seen to be more relevant. Second, the higher the last observation in the time series, the greater will be the adjustment. Third, a high or low observation in the most recent period is likely to be more salient when it lies well above or below the plot of the statistical forecast for that period on a graph. Thus, a large positive forecast error in the last period is likely to increase the upwards adjustment made for the latest observation and the opposite may be true of a large negative error. Finally, higher levels of noise in the series may have conflicting effects on the size of adjustment. High noise will tend to mask the effects of the earlier promotion but the greater variation of the past time series values will provide support for the belief that high sales are possible. These arguments lead to the following hypotheses.

H1: Larger sales achieved in the previous promotion will be associated with larger upwards adjustments.

H2: The closer the timing of the previous promotion to the current period the larger the adjustment will be.

H3: The higher the last observation the higher will be the adjustment.

H4: Statistical forecasting errors for the most recent period will be positively associated with judgmental adjustments.

H5: Larger statistical baseline forecasts will be associated with smaller adjustments.

H6: Levels of noise in a series will not be associated with the size of upwards adjustments.

2.3 Assessing the implications of multiple items of contextual information

The ways in which people make assessments of the effects of multiple items of contextual information, or cues, has been the subject of much research by psychologists (e.g. see Karelaia & Hogarth 2008). However, the cues have usually been presented as numeric values, rather than the verbal arguments or narratives which are commonly presented during the S&OP process to supply chain forecasters. How is a forecaster likely to react when faced with a set of multiple statements, some of which will be positive in relation to the expected promotional effect and some of which will be negative?

To produce an adjustment to a statistical forecast, ideally, s/he would assess the reliability and relevance of the information in each statement, filtering out any information that is judged not to be sufficiently reliable or relevant. The forecaster should then assess each statement's likely implications for sales before finally aggregating these estimated effects, together with any estimated interactive effects. However, assessing reliability and the implications of the different statements may be hindered by a sparsity of past cases. For example, promotion campaigns may be relatively rare so a forecaster will have few opportunities to assess the extent to which particular factors will be associated with different levels of sales uplift. In addition, before considering the specific features of an event such as a promotion, forecasters may have a prior view of its likely impact that may be based on their recall (which may be imperfect) of earlier promotions (Reimers and Harvey 2011), or on industry beliefs. Thus there is likely to be a tension between the prior estimate, and the assessment of the implications of the specific pieces of available information relating to a forthcoming promotion. Anchoring on the effects of an earlier promotion or a prior view would suggest that the latest information will be underweighted, but Kahneman and

Lovallo's (1993) notion of the 'inside view' would also suggest that attention will be devoted to the specifics of the particular circumstance being studied.

When it comes to aggregating the effects implied by the different statements, research in other domains, such as multiattribute decision making, suggests that forecasters may employ either compensatory or non-compensatory strategies (Payne et al. 1993). In a compensatory strategy the effects of the full set of available statements are considered and the implications of reasons pointing to low sales counted against the implications of those favouring high sales. In principle, the aggregation may involve a weighted sum strategy where the estimated effects are weighted depending on the reliability, and relative importance of the underlying arguments. There is some evidence that, in general, negative information is perceived to be more potent and more salient than positive information. When combinations of positive and negative information are provided, the holistic perception is more negative than the aggregate of the evaluations of the individual pieces of information (Kahneman & Tversky 1979; Rozin & Royzman 2001). However, attaching differential weights to items of information is cognitively demanding and it seems possible that forecasters will an equal weight strategy (e.g. Dawes 1979). Indeed, when faced with multiple items of information it seems most likely that supply-chain forecasters will seek to minimize the cognitive demands placed on them by adopting a non-compensatory strategy. At its simplest, this may involve basing their judgment on a single statement – for example, the one which is the most salient (such as a negative statement) or is judged to be the most important or reliable – while ignoring the others. This leads to the following hypotheses:

H7: Given the cognitive demands placed on them, forecasters will adopt a non-compensatory strategy when considering the estimated effects of multiple items of contextual information.

H8: Negative information will have a greater influence on forecast adjustments than positive information.

2.4 Combining time series information, a statistical forecast and contextual information

The human mind has limited information processing capacity (Hogarth, 1987) and when large amounts of information from different sources needs to be integrated using judgment, it seems unlikely that the information will be combined efficiently (Leitner & Leopold-Wildburger, 2011). In particular, problems may occur when there is a mixture of statistical base-rate data and contextual information. Information presented verbally in the form of anecdotes, stories or scenarios may attract more attention than the relatively dull numeric statistical forecast or past time series data (Önkal, Sayım, and Gönül 2013). For example, in a classic study Tversky and Kahneman (1974) showed that information on statistical base rates is often neglected or discounted even when unreliable narrative information is supplied to people making judgments. Hence we hypothesize:

H9: Statistical base-rate information on the average uplift achieved by promotions will tend to be neglected.

2.5 The effects of prior understanding and motivation on information use

The motivation of the forecaster is also likely to affect the way in which sets of information in verbal statements will be assessed and aggregated in forecasting (Eroglu and Croxton 2010). In some situations forecasters may prefer the variable-to-be-forecast to take on high or low values (e.g. a desire for high sales). Such desirability of outcomes may lead to an overblown optimism (referred to as 'desirability bias'), potentially influencing the relevant predictions (McGregor 1938, Olsen 1997; Windschitl et al. 2010).

Valence priming and differential scrutiny are two mechanisms that may be associated with inducing optimistic forecasts Krizan and Windschitl (2007). In *valence priming,* an event that is intrinsically attractive, like high sales, activates attention to positive factors that are consistent with the desirable event. For example, attention may be directed to factors such as an enthusiastic reaction to a product's packaging by focus groups, rather than the news that a rival is launching a competing product. *Differential scrutiny* applies when evidence in favour of the desired outcome is accepted at a lower quality threshold. Evidence against the desired outcome is more carefully scrutinised so that weaknesses in this evidence are more likely to be uncovered. In contrast to these mechanisms, *negativity bias* can make people more pessimistic because the desire for a positive outcome actually makes negative information more salient.

Despite these potential biases forecasters in many organisations are also likely to be motivated to produce accurate forecasts. Indeed, supply chain forecasters identified accuracy as their most important objective in the survey by Fildes and Goodwin (2007); so aiming for accuracy may lead to a more thorough assessment of available information, increasing the chances that a compensatory strategy will be employed (Kunda 1990). Moreover, prestige concerns and/or the knowledge that one's forecast will be evaluated may lead to a 'reality constraint' so that factors favouring optimism bias, for example, may be tempered. However, when large adjustments are required, accuracy rewards may lead to risk averse behaviour and insufficient adjustment. Errors, and associated losses in rewards, resulting from wrongly making a large adjustment (i.e. an act of commission) may be regarded as worse than those resulting from failure to make a large adjustment (an act of omission) (e.g. see Ritov and Baron 1992).

In addition to motivational influences, forecasters all come to the task with relevant past experience which may affect the weightings they give to the different pieces of information they are presented with, whether in a real S&OP process or a simulated process. Individual forecasters typically face the task of forecasting hundreds of products (Fildes et al. 2009) so that their experience of the accuracy and reliability of the statistical forecasts which are provided to them and their prior understanding of the context may well affect the individual forecasts they make. For example, Franses (2014, p.86) found that more experienced forecasters in a pharmaceutical company produced more accurate adjustments. It might also be hoped that their knowledge of forecasting and promotions would lead to improved accuracy!

In summary, little is known about the way forecasters use information to produce their judgmental adjustments of statistical baseline forecasts. Yet it is an important issue in that judgmental adjustment has been shown to improve accuracy quite substantially but sometimes also to diminish it, inducing bias and inefficiencies not present in the statistical forecasts. In the remainder of this paper, we test our hypotheses by investigating the influences of the following factors on judgmental adjustments to statistical forecasts when a special event is imminent and when the different types of information are simultaneously available:

i. The statistical time series evidence, including the statistical time series forecast.

ii. The effect on sales of an earlier promotion.

iii. The number of positive and negative verbal statements relating to the potential success of a promotion.

We examine these influences while controlling for prior expectations of promotion effects, self-reported knowledge of forecasting and different types of motivation.

3. Methodology and experimental design

In order to test the hypotheses developed above, we have adopted a behavioural experimental approach, common now in the operations literature (see for example, Gans and Croson 2008, Croson et al. 2012, Siemsen 2011, Zhao et al 2013). Our experiments were designed to replicate the demand forecasting task that is common in supply-chain companies. From these summaries of the field of behavioural operations we see that, while experiments are a commonly adopted methodology, research into behavioural forecasting issues has been relatively rare (estimated as under 4% of behavioural research into operations in Croson et al. 2012). The advantages of using behavioural experiments here are (as usual) the ability to control the various factors hypothesized from the literature review as having possible effects on forecasters' adjustments.

We examine the forecast of the effect of a sales promotion which Fildes and Goodwin (2007) identified as the most common reason for forecasters making judgmental adjustment to statistical forecasts. From the field case studies we know that the final adjusted forecast summarises an often complex S&OP process where various pieces of information are gathered by the forecaster and used as a basis to produce the final forecast: the information used by the forecasters include market research, promotional features (size, coverage etc.) and weather factors (which might damp or amplify, Nikolopoulos et al. 2013).

The participants in the experiments were business and management students, studying for either bachelors, masters or doctoral degrees at the same universities as this paper's authors, Bath (UK), Bilkent (Turkey), Lancaster (UK), They had all studied some forecasting. While they obviously do not have the same experience as commercial forecasters, they have at least as much statistical training as many practicing forecasters.

The participants were asked to assume the role of a forecaster for a large company which supplies a wide range of products to supermarkets. They were told that their task was to predict the sales of a number of these products that would be subject to a sales promotion. Each participant was given a brief (see Appendix) describing the task and providing some base line information, including the average impact of a promotion at this supermarket. After consenting to the terms of the experiment, once started on the experiment proper, information potentially relevant was provided which included a statistical baseline forecast together with information on the factors that may affect the promotion's outcome. The information was delivered through a forecasting support system (FSS) (see Figure 1 for a typical screenshot). The screen has been designed to have features and a format that is similar to those found in some widely used commercial forecasting systems (e.g ForecastPro™) including a graphical display. This is important as some researchers have found that presenting time series data in a tabular, rather than a graphical format can have a strong influence on the accuracy of judgmental forecasts (e.g. Harvey and Bolger, 1996). It was hoped that the realism of both the system and the participants' task in interacting with the system would increase the ecological validity of our findings including the motivation of the participants (this has been the case in other fields such as marketing research involving choice experiments (see e.g. Rogers and Soopramanien).

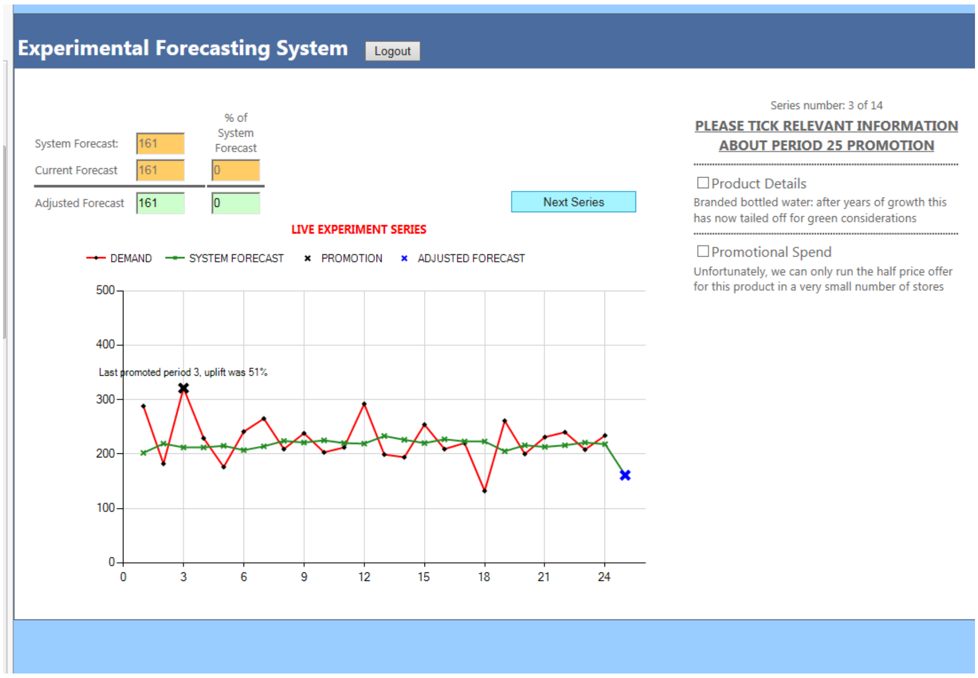


Figure 1 Screenshot of experimental forecasting support system

The participants first saw product details for a particular SKU, (the SKU's were presented in random order), a corresponding time series sales history of 24 periods and the corresponding statistical forecast for all periods including the 25th. The data were generated according to the rules:

$$Sales_t = 200 + \beta time + \varepsilon_t + \text{Pr}omotional\ Effect_t$$
$$\text{Pr}omotional\ Effect \sim Uniform(40,60)$$
$$\varepsilon_t \sim Normal(0, stddev)$$

The time trend parameter took three values: 0 and ± 1%, while the standard deviation had values of 40 and 80. On the rare occasions where the simulated observation turned out negative, a value of 0 was substituted.

The FSS provided a simple exponential smoothing forecast (based on a smoothing parameter of 0.2) so that the baseline forecast for period *t* is given by:

$$BaseLine\ Forecast_t = 0.2 * \text{Sales}_{t-1} + 0.8 * BaseLine\ Forecast_{t-1}$$

For promoted periods, *t*, the previous baseline forecast is not updated:

$$BaseLine\ Forecast_t = BaseLine\ Forecast_{t-1}$$

For the period to be forecast (period 25), a random perturbation was added to the forecast. This was done by assigning each series a value of 0, or ±50*U(0.4,0.6), i.e., a random perturbation of between 20 and 30 in absolute value. This limited the collinearity between the forecast, previous sales observation and previous error, allowing its influence on the adjustment to be estimated more precisely. It was made clear that the baseline forecast did not include any promotional effects.

For each SKU, the participants were invited to use their judgment to adjust the baseline forecast to take into account the promotion. To inform their judgment they were supplied on the screen with several items of information. First the time series history included the effect of an earlier promotion. The timing and effect of this promotion varied across SKUs but the mean sales uplift (relative to the baseline forecast) was 50%. Secondly, each screen carried a reminder that the mean (base rate) for promotional uplifts at this supermarket was 50%.

Finally, for each SKU the screen displayed between zero and four written statements which gave reasons suggesting why the level of sales uplift achieved by the forthcoming promotional effect would be above or below the average ('positive' and 'negative' reasons). These reasons related to the amount spent on the promotion (e.g. "Over £1m is being spent on the promotion, double the usual size"), market research (e.g. "Focus groups have been quite negative about the promotional packs, but we can't change these at this late stage."), weather factors (e.g. "This product is mainly sold in the North where the weather conditions should be good for high sales according to the latest forecast.") and campaign effectiveness (e.g. "We were hoping for a celebrity endorsement of our product as part of the campaign, but negotiations have not been successful and, unfortunately, we will have to run the campaign without this endorsement"). A full list of reasons is available from the authors. Half of these were positive and half were negative. The number of reasons displayed at any one time, the appearance of positive or negative reasons and the order of their display were all randomized.

Because of the complexities of designing a realistic experiment, a number of preliminary experiments were run, involving over 200 participants, to fine tune the design and screen display, to eliminate potential confounded factors and to identify the key issues that merited further investigation. In the early experiments 24 data series were used, including, 6 trial series. In order to lessen the experimental burden the number of series was later lowered to 14, including two trial series, and checks confirmed that this had no significant effect on the results. Further variations included having a fixed number of four reasons, of which 0 to 4 were positive, displaying just 2 reasons, colour-coding these reasons (green for positive and red for negative), forcing participants to select a primary and secondary reason to support their adjustment, having an average promotional uplift of 80%, providing a baseline forecast for period 25 without the random perturbation and having treatments that reversed the sequence of the reasons provided to another treatment or counterbalanced these reasons by displaying negative reason for each positive and vice versa. In the main, the results of these experiments were consistent with those that we discuss next so, for brevity, they will not be reported here. We will refer to the few cases where there were differences in the Discussion Section 5.

In the final experiment, which we report on in detail here, participants were randomly assigned to three treatments that were designed to provide different types of motivation. The first group were told that they would be rewarded when a promotion uplift exceeded 50% (although this was beyond their control, it was thought that the possibility of this reward might lead to desirability bias). The second group were told that they would be rewarded for the accuracy of their forecasts. A third (control) group were given a reward merely for participating in the experiment. This led to a 3 (motivation type) between subjects x 12 (SKUs) within subjects design.

Before embarking on the experiment, participants were asked to indicate what they thought a typical percentage sales uplift would be for a fast-moving consumer good that was being promoted. They then made forecasts for the 2 SKUs that were used as a trial run to familiarise themselves with the FSS. For each SKU they had the option of indicating which, if any, of the displayed reasons had led them to make their adjustment. During the trial run they were provided with an assessment of why the earlier promotion had, or had not been, a success with overall feedback on their accuracy given after forecasts had been made for both SKUs. No feedback was provided in the main part of the experiment. The best two forecasters in each treatment received an Amazon voucher or in the case of the control group, a prize draw was used to select the two winners. At the end of the experiment participants completed a questionnaire designed to assess their knowledge of forecasting, their engagement in the task, their expectations regarding the accuracy of their judgmental adjustments and their interpretation of the reasons that were provided.

4. Experimental analysis and results

4.1 Exploratory data analysis

133 participants took part in the experiment. We excluded respondents who did not make any but the very smallest average adjustments (i.e. their mean adjustment was less than 0) as this suggests either a limited understanding of promotional effects in retailing or no engagement with the experiments. The results are therefore based on a sample of 112 participants.
As indicated above, participants responded to a post-experimental questionnaire. The main results of interest are summarised in Table 1.

Table 1 Questionnaire responses

Question	Mean	Std.Dev.
Rating of overall knowledge of demand forecasting	2.77	0.86
Expectations of statistical forecast performance	3.03	0.77
The provided reasons had a direct influence on my forecasts	3.46	1.07
Confidence in my final adjusted forecast	2.66	0.94
Motivation to engage with the task	3.40	0.98

Scale: (1) None / low expectations, to (5) High / high expectations - depending on question

The results show participants were generally motivated by the experiment and responded to the reasons provided. Typically, they did not 'write-off' the potential performance of the statistical baseline forecasts, despite the fact that they were bound to have large errors in a promotion period. This may reflect some acknowledgment of the statistical forecasts' usefulness in establishing a reliable baseline for judgmental adjustment. The participants also indicated a lack of confidence in the accuracy of their adjusted forecasts which is reasonable given the level of uncertainty associated with the promotion effects and the relatively small amount of information that was available to support their judgments.

We next look at the change in the estimated uplift from the participant's prior estimate of promotional effects (that would be expected in similar circumstances to those simulated in the experiment) to their actual adjustments during the experiment. Table 2 presents 95% trimmed means and medians for the participants' prior estimates, the adjustments they made during the experiment and the differences between these two values (the trimming being based on the adjustment percentage).

Table 2 Prior expectations of uplift and uplift during experiment

	Prior estimate	Adjustment during experiment	Change in estimate
Mean	50.8%	30.8%	-20.0%
Median	50.0%	30.0%	-15.0%

The distribution of the percentage adjustment is broadly normal with a few positive outliers. 25% of the adjustments were greater than the advertised uplift of 50%. Some were as high as 200%, which is quite reasonable for the sorts of products we have included in our experimental design. Overall the results in Table 2 indicate that the participants tended to underestimate the promotional effects both with regard to the information provided on promotional effects and also the past promotional evidence in the time series. Moreover, they provided lower estimates during the experiment, despite the information provided as to the typical promotional effects, thereby neglecting the base rate and hence providing support for H9. The possible reason for this will be considered in the Discussion section.

4.2 Statistical modelling

The nature of the experiment where each respondent is sequentially given a number of series in random order requires a more sophisticated analysis than a standard ANOVA or regression. We use the following linear mixed-effects model (Verbeke and Molenberghs, 2000).

$$\mathbf{Y}_i = X_i\beta + Z_i\mathbf{b}_i + \varepsilon_i$$
$$\mathbf{b}_i \sim N(0, D)$$
$$\varepsilon_i \sim N(0, \Sigma_i)$$

where $\mathbf{Y_i}$ **is** the n_i dimensional response vector for respondent *i*, representing the promotional estimates for the ith series. X_i and Z_i are the $n_i\ x\ p$ and $n_i\ x\ q$ of the factors influencing the response while β is the p dimensional vector of fixed treatment effects and $\mathbf{b_i}$ is the q dimensional vector of random effects. The covariance matrices are potentially important to the model building. D and Σ are assumed independent. A repeated measures design is needed as the observations of the promotional uplift estimates from a given subject cannot be assumed independent of each other, for example in the sequence in which they were made.

The standard assumption made for the variance-covariance matrix of the random effects, *D*, is that the respective variances of the $\mathbf{b_i}$ differ but are independent of each other – labelled the variance component assumption. In addition the sensitivity of the estimated effects to changes in this assumption has been tested through an assumed autoregressive structure to capture any carry-over effect between the repeated observations, i.e an AR(1) structure was assumed for D[1].

The key features of the linear mixed effects model are set out below.

- The dependent variable is the adjustment percentage transformed into log (100+Adjust_Percent).
- The effects of variables relating to the past forecast history were assumed to be random effects as they depend on the individual participant. These variables were: the log of the respondent's prior estimates of promotional effects, the log of the last forecast error, the log of the uplift achieved in the last promotion (i.e. actual promoted sales over the baseline forecast), the log of latest forecast for the promoted period and the timing of previous promotion.
- The effects of the series characteristics, the trend [2]and noise variance were treated as fixed effects class variables.
- The number of positive and negative reasons were treated as fixed effects class variables, i.e. treatments.

In addition, the results presented have points of high leverage removed. Leverage was measured using Cook's D (eliminating points with D>.002 – approximating one of the recommended cut-offs of 4/n). Various modelling choices needed to be resolved, in particular how to characterise the number of negative and positive reasons. Several alternatives were considered, including using both variables (with an interaction) and one variable together with the difference between positive and negative reasons. Using the variable *Reasncat* (defined as the number of positive reasons minus the number of negative reasons) proved the most parsimonious specification with minimum BIC. A sensitivity check on the assumption

[1] SAS 9.3 has been used in estimating this model using Restricted Maximum Likelihood. Details of the various defaults used are given in http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mixed_sect007.htm.

[2] The trend was dropped in the results we report as it proved to have little impact on the results from the preliminary experiments

of the correlation structure of the repeated measures did not show any substantive differences.

4.3 Results of modelling

The results from the model are shown in Table 3.

Table 3 Model of the percentage adjustment ($\log_e$(100+Adjustment%))

Effect	Estimate	p-value†
Intercept	0.5048	<.0001
ln(last uplift)	0.2748	<.0001
ln (last actual)	0.03696	0.0015
ln (last stats forecast error)	0.03994	0.1051
ln(current stats forecast)	-0.1281	<.0001
ln(Prior)	0.0347	0.0141
Noise Variance	0.0214	0.0179
Timing	.0010	0.0357
Reasncat = -4	-0.1067	<.0001
Reasncat =-3	-0.1356	<.0001
Reasncat = -2	-0.1140	<.0001
Reasncat = -1	-0.0767	<.0001
Reasncat = 0	-0.0775	<.0001
Reasncat = 1	-0.0517	0.0014
Reasncat = 2	-0.0228	0.1120
Reasncat = 3	-0.0272	0.1004

[Available n= 1560; sample size after deleting high leverage points=1309]
[*Reasncat* = No. of positive reasons supplied – No. of negative reasons]

† All tests are one-sided apart from that for the noise variance

Effect of time series characteristics and baseline forecast

The results provide support for H1 (higher sales in the previous promotion will be associated with a larger upwards adjustment), H2 (the more recent the previous promotion, the higher the adjustment -though the effect size was relatively small), H3 (the higher the most recent sales figure, the higher the adjustment), H5 (the higher the statistical baseline forecast, the smaller the adjustment will be). H4 (larger statistical baseline forecast errors will be associated with higher adjustments) was not supported. This may because larger values of the most recent observation will be correlated with the most recent forecast error (r = 0.55) so that, rather than amplifying the salience of the most recent observation hypothesis, the most recent error was simply seen as a duplicate cue and was therefore regarded as redundant by the participants. H6 (levels of noise in the series will not be associated with the extent of the upwards adjustment) was also rejected with the higher observed variation leading to higher adjustments.

Effect of Reasons

Figure 2 shows the relationship between $\log_e$(100+Adjustment%) and the difference between the number of positive and the number of negative reasons. The effects are compared with situations where there are an equal number of positive and negative reasons. Figure 3 shows the effect of marginal increases in the difference between the number of positive and negative reasons.

Figure 2 Effects on $\log_e$(100+Adjustment%) of differences between the number of positive and negative reasons relative to situations where there are an equal number of positive and negative reasons

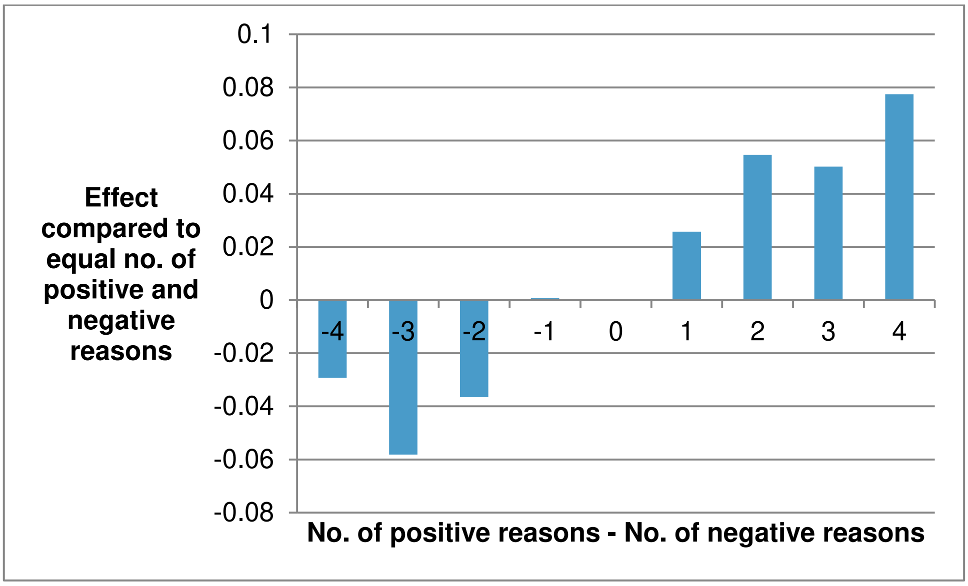
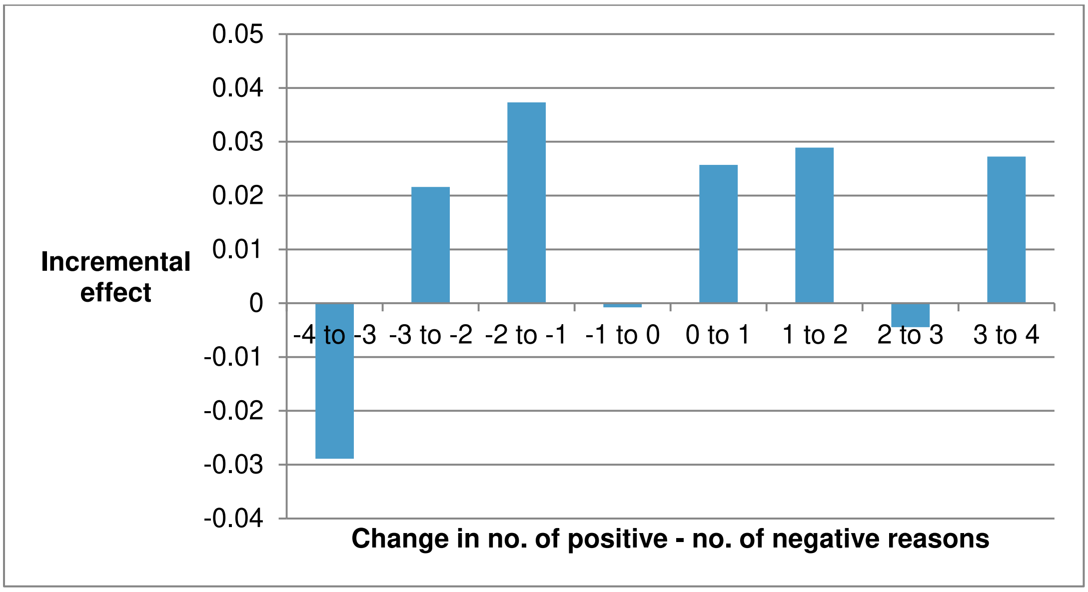


Figure 3 Incremental effect on $\log_e$ (100 + adjustment %) of change in number of positive minus number of negative reasons



It can be seen that, in general, the greater the number of positive reasons relative to the number of negative, the larger the upwards adjustment. This suggests that participants were balancing the reasons against each other, indicating that they were using a compensatory strategy, which is contrary to H7. However, there is little evidence of a difference between adjustments made when the number of negative reasons exceeds the number of positives by 4,3,or 2 (i.e., -4,-3 and -2 in Figure 2). In these cases there is an average reduction in the upwards adjustment of 4% compared to the neutral 0 category. This suggests that once the number of negative reasons have exceeded the number of positives by 2, any extra negative reason will have no effect (we discount the non-monotonic estimate for -3). This is mirrored when the majority of reasons are positive: an excess of 2, 3 or 4 positive reasons also leads to similar sized adjustments (with an average increase of 6% compared to the neutral category although four positive reasons leads to a significantly.larger increased adjustment). We have tested this out further by considering the contrast between the 0 category and the average effect of experiencing positive reasons alone. The effect is approximately 5% (this is significant:, p-value <.0001).

Did negative reasons have a greater influence than positive reasons? In part an answer to this depends on the mix of reasons available (i.e. a change from the 0 position with 1 positive and 1 negative reason to a situation where there are 2 positive reasons and 1 negative differs in impact compared to the difference between 0 positives, 0 negatives and 1 positive, 0 negative). The following test illuminates the comparison further by examining the difference in the sizes of the logs of the upwards adjustments between having one (two) more positive reason versus one (two) more negative. This yields the results in Table 4.

Table 4 Relative effect of extra positive and negative reasons

Contrast	**$\log_e$(100+Adjustment%)**	t Value	Pr > |t|
1 more positive vs 1 more negative reason	0.026	1.87	0.0617
2 more positive vs 2 more negative reasons	0.018	1.07	0.2834

This shows that having one more positive reason is somewhat more impactful than one more negative reason (although the p-value is 0.06), but there is little difference between having 2 more positive compared 2 more negative reasons. Overall, the results suggest that positive reasons have slightly more effect than negative ones, which is contrary to H8.

Motivation and participants' characteristics

There is an apparent country effect between the participants based in the UK and Turkey ($p<.001$) with the latter providing lower forecasts of uplifts. Once individual priors were included the effect was insignificant. This probably reflects the different retail environments that the participants were familiar with. There were no other substantive or significant effects on the size of the adjustment relating to the different motivation treatments or the characteristics of the participants, such as their knowledge of statistical forecasting, apart from the finding that participants' motivation in the task proved significant in increasing their average uplift. Their initial estimates of promotion effects were also significant but the carry-over effect was small.

Overall, these results suggest that when making their forecasts, participants were aiming to match the previous promotion's sales by adjusting the statistical baseline forecast until it was similar to this value. However, this strategy was moderated by the recency of the previous promotion, the level of sales achieved in the most recent period, the level of noise in the series and the difference between the number of positive and negative reasons that were presented to them. The advertised base rate (i,e. an average uplift of 50%) appeared to have been discounted, heavily influenced by experimental information.

5. Discussion

In general, the results reported in the last section were consistent with those obtained in the preliminary experiments, which were used as part of the experimental design process, suggesting that the results are robust. The only noteworthy difference was that the log of the previous statistical forecast error had a significant effect on the size of the upwards adjustment in the preliminary experiments, but in the main experiment the effect was not significant. In the preliminary experiments the forecasts were obtained though exponential smoothing so a positive error in the most recent period would be followed by a higher baseline forecast for the forthcoming period. Thus a large positive error would be likely to amplify the effect of a higher forecast as it enhanced the impression that the baseline had increased. In the main experiment the random perturbations that were applied to the statistical baseline forecast may have nullified this potential amplification. For example, a positive error could be followed by a forecast that was lower than its predecessor.

Three findings are of particular interest. The first is the large range of information that participants drew upon to make their forecast adjustments. Almost all the cues available to them had a significant effect on the size of the adjustments despite the large amount of information presented and the mix of quantitative ('hard') and verbal ('soft') information. Surprisingly, despite the volume of information presented and the number of SKUs they had to produce forecasts for, they appeared to apply a compensatory strategy when using the reasons to assess the required adjustment.

In many cases the cues were used inappropriately. For example, the timing of the last promotion should have had no influence on the judgments and the optimal estimate of promotional uplift was not related to the time series history. In particular, the most recent observation was found to carry undue influence, thereby replicating the findings of time series extrapolation studies where no special events are present. Moreover, the participants' attempts to match a single past observation (i.e. the sales achieved by the previous promotion) ignored the lack of diagnosticity that is inherent in a single case - using the advertised base rate of 50% uplift as a starting point would have been a superior strategy. Anyway, the sales achieved in the previous promotion were less relevant than the extent to which these sales exceeded the statistical baseline forecasts for that period. Nevertheless, it is possible that the use of a wide range of information was made possible through the use of a forecasting support system which enabled all the information to be presented simultaneously on single screen. It is possible that this information-rich presentation enabled the participants to adopt a simple cumulative additive strategy : starting with the previous promotion's sales, an upward or downward adjustment could be made from the current estimate as each new cue was examined. Only the estimate based on the cues examined so far and the effect of the new cue would need to be considered at any stage during the process.

The second interesting finding is that, despite the wide range of information employed by the participants, there was a consistent bias towards under estimating the effect of the forthcoming promotions (this was also a consistent finding in all of the preliminary experiments). The bias occurred despite the fact that the 50% average uplift was highlighted both in the cover story and in information presented on the computer screen during the trial run. Telling participants that they would be rewarded if the uplift exceed 50% (in an attempt to induce desirability bias) had no effect, nor did rewarding accuracy.

We considered four possible reasons for this: (i) the participants were conservative in that their prior expectations of typical uplifts caused them to discount the provided information on the base-rate for promotions, (ii) the negative reasons carried more weight than the positive reasons, (iii) the participants regarded the advertised average uplift as a best-case scenario that only applied when positive reasons were provided or (iv) the previous promotion effect and the statistical forecast both acted as anchors so the estimated uplift tended to be set at a point between them. The first explanation seems implausible, given the results in Table 1, where it can be seen that the average adjustments made during the experiments were actually substantially lower than the prior estimates. While participants' prior estimates were significantly associated with the ones made during the experiment the carry-over effect was low (and insignificant in the preliminary experiments). Regarding the second explanation, as indicated earlier, there was no evidence that negative reasons carried greater weight. In fact there was a slight tendency for positive reasons to have more impact on the adjustments. The third explanation is also not supported: the mean adjustment in cases where there were no negative reasons were supplied was only 30.8%. Even when there were 4 positive reasons is was only 43.9%, still short of the base-rate.

In relation to the fourth explanation, Hoch and Schkade (1996) found that, when people adopted a pattern matching strategy, the point they were trying to match acted as the anchor with (insufficient) adjustments being made based on the conditions that applied to the case being forecast. Given that the task required an adjustment to the statistical forecast this could have acted as a second anchor. Anderson's (1965) integration model suggests that anchoring and adjustment can be modelled as a weighted average of a starting or initial value (i.e. an anchor, in this case the statistical baseline forecast), and an estimate that the person would have made had they not seen the anchor (in this case the sales in the previous promotion). This would tend to place the estimated uplift below the 50% base rate, as was observed. Thus this explanation appears to be the most plausible.

Is the underestimation of promotion effects we found in our laboratory experiments typical of what happens in the field? Several field studies have reported that judgmental adjustments tend to suffer from a optimism bias (Fildes et. al. 2009, Franses and Legerstee 2011), the opposite to what we have found here. However, neither of these studies confined their analysis to specifically forecast periods when a special event was due to occur. Moreover, a study of the sales forecasts of German companies, again not limited to special periods and not limited to adjustments of statistical forecasts, found a pessimism bias (Muller 2011). It was argued that because these forecasts were made anonymously to a panel there was no incentive to 'overgloss' the forecasts and instead the forecasts were the result of loss aversion where a negative surprise was disliked more than positive ones were liked. One study that focused on promotion periods did find an optimism bias, but in the manufacturing company studied, the sales in promotion periods were on average only 8.7% higher than those of the statistical baseline forecast (Trapero et al. 2013). It is unclear whether over-optimism would still have been evident if the typical uplift had been as high as 50%, as in our experiment.

The third interesting finding is that a larger average upwards adjustment was made by those who chose to indicate which reason had the greatest influence on their adjustment (the median adjustments were 30% compared to 25%.) For those participants who also included the SKU characteristics as one of their reasons behind their adjustments, predicted adjustments were 30% compared to 20%. This suggests that a lower uplift is predicted when neither the reasons nor the SKU characteristics resonate, a sensible strategy though it ignores the base rate of 50%. Table 4 gives the overall median uplifts where at least one positive reason was selected as a reason for the uplift compared to a negative reason and where no reasons were selected. Interestingly, the largest adjustments tended to be made when both positive and negative reasons were selected. It is unclear why this was the case. It might have reflected greater engagement with the task so that greater attention was paid to the base-rate; or alternatively, the heavier cognitive load of attempting to consider both types of reasons may in some way have led to larger adjustments. Typically with more positive (negative) reasons available participants were more likely to select a positive (negative) reason. This has implications for how information is captured in the S&OP process and the design of the FSS.

Table 4 The effects of selecting a positive (negative) reason on the predicted uplift. The number of times the particular cell was selected is shown in brackets.

Available reasons	Selected			
	Positive selected	Negative selected	Both positive and negative selected	None selected
More positive	18.5 (46)	8 (10)	49 (30)	4 (15)
More Negative	15.5 (20)	14.5 (42)	34.5 (44)	30 (30)
No difference	21.5 (34)	18 (27)	41 (32)	30 (307)
Only positive	42 (241)	n.r.	n.r	12 (85)
Only Negative	n.r	25 (213)	n.r.	23 (120)
Overall	37 (341)	24 (292)	40 (106)	25 (557)

The analysis using the mixed model can also illuminate the effects of selection. Where there are more positive than negative reasons available and a positive one is selected, there is a higher (significant) uplift than the case where no reason is identified.

Finally, none of our motivation treatments had a significant effect on the size of adjustments. While these were not the main focus of the study –they were intended to control for possible motivational effects - the absence of a desirability bias for those who were rewarded for higher than average uplifts was surprising. This once again demonstrates that it is difficult to replicate motivational and associated political effects that occur in the field in the laboratory. A small reward of a voucher for a higher sales forecast or an accurate one is not the same as the incentive to please the boss with a high forecast or the incentive to bring kudos and resources to one's department by producing reliable forecasts. In particular, in the field there may be a sense of personal responsibility for the success of a promotion leading to wishful thinking and advocacy bias (i.e. 'the tendency of product planners to champion their project by overpromising on forecasts' (Tyebjee 1987)), and there could also be prestige-related motivations leading to various distortions and biases

6. Conclusions

Information from actors engaged in supply chain planning is potentially crucial to effective demand forecasting. In the S&OP process, information is exchanged between production, logistics, finance, sales, marketing and the demand planners in attempts to reconcile demand with supply availability. Given their supply chain and financial repercussions, promotions pose particularly sharp challenges to S&OP decision-makers. This experimental study has examined how forecasts of the effects of promotional events are produced using a forecasting support system that closely replicates those observed in practice. The results match those seen in the field with various heuristics being adopted by the forecasters. In particular, past observations on the SKUs (Fildes et al. 2009) as well as the exemplar previous promotions (Cooper et al. 1999) are used by participants despite leading to inefficiencies; that is to say, reweighting these pieces of information could potentially lead to more accurate forecasts of uplift. A critical element in the S&OP process is how external sources of information (provided through either face-to-face meetings or, as is becoming more mainstream, electronically (Weller and Crone 2102), as in this study) are used to improve on a baseline statistical forecast. The experimental evidence reported in this work showed that participants took into account the positive (and negative) pieces of information using a compensatory strategy, though at the extremes, additional pieces of positive or negative information had less impact. Furthermore, participants also appeared to take into account extraneous information damaging to forecast accuracy. The validity of the results from the experiment presented in detail here is demonstrated by their robustness over the two earlier experiments where information was supplied in a somewhat different form and with different sized promotional effect.

The research has important implications for the design of forecasting support systems and also the S&OP process itself. The participants (and from the field evidence, practicing forecasters) mis-interpret the time series history: they choose not to accept the accuracy of the statistical baseline forecast, using this and past observations to reweight the forecast, introducing additional forecast error and displaying a case of ‘algorithm aversion’ (Dietvorst et al.,2015). They also appear to adopt a version of the ‘last-lift’ heuristic, the most common promotional forecasting method used in practice. Their mistake is to ignore the average uplift, instead focussing on the last observed value. It may be thought that a possible explanation for this lies in the experimental subjects trying to outguess the experimenters, but what we have observed reflects what has already been seen in the field. Both facets of the participants sub-optimal forecasting suggest that FSSs need to be redesigned: the baseline forecasting algorithm would need to be carefully constructed and in particular, evidence as to its effectiveness provided; though, as Dietvorst et al. (2015) and Lim and O’Connor (1995) have shown, changing the habit of misweighting remains difficult. Second, evidence on promotional effects needs to be presented in a form that remains salient but overrides the randomness of a particular past promotion, perhaps following some of Lee et al.’s (2007) suggestions.

Like most experimental studies, this work has limitations in its use of students, despite their motivation and knowledge which closes some of the gap with practicing forecasters. In addition, while the on-screen simulation mirrored the operational realities of forecasting closely, the demand model and the promotional effects were based on a simple statistical model. It is both of theoretical interest and practical significance to move one step closer to practice with a demand model that captures some promotional drivers: the issue confronting experimental participants and forecasters then becomes whether a forecaster facing the diversity of information that the S&OP process delivers through a real FSS (such as SAP-APO) can see the wood for the many trees. With the limitations of current systems (Fildes et al. 2006), it seems there is substantial scope for design innovations.

References:

Allen, M., & Preiss, R. W. (1997). Comparing the persuasiveness of narrative and statistical evidence using meta-analysis. *Communication Research Reports*, *14*, 125-131

Aviv, Y. (2001). The effect of collaborative forecasting on supply chain performance, *Management Science*, *47*, 1326-1343.

Ben-David, I., J.R. Graham, C.R. Harvey (in press). Managerial miscalibration, *The Quarterly Journal of Economics*, doi: 10.1093/qje/qjt023

Cooper, L. G., Baron, P., Levy, W., Swisher, M., & Gogos, P. (1999). PromoCast (TM): A new forecasting method for promotion planning. *Marketing Science, 18*, 301-316.

Croson, R., Schultz, K., Siemsen, E., & Yeo, M. L. (2013). Behavioral operations: The state of the field. *Journal of Operations Management, 31*, 1-5.

Danese, P., M. Kalchschmidt (2011a). The role of the forecasting process in improving forecast accuracy and operational performance, *International Journal of Production Economics, 131*, 204-214.

Danese, P., M. Kalchschmidt (2011b). The impact of forecasting on companies' performance: Analysis in a multivariate setting, *International Journal of Production Economics, 133*, 458-469.

Dawes, R.M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571-582.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology-General, 144*, 114-126.

Eksoz, C., S.A. Mansouri, M. Bourlakis (2014). Collaborative forecasting in the food supply chain: A conceptual framework. *International Journal of Production Economics. 158*, 120-135.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces, 37*, 570-576.

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 judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25*, 3-23.

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

Franses, P.H. (2014) *Expert Adjustment of Model Forecasts*, Cambridge: Cambridge University Press.

Franses, P.H., R. Legerstee (2010). Do experts' adjustments on model-based SKU-level forecasts improve forecast quality? *Journal of Forecasting*, 29, 331-340.

Franses, P.H., R. Legerstee (2011). Experts' adjustment to model-based SKU-level forecasts: Does the forecast horizon matter? *Journal of the Operational Research Society*, 62, 537-543.

Franses, P.H. and Legerstee, R. (2011). Combining SKU-level sales forecast from models and experts. *Expert Systems with Applications. 38,* 2365-2370**.**

Franses, P.H., R. Legerstee (2013). Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting*. 29, 80-87.

Goodwin, P. (2005). Providing support for decisions based on time series information under conditions of asymmetric loss. *European Journal of Operational Research. 163*, 388-402.

Goodwin, P., Lee, W.-Y., Fildes, R., Nikolopoulos, K., & M. Lawrence, M. (2007). Understanding the Use of Forecasting Systems: An Interpretive study in a Supply-Chain Company. University of Bath, School of Management Working Paper Series, 2007.14.

Harvey, N, (2007). Use of heuristics: Insights from forecasting research. *Thinking and Reasoning. 13,* 5-24.

Harvey, N., F. Bolger. and A.G.R. McClelland. 1994. On the nature of expectations. *British Journal of Psychology. 85*, 203-229.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting. 12*(1), 119-137.

Hoch, S. J., & Schkade, D. A. (1996). A psychological approach to decision support systems. *Management Science*. *42*(1), 51-64.

Hong, S., & Park, H. S. (2012). Computer-mediated persuasion in online reviews: Statistical versus narrative evidence. *Computers in Human Behavior. 28*, 906-919.

Hogarth, R. (1987) *Judgment and Choice: The Psychology of Decision* 2nd edition, Chichester: Wiley.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review. 80*, 237-251.

Kahneman, D., & Tversky, A. (1979). Prospect theory – Analysis of decision under risk. *Econometrica. 47*, 263-291.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin. 134*, 404-426.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin. 133*, 95-121.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin. 108*, 480-498.

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

Lapide, L. (2007). Sales and operations planning (S&OP) mindsets, *Journal of Business Forecasting.* Spring, 21-31.

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

Lee, H.L., K.C. So, C.S. Tang (2000). The value of information sharing in a two-level supply chain, *Management Science. 46*, 626-643.

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting. 23*, 377-390.

Legerstee, R., P.H. Franses (2014). Do experts' SKU forecasts improve after feedback? *Journal of Forecasting. 33*, 69-79.

Leitner, J., & Leopold-Wildburger, U. (2011). Experiments on forecasting behavior with several sources of information – A review of the literature. *European Journal of Operational Research. 213*, 459-469.

Lim, J.S. and O'Connor, M. (1995), Judgemental adjustment of initial forecasts: its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*, 149-168

McGregor, D. (1938). The major determinants of the prediction of social events, *Journal of Abnormal Psychology*, *33*, 179-204.

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

Moore, D. A., Tenney, E. R., & Haran, U. (in press). Overprecision in judgment. In G. Wu and G. Keren (Eds.), *Handbook of Judgment and Decision Making*. New York: Wiley.

Müller, H. C, (2011). *Forecast Errors in Undisclosed Management Sales Forecasts: The Disappearance of the Overoptimism Bias*, DICE discussion paper, No. 40

Nikolopoulos, K., & Fildes, R. (2013). Adjusting supply chain forecasts for short-term temperature estimates: a case study in a Brewing company. *IMA Journal of Management Mathematics, 24*, 79-88.

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

Oliva, R., N. Watson (2011). Cross-functional alignment in supply chain planning: A case study of sales and operations planning. *Journal of Operations Management. 29,* 434-448.

Olsen, R.A. (1997). Desirability bias among professional investment managers: Some evidence from experts, *Journal of Behavioral Decision Making. 10*, 65-72.

Önkal, D., and E. Aktas (2011). Supply chain flexibility: Managerial implications. In D. Önkal and E. Aktas (eds.), *Supply Chain Systems-Pathways for Research and Practice*, Croatia: Intech Publ., pp 75-84.

Önkal, D., K. Z. Sayım, and M.S. Gönül (2013). Scenarios as channels of forecast advice, *Technological Forecasting & Social Change. 80*, 772-788.

Payne, J.W., Bettman, J.R. & Johnson, E.J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.

Reimers, S., & Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting. 27*, 1196-1214.

Reinard, J. C. (1988). The empirical study of the persuasive effects of evidence – The status after 50 years of research. *Human Communication Research. 15*, 3-59.

Ritov, I., & Baron, J. (1992). Status-quo and omission biases. *Journal of Risk and Uncertainty. 5*, 49-61.

Rogers, G., & Soopramanien, D. (2009). The truth is out there! How external validity can lead to better marketing decisions. *International Journal of Market Research. 51*, 163-180.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review. 5*, 296-320.

Seifert, M., Siemsen, E., Hadida, A. L., & Eisingerich, A. B. (2015). Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management. 36*, 33-45.

Siemsen, E. (2011). The usefulness of behavioral laboratory experiments in supply chain management research. *Journal of Supply Chain Management. 47*(3), 17-18.

Stahl, R.A. (2010). Executive S&OP: Managing to achieve consensus. *Foresight: The International Journal of Applied Forecasting.*, Fall, 34-38.

Thomé, A.M.T., L.F. Scavarda, N.S. Fernandez, A.J. Scavarda (2012). Sales and operations planning: A research synthesis, *International Journal of Production Economics. 138*, 1-13.

Tyebjee, T. T. (1987). Behavioral biases in new product forecasting. *International Journal of Forecasting, 3*, 393–404.

Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting. 29*, 234-243.

Tuomikangas, N., & Kaipia, R. (2014). A coordination framework for sales and operations planning (S&OP): Synthesis from the literature. *International Journal of Production Economics, 154*, 243-262.

Zhao, X., Zhao, X., & Wu, Y. (2013). Opportunities for research in behavioral operations management. *International Journal of Production Economics. 142*, 1-2.

Weller, M., & Crone, S. F. (2012). Supply Chain Forecasting: Best Practices & Benchmarking Study. Lancaster University. http://www.lancaster.ac.uk/lums/forecasting/material/

Windschitl, P.D., A.R. Smith, J.P.Rose, Z. Krizan (1020). The desirability bias in predictions: Going optimistic without leaving realism, *Organizational Behavior and Human Decision Processes. 111*, 33-47.

Yu, Z., H. Yan, T.C.E. Cheng (2001). Benefits of information sharing with supply chain partnerships, *Industrial Management & Data Systems. 101*, 114-121.