Lancaster University Management School
Working Paper 2016:4

Use and misuse of information in supply chain forecasting of promotion events

Robert Fildes*, Lancaster University: Paul Goodwin, University of Bath and Dilek Önkal, Bilkent University, Turkey

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

Revised, June 2016

Use and misuse of information in supply chain forecasting of promotion events

Abstract:

Demand forecasting is a critical component of sales and operations planning (S&OP) and is pivotal in supporting inventory and production planning in supply chains. Because of their relative infrequency the effects of sales promotions can be particularly difficult to forecast - yet these are events where production and inventory planners need clear guidance on the probable uplifts in demand. A widely-documented practice involves judgmentally adjusting a baseline statistical forecast on receipt of shared information from sales, marketing and logistics. However, much of this information will either have no predictive value in estimating demand uplift resulting from the promotion or its predictive diagnosticity will be unknown. Theoretical arguments on 'system neglect' and 'base rate discounting' suggest that the provision of information with no or unknown diagnosticity would lead to the forecasters being distracted from the underlying base-rate uplift with deleterious effects on forecast accuracy. This study investigates this possibility when forecasters made judgmental adjustments to forecasts via a forecasting support system (FSS) in advance of forthcoming sales promotions. In experiments forecasters were provided with the mean rate of sales uplift achieved through promotions (the base rate), and a baseline statistical forecast, together with both quantitative and qualitative information relating to a range of products that were due to be promoted. The results revealed that forecasters were distracted from the base rate, misinterpreting the diverse information available to them, and this led to underestimates of the uplift achieved by the promotions. By extending earlier findings from field observation to a representative experimental setting, these findings have important implications for the quality of inventory decisions, for the design of organizational S&OP processes, and for the implementation of the FSSs that such processes rely on.

Key Words: Sales and Operations Planning; behavioral operations; system neglect hypothesis; information effects; forecaster behavior.

1. Introduction

Production and inventory planning, scheduling, logistics, marketing and finance in supply chain companies all rely on short-term disaggregate forecasts at the 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 (e.g., Lapide 2007, Stahl 2010). The sales uplifts achieved through sales promotions campaigns are particularly difficult to forecast because of the relative infrequency of such events. When promotion campaigns are due to take place the forecasts within S&OP are usually produced as a combination of a simple baseline statistical forecast and a judgmental adjustment which is an estimate of the promotion effect (Fildes and Goodwin 2007). The adjustments are made to reflect information received from different departments such as production and marketing. Such adjustments may mirror individual and functional biases stemming from informational blindspots as well as other organizational misalignments in supply chain processes (Oliva and Watson, 2009, 2011).

In one of the few detailed case studies of forecasting practice, Goodwin et al. (2007) found that the benefits of judgmental adjustments based on additional information in a pharmaceutical company were slight and often negative. Other studies have found evidence that information use is inefficient and biased (Fildes et al. 2009, Franses and Legerstee 2010, 2011, 2013) and, where promotions are concerned, adjustments can therefore have a deleterious effect on forecast accuracy (Trapero et al. 2013). While the consensus is that integrating diverse sources of information is valuable (Kremer et al. forthcoming) and that forecast information sharing affects supply chain performance (Ozer and Raz, 2011; Ozer et al, 2011), no studies have examined the nature of the information that was available and how it was used. In the case we examine here the incorporation of promotion information is considered. Promotions were shown by Fildes and Goodwin (2007) to be the most important reason behind judgmental adjustments of demand forecasts.

This paper aims to help address this research gap by identifying how supply-chain-based forecasters respond when they have base-rate information indicating the average sales uplift achieved during sales promotions, together with other information that has either no, or unknown, diagnosticity[1] - a situation

[1] The diagnosticity of a piece of information is a measure of its helpfulness and usefulness for making a judgment (or forecast) in empirical studies." Qiu, L., Pang, J., & Lim, K. H. (2012).

that is common in S&OP settings. Using controlled experimentation with realistic simulations of the supply chain forecaster's task environment when sales promotions are imminent, we investigate whether the deleterious effects on accuracy predicted by the system neglect hypothesis (Massey and Wu 2005) and the literature on base-rate neglect (e.g. Kahneman and Tversky 1973) are applicable to promotion forecasting.

The paper is divided into five further sections. Following a review of the relevant literature in Section 2, we set out our hypotheses. Section 3 describes our methodology, the participants and the experimental setting. The fourth and fifth sections contain the detailed results of two experiments and associated discussion. Finally, Section 6 summarizes our conclusions as well as providing suggestions for further work and implications for practice.

2. The use of information in promotion forecasting

2.1 Base rate distractors

Previous work has acknowledged the importance of an expanded information set in order to enhance supply chain forecasting performance. Such information may come from internal sources such as marketing and operations (Fildes and Hastings 1994) as well as from information shared by other supply chain partners (e.g., Eksoz et al. 2014, Önkal and Aktas 2011). Such an extended information set has been shown to be valuable in enhancing accuracy with consequential stock service level improvements (Cui et al. 2015, Trapero et al. 2013). However, there appears to be surprisingly little empirical work on the actual *use* of information in supply chain forecasting, particularly in the important case of sales promotions. When determining the extent to which a baseline statistical forecast should be adjusted to take into account the effect of a forthcoming sales promotion, forecasters will usually have access to a diverse range of information, both quantitative and qualitative. In addition to a base rate figure that shows the average sales uplift achieved by past promotion campaigns, this will typically including historic demand data, past and current baseline statistical forecasts, quantitative information on the most recent sales promotion as well as qualitative information on factors judged to be relevant to the success of the forthcoming promotion. A potential problem is that some of this information, which may superficially appear to be relevant, will have little or no predictive value. For example, while a base rate based on a large sample of previous promotions is likely to provide a reliable estimate of the typical sales uplift, the uplift achieved in the most recent promotion is a sample of just one observation. Yet there is evidence that it is mostly this latest uplift that gets used in promotion forecasting. For example 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 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).

A second problem is that some information, such as an announcement that a particular celebrity has been recruited to head the promotion campaign, *may* be relevant, but in the absence of relevant data, its likely effect on sales uplift remains unknown. In these and similar circumstances, the safest strategy will be to adjust the statistical baseline forecast by estimating an uplift equivalent to the base rate. After all, some of the promotions that were used to estimate the base rate may also have employed celebrities (or had other characteristics that are similar to the forthcoming promotion), but such information may not be immediately accessible. Deviating from the base rate in such cases implicitly relies on the unsupported assumption that the celebrity effect (or the effect of any similar characteristic of the promotion that is being forecast) is not already embedded in the base rate.

These problems raise the question of the extent to which forecasters are likely to be sidetracked from using the relevant base rate. The literature suggests two main reasons why forecasters may be distracted. In the case where a salient item of quantitative information is available, such as the uplift achieved in a previous promotion, this may eclipse the base rate in accordance with the system neglect hypothesis proposed by Massey and Wu (2005). Massey and Wu found that even when people were informed of the parameters underlying the generation of observations in a system, they neglected these and formed judgments based on recent, and hence salient, observations which were subject to noise. Kremer et al. (2011) have shown that the same principles apply when people are asked to make judgmental time series forecasts. This tendency to focus on recent observations –particularly *the* most recent observation - rather than underlying nature of the system that produced these observations, has been found in several other judgmental forecasting studies (e.g. Andreassen and Kraus 1990, Lawrence and O'Connor 1992, 1995, Bolger and Harvey 1993). The system neglect hypothesis predicts that people will underreact to a changed environment because they underweight the underlying parameters that indicate change and hence fail to realise that a change is occurring. However, there is another reason why focusing on a single most recent promotion at the expense of the base rate will tend to lead to underestimation of a forthcoming promotion effect. 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. When the tasks requires an adjustment to the statistical forecast it is possible that this will act 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 and an estimate that the person would have made had they not seen the anchor.

In the context of adjusting a statistical baseline forecast for a promotion, this might be reflected by people taking a weighted average of the baseline statistical forecast and the previous promotion. The effect of this would be a forecast that fell between the two values leading to an underestimate of the promotion effect. Ironically the use of a weighted average would imply that higher statistical forecasts lead to smaller percentage adjustments because in this case the statistical forecast will be closer to the uplifted sales achieved in the previous promotion.

The tendency to use the previous promotion effect in the adjustment to the statistical forecast, thereby neglecting the base rate, is likely to depend on the salience of this effect. Goodwin and Fildes (1999) found that previous promotion effects had no influence on judgmental time series forecasts for promotion periods when the time series was highly noisy. Under these conditions, the previous effects were submerged in the large random movements of the series and hence were not salient. In contrast, given the prevalence of a recency bias in judgmental forecasting, it seems likely that the proximity the latest promotion to the current period will increase its salience and hence amplify its influence on the forecast.

The second reason why people might neglect the base-rate uplift is the presence of qualitative information in the form of reasons why the promotion may or may not be a success. Although there may be no evidence to establish the diagnosticity of these reasons, information in a narrative form is likely to act as a powerful distractor from the base rate (Önkal, Sayım, and Gönül 2013). For example, in a classic study of judgmental decision making, Tversky and Kahneman (1974) showed that information on statistical base rates is often neglected or discounted and unreliable narrative information is preferred. Kahneman and Lovallo's (1993) notion of the 'inside view' would also suggest that the availability of reasons will cause attention to be devoted to the specific characteristics of the particular promotion that is being forecast so that the focus on average sales uplifts (the 'outside view') will be lost.

When a forecaster has access to multiple items of qualitative information some of which are positive - suggesting that the promotion will be a relative success - and some of which are negative there is some evidence to suggest that the negative information may be more potent (Rozin and Royzman 2001). This is consistent with prospect theory which assumes that people have an aversion to losses (Kahneman and Tversky, 1979) and hence may be more vigilant in their response to negative information indicating a potential loss than they would be to positive information. The reasons for the greater influence of negative information are complex but researchers such as Peeters and Czapinski, (1990) have suggested that in the environment negative events are rarer but can have more important implications for survival, so it pays to be especially watchful for the dangerous negative event. In our context this negativity bias implies that in situations where positive and negative reasons are equally likely to be present and where their diagnosticity is unknown, there will be a tendency to under estimate future promotion effects.

In summary, in the S&OP forecasting process where information relating to a forthcoming promotion is provided to a demand forecaster in quantitative, graphical and qualitative forms, information with zero or unknown diagnosticity is likely to distract them from the normative base rate for promotional events.

2.2 Moderating factors

In practical contexts a number of other factors may moderate the extent to which both the latest promotion and qualitative information arising from S&OP discussions (Oliva and Watson 2009) lead to a tendency to under forecast future promotion effects. Before commencing the forecasting task, forecasters may have a prior view of the likely impact of promotions based on their recall (which may be imperfect) of earlier promotions (Reimers and Harvey 2011), or on industry beliefs. This may also serve to reduce the weight that is attached to the base rate. Secondly, forecasters in organizations may be subject to motivating factors that cause them to consciously or unconsciously bias their forecasts. The motivation of the forecaster may also 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'), (e.g., Windschitl et al. 2010). As the Oliva and Watson (2009) case study shows such a bias is a common feature of the S&OP forecasting process.

Despite these potential biases, forecasters in many organizations 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). Moreover, prestige and reputation concerns and the knowledge that one's forecast will be evaluated may lead to a 'reality constraint' so that factors favoring optimism bias, for example, may be tempered (Lerner and Tetlock 1999).

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. For example, Franses (2014) found that more experienced forecasters in a pharmaceutical company produced more accurate adjustments.

Finally, the most recent sales figure (as opposed to sales in the most recent promotion period) and most recent forecast error may have an influence on the size of adjustment made for the forthcoming promotion. For example, an additional upwards adjustment might be made to reflect a relatively high 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). If this observation is well above the forecast for that period –leading to a large positive forecast error – its salience, and hence its influence on the adjustment is likely to be enhanced.

The above discussion suggests the following hypotheses:

H1: Adjustments to statistical baseline forecasts to take into account forthcoming promotion effects will deviate from base rates when information with no, or unknown, diagnosticity is provided

H2: Adjustments made to statistical baseline forecasts to take into account the effects of forthcoming promotions will tend to underestimate these effects when the uplift obtained by a previous promotion is displayed on a time series graph.

H3: Adjustments made to statistical baseline forecasts to take into account the effects of forthcoming promotions will tend to underestimate these effects when positive and negative qualitative information, with unknown diagnosticity is equally likely to be presented

In summary, little is known about the way forecasters use information to produce their judgmental adjustments of statistical baseline forecasts when products are due to be promoted. Yet it is an important issue in that inaccurate forecasts can be costly in terms of surplus inventory or the loss of customer goodwill and sales. It is also important theoretically in that little research has examined the interpretation of information in a time series context. In the remainder of this paper, we investigate whether information of different types distracts forecasters from using base rate information and whether this leads to a tendency to under forecast promotion effects.

3. Methodology and design of experiment 1

In order to test the hypothesis developed above, while controlling for prior expectations of promotion effects, self-reported knowledge of forecasting and different types of motivation we have adopted a behavioural experimental approach. Controlled laboratory experiments are increasingly used to investigate demand forecasting behavior and related biases (Kremer et al 2011, Moritz et al. 2014, Siemsen 2011), as they allow for systematized examinations of crucial factors affecting forecasting performance. They are also now common in the operations literature (Gans and Croson 2008, Croson et al. 2013, Zhao et al. 2013).

We report in detail on one experiment which built on the experience gained from a number of preliminary experiments. The participants in the first experiment were management students, studying for either bachelors, masters or doctoral degrees at the Universities of Bath (UK), Bilkent (Turkey) and Lancaster (UK) They had all studied some forecasting. While they do not have the same experience as commercial forecasters, they have at least as much statistical training as many practicing forecasters. Evidence provided by earlier studies, and recently by Kremer et al (forthcoming), strongly suggest there are few differences between the two groups in a context such as that simulated here. (We pick up on this issue in the second experiment we report on which involved executives.)

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 briefing describing the task and base rate information, on the average percentage uplift in sales achieved by a promotions at this supermarket. The 50% average uplift was highlighted both in the cover story and in information presented on the computer screen during the trial run.

Once the experiment started other information was provided through an FSS (see Figure 1 for a typical screenshot) 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. The realism of both the system and the participants' task were intended to increase the ecological validity of our findings (Rogers and Soopramanien 2009).

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:

$$
\begin{aligned}
&Sales_t = BaseLineForecast_t + Promotional\ \ Effect_t + \varepsilon_t \\
&BaseLineForecast_t = 0.2 * Sales_{t-1} + 0.8 * BaselineForecast_{t-1} + (perturbation\ for\ \ t = 25) \\
&Promotional\ \ Effect \sim Uniform(80,120) \\
&\varepsilon_t \sim Normal(0, stddev)
\end{aligned}
$$

The initial BaselineForecast was set to 200. The standard deviation (*Stddev*) 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 as shown, apart from a random perturbation in period 25. 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 promoted periods, *t*, the previous baseline forecast was not updated. The single promotion in the historical data was generated uniformly for an integer period between 1 and 24. The timing and effect of this promotion varied across SKUs but the mean sales uplift (relative to the baseline forecast) was 50%

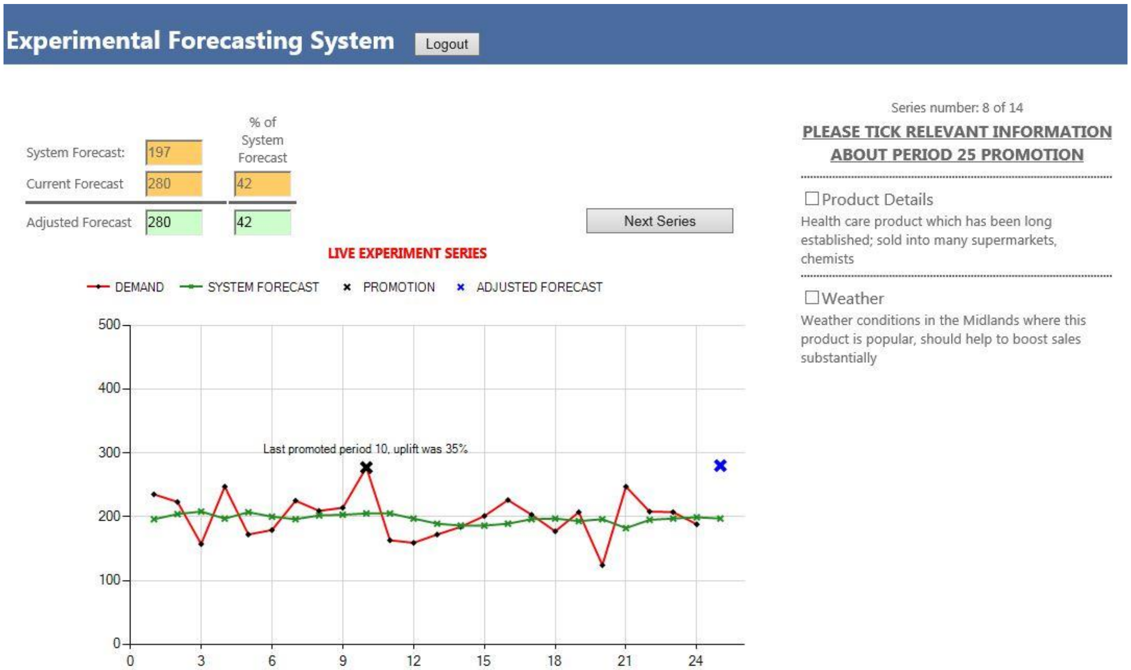


Figure 1 Screenshot of experimental forecasting support system

In addition to the historical sales series and statistical forecast . the screen displayed, for each SKU, 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. Having been presented with all of this information the participants were invited to use their judgment to adjust the baseline forecast for each SKU to take into account the forthcoming promotion for that product.

In order to control for the possible moderating effect of motivation 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. The best two forecasters in each treatment received an Amazon voucher in the first and second groups; a prize draw was used to select the two winners in the control group. 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 their prior view of what 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 familiarize 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 –though no empirical evidence was produced to support the assessment. They also received overall feedback on their accuracy after forecasts had been made for both trail-run SKUs. No feedback was provided in the main part of the experiment. 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.

Because of the complexities of designing experiments that provided a realistic simulation of the supply-chain forecaster's task, a number of preliminary experiments were run, involving over 200 participants. These enabled us to fine tune the design and screen display, to eliminate potential confounded factors and to identify the key issues that merited further investigation. These experiments included different numbers of series, fixed numbers of reasons, forcing participants to select a primary and secondary reason to support their adjustment, having an average promotional uplift of 80% and including a trend in the data. The results of these experiments were consistent with those that we discuss next suggesting that our results are robust. For brevity, these earlier results will not be reported here but they are available from the authors.

4. Analysis and results of Experiment 1

126 participants completed the experiment. We then 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 summarized in Table 1.

Table 1 Questionnaire responses

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

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

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.

The participant's median prior estimate of the percentage uplift achieved in supermarket promotions was 50%. However, their median estimated uplift during the experiment was only 30%, significantly lower ($p<.001$) than the base rate of 50%, a. result consistent with a neglect of the base rate. This provides support for both H1 and H2. The distribution of these percentage adjustments was broadly normal with a few positive outliers. Only 25% of the adjustments were greater than the base rate of 50%. However, some were as high as 200%, which is quite reasonable for the sorts of products we have included in our experimental design.

4.1 Statistical modelling

Statistical modelling was used to identify the cues that were determining the sizes of adjustments made by participants and ,in particular, whether the previous promotion effect and reasons were distracting them from the base rate. The nature of the experiment, where each respondent is sequentially given a number of series in random order, together with random information cues, requires a more sophisticated analysis than a standard ANOVA or regression. Individual participants can be expected to have a random response to the series and cues. The advantages of using linear mixed effects models for this situation (Verbeke and Molenberghs 2000) have been summarized as "they allow the researcher to simultaneously consider all factors that potentially contribute to the understanding of the structure of the data….including standard fixed effects ….. and covariates" compared to standard approaches (Baayen et al. 2008)

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

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 $n_i \times p$ and $n_i \times q$ 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. SAS 9.3 has been used to estimate the equations using restricted maximum likelihood.

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

- The dependent variable is the adjustment percentage transformed into log (1+Proportional_uplift) to ensure better error distributional characteristics (Davydenko and Fildes 2013).
- 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 percentage error (measured as log(Forecast/Actual)), 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 noise variance was treated as a fixed effects class variable.
- Participants' responses to the information cues were treated as random effects specific to the individuals.
- 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 characterize 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. In addition, various interactions were included but did not add any explanatory power. Any non-linear effect of timing was also checked but a linear model proved adequate. A sensitivity check on the assumption of the correlation structure of the repeated measures did not show any substantive differences.

4.2 Results of modelling

The results from the model are shown in Table 2 which excludes observations of high leverage. However, as a check on the robustness of our findings, the results from estimating the model with the full set of observations remained broadly the same (1560 observations were reduced to 1309 after excluding high leverage points and non-compliant responders). The parameter coefficients are interpreted as percentage effects so for example, four negative reasons (Reasncat=-4) lowers the average adjustment by 10.1% (=100[1-exp(-0.1067)]). It can be seen that both the previous promotion uplift and the reasons were significantly associated with the adjustments made by the participants, consistent with H1. Higher uplifts in the previous promotion were associated with higher adjustments. This effect was slightly greater the more recent the promotion. As expected, lower levels of noise were also associated with higher estimated uplifts suggesting that high noise was making the effects of the previous promotion less salient. The significant negative coefficient for the statistical baseline forecast is consistent with participants placing their estimate of the uplifted sales between the baseline forecast and the previous promotion. This would account for the tendency to underestimate the expected uplift of 50% and is consistent with H2.

Table 2 Model of the adjustment ($\log_e$(1+Proportional_uplift)

Effect	Estimate	p-value†
Intercept	0.505	<.0001
ln(last promotion uplift)	0.275	<.0001
ln (last actual)	0.037	0.001
ln (last stats forecast error)	0.040	0.105
ln(current stats forecast)	-0.128	<.0001
ln(Prior)	0.035	0.014
Low Noise	0.021	0.018
Timing of past promotion	0.001	0.036
Reasncat = -4	-0.107	<.0001
Reasncat =-3	-0.136	<.0001
Reasncat = -2	-0.114	<.0001
Reasncat = -1	-0.077	<.0001
Reasncat = 0	-0.078	<.0001
Reasncat = 1	-0.052	0.001

Reasncat = 2	-0.023	0.112
Reasncat = 3	-0.027	0.100

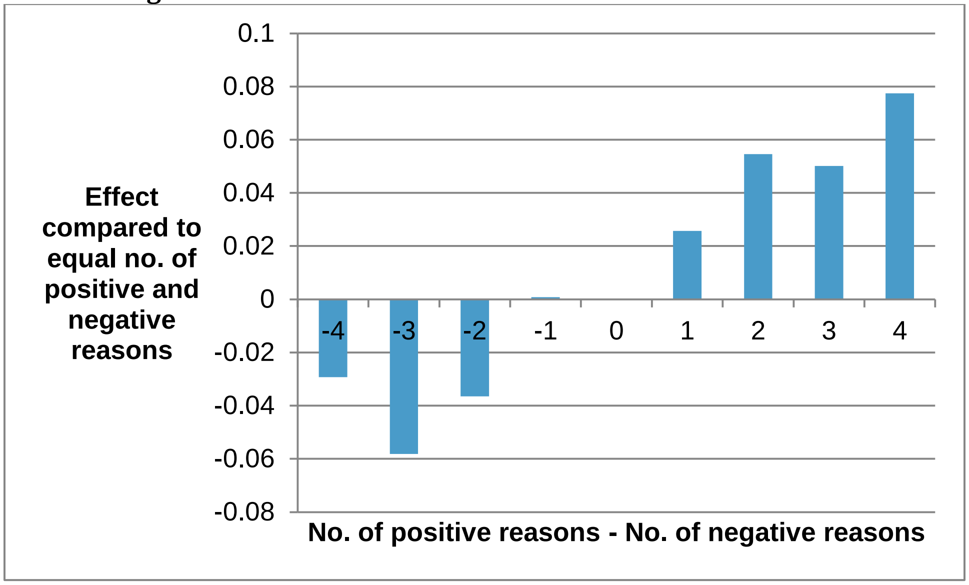
[Available n= 1560; sample size after deleting high leverage points=1309]
[*Reasncat* = No. of positive reasons supplied – No. of negative reasons]
† All tests are one-sided apart from that for low noise (and intercept).

Figure 2 shows the relationship between $\log_e$(1+Proportional_uplift) 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. 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. Broadly speaking, the greater the balance of reasons in one direction then the greater was the distraction from the base rate estimate, despite the unknown diagnosticity of these reasons which is consistent with H1.

Did negative reasons have a greater influence than positive reasons? An analysis of ‘contrasts’ showed that having one more positive reason is more impactful than one more negative reason, 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 on the adjustment which is contrary to H3. This suggests that the propensity to underestimate the promotion effects resulted from the tendency to place the adjusted forecast between the baseline forecast and the previous promotion, rather than greater weight being attached to negative reasons.

We investigated whether a number of variables had a moderating effect on the results. Table 2 shows that participants who came to the experiment with higher prior expectations of promotion uplifts tended to make larger upwards adjustments to the baseline forecasts. However, the carry-over effect to their individual SKU adjustments was small. There was also 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

Figure 2: Effects on $\log_e$(1+Proportional_uplift) of differences between the number of positive and negative reasons relative to situations where there are an equal number of positive and negative reasons



the participants were familiar with. In addition, Table 2 shows that adjustments tended to be higher when the most recent actual sales figure was higher (this was always a non-promotion period). As discussed earlier, a high recent sales figure might be interpreted as a signal that the underlying level of sales has increased so that a greater adjustment to the statistical forecast is needed.

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. Telling participants that they would be rewarded if the uplift exceed 50% (in an attempt to induce desirability bias) did not reduce this underestimation, nor did rewarding accuracy.

4.3. Discussion of Experiment 1

Overall, the results of experiment 1 suggest that when making their forecasts, participants were distracted from the 50% base rate by the previous promotion uplift and the reasons, together with a number of moderating factors, despite this information having either no or unknown diagnosticity. In particular, they appeared to set their adjustment between the baseline statistical forecast and the previous promotion resulting in a tendency to underforecast the forthcoming promotion effect. These findings are consistent with the system neglect hypothesis and earlier findings on base rate neglect which suggest that base rates are eclipsed by more salient information despite its low predictive value.

None of our motivation treatments had a significant effect on the size of adjustments: the treatments 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 perhaps 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 prestige and resources to one's department by producing reliable forecasts.

5. Experiment 2

In Experiment 1 participants had access to both potential distractors from the base rate, namely the previous promotion uplift and the reasons. This did not allow the effects of information with zero diagnosticity and information with unknown diagnosticity to be examined separately. Experiment 2 had a simpler design. Participants were randomly assigned to one of two treatments. In the previous promotion treatment, the series contained sales obtained in a previous promotion but no reasons relating to the forthcoming promotion were displayed. In the 'reasons' treatment reasons for the success or otherwise of the forthcoming promotion were displayed but no previous promotion effects appeared in the time series. In this case, for each SKU, either two negative, two positive or zero reasons were displayed with the number of reasons being selected at random. The design was based on the results of the previous experiment where the influential variable proved to be the difference between the number of positive reasons offered and the number of negative reasons. The effects had proved approximately linear between -2 and 2 reasons. In the 'reasons' treatment the promotion always appeared in period 18 eliminating any possible timing effects. Unlike Experiment 1 there was no random perturbation of the statistical baseline forecast and no motivation treatments were included. Each participant made forecasts for period 25 for 14 series, including two trial series. In all other respects the experiment was the same as Experiment 1.

The participants were 30 executives undertaking an Executive MBA module on forecasting so the experiment also enabled us to test whether the effects observed in Experiment 1 are also valid for experienced executives.

5.1 Results of Experiment 2

While the two sets of cues, past promotions and reasons, can be embedded in a single analytical model, analysis shows there are interaction effects that annul any gains in efficiency in the estimation of coefficients. Hence separate mixed linear effects models, were estimated for the adjustments made by participants in the two treatments. The models had the same underlying structure as the one used to analyze Experiment 1, except that the number of positive minus the number negative reasons was represented by a single variable, rather than a series of dummy variables. This was because, as indicated above, the association between the number of positive minus the number of negatives and the estimated uplift was approximately linear for the range -2 to +2 reasons in experiment 1. As before the dependent variable was $\log_e(1 + \text{Proportional\_uplift})$ and high leverage points were removed. This time ln(last actual) and ln(last stats forecast error) were not included in the list of independent variables because the lack of random perturbation in the statistical forecast meant that they would be collinear with that forecast.

The results from the group of executives taking part in this experiment generally support those reported earlier. As shown in Table 3, for the 'previous promotion' group, the previous promotion had a highly significant effect on the estimated uplift for the forthcoming promotion. Also, as in Experiment 1, a lower level of noise led to higher estimated uplifts consistent with the notion that the effect of the previous promotion was less salient under conditions of high noise. However,, unlike Experiment 1 a higher statistical forecast was associated with a larger upwards adjustment, probably do to multicollinearity with the last actual (which was broken experiment 1). The effect of these factors was the effect that the overall adjustment had a (trimmed) mean of 43%, and a median of 40%, (insignificantly different from 50% but substantially and significantly less than the observed means of the past promotions, 57.5%).

Table 3 shows that participants in the ‘reasons’ group were significantly influenced by the provided reasons when they estimated the promotion uplift. The greater the balance in favor of positive reasons then the greater was the upwards adjustment they tended to make. On the occasion when no reasons were provided their mean estimate was 49% (close to the base rate), although the median estimate was only 34%. As with the ‘previous promotion’ group, higher statistical forecasts were associated with larger upwards adjustments as there was no previous promotion to encourage a pattern matching strategy. But, as expected, the level of noise did not have a significant effect as there was no previous promotion uplift that was potentially submerged in high noise. Also, while the participants’ prior estimates had a significant effect for the ‘previous promotion’ group, the effect was not significant when reasons were supplied. This might suggest that reasons are more powerful than data on a previous promotion in causing people to revise their prior estimates.

Table 3 p-values for coefficients in linear mixed effects model for the two treatments

	Group1: Previous promotion		Group2: Reasons	
Effect	**Estimate (n=162)**	**p-value**	**Estimate (n=162)**	**p-value**
Intercept	-4.830	<0.001	-2.946	0.001
ln(last promotion uplift)	0.678	<0.001	n/a	n/a
ln(current stats forecast)	0.889	0.001	0.598	0.003
ln(Prior)	0.418	0.153	1.160	0.133
Low Noise	0.190	0.001	-0.022	0.608
No. of pos. minus no. of neg. reasons	n/a	n/a	0.032	0.0416

All tests are one-sided except for that for low noise (and intercept)

In conclusion, Experiment 2 has clearly demonstrated that, when they were presented separately, both past promotions with zero diagnosticity and qualitative information with unknown diagnosticity adversely affected estimates of promotional adjustments (compared to a normative adjustment of 50%. In addition the effects observed earlier with a diverse group of business students have been replicated with experienced executives.

6. General Discussion and Conclusions

The efficient use of information by demand forecasters can be crucial given the negative effects of forecast errors on production, distribution and inventory planning. For example, Kremer et al. (forthcoming) estimate that a percentage improvement in accuracy translates into a similar percentage reduction in safety stock. Given their supply chain repercussions, promotions pose particularly sharp challenges to S&OP decision-makers. The results of this experiment-based study suggest that the provision of information relating to promotions can be detrimental to forecast accuracy when, despite its salience, it has either no or unknown diagnosticity. This finding has important implications both for the design of forecasting support systems that are commonly used in supply-chain-based organizations and for the extent to which supply chains can operate efficiently. The systems typically place an emphasis on the provision of information to the forecaster in an amenable and accessible format, irrespective of its predictive value. Our results suggest that these 'passive' systems may be inimical to accuracy. The participants misinterpreted the time series history: they chose not to accept the accuracy of the statistical baseline forecast, using this and past observations to reweight the forecast, introducing additional forecast errors They also appeared to adopt a version of the 'last-lift' heuristic, the most common promotional forecasting method used in practice. Their mistake was to ignore the average uplift, instead focusing on the last observed value.

Both facets of the participants' sub-optimal forecasting suggest that FSSs need to be redesigned. Systems that actively evaluate and filter information before presenting it may lead to improved accuracy. Parikh et al. (2001) found that the provision in an FSS of informative guidance, which they defined as the provision of unbiased, relevant information without a specific suggestion, was superior in promoting learning than systems that suggested how information should be used. However, the emphasis needs to be on the provision of relevant and salient information. For example, in promotion forecasting, a system that identifies analogous past promotions and provide estimates of their average effect has been found to improve forecast accuracy in a laboratory study (Lee et al. 2007). But as Dietvorst et al. (2015) and Lim and O'Connor (1995) have shown, changing the habit of mis-weighting remains difficult. Such intentional and unintentional misuse of information and the prevalence of habitual (mis) weighting schemes further support the call for effective redesign of FSSs to aid predictive performance.

Is the underestimation of promotion effects we found in our laboratory experiments typical of what happens in the field? The evidence is sparse. Our study suggests the bias results from the poor design of FSSs but no field study has given details of the characteristics of the FSS that was used in promotion forecasting and the role it played, if one was used at all. Some field studies have reported that judgmental adjustments tend to suffer from optimism bias (Fildes et. al. 2009, Franses and Legerstee 2011), the opposite to what we found here. However, neither of these studies confined their analysis specifically to forecast periods when a promotion was due to occur. Moreover, a study of the sales forecasts of German companies, again not limited to promotions and not limited to adjustments of statistical forecasts, found a pessimism bias (Muller 2011). 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. Further work on forecasting of promotion events is clearly needed to disentangle the confounding factors affecting predictive performance in supply chains.

Like most experimental studies, this work has limitations. One issue is whether the participants felt obliged to deviate from the 50% base rate, otherwise why were they being invited to take part in the experiment. Simply entering a 50% uplift for every SKU may have seemed too easy or may have been perceived as signaling disengagement with the forecasting task. Yet, in this respect the experiment was probably an accurate reflection of the field. For example, Fildes et al. (2006) found that forecasters in companies tended to make lots of small gratuitous adjustments to statistical forecasts apparently simply to justify their role. The participation of students in Experiment 1 may be regarded as another limitation despite their motivation and knowledge of forecasting. However, this is unlikely to affect the substantive conclusions, as others, including Experiment 2 here and Kremer et al. (forthcoming) illustrate, also in the context of demand forecasting. 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. The results may also depend on the features of the base line statistical model, where the smoothing parameter is known to affect responses (Kremer et al. 2011). Future research could usefully examine the behavior of forecasters when the statistical model captures some promotional drivers.

In summary, given the limitations of current forecasting systems (Fildes et al. 2006), there appears to be substantial scope for design innovations. These may include structured support on filtering and integrating qualitative and quantitative information, targeted to individual forecasters, as well as support on design of collaborative forecasting systems that reach across different supply chain partners operating under diverse information platforms. Further work on such innovative designs promises to enhance communication between forecasters and decision makers with extensive impact on overall supply chain performance.

References:

Andreassen, P. B., S.J. Kraus. 1990. Judgmental extrapolation and the salience of change. *Journal of Forecasting*, **9**: 347-372.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Bolger, F., & Harvey, N. 1993. Context-sensitive heuristics in statistical reasoning. *The Quarterly Journal of Experimental Psychology*, **46**: 779-811.

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

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

Cui, R., G. Allon, A., Bassamboo, J.A.V. Mieghem. 2015. Information Sharing in Supply Chains: An Empirical and Theoretical Valuation. *Management Science*. **61**, 2803-2824.

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

Davydenko, A., R. Fildes, 2013. Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*. **29**: 510-522.

Dietvorst, B. J., J.P. Simmons, C. Massey. 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.

Eroglu, C., K.L Croxton. 2010. Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*. **26**: 116-133.

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

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

Fildes, R., P. Goodwin, M. Lawrence,K. Nikolopoulos. 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., R Hastings. 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 University Press, Cambridge.

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. 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.

Gans, N., & Croson, R. 2008. Introduction to the special issue on behavioral operations. *Manufacturing & Service Operations Management*. **10**(4): 563-565.

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., W-Y Lee, R. Fildes, K. Nikolopoulos, M. Lawrence. 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.

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

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

Kahneman, D., D. Lovallo. 1993. Timid Choices and Bold Forecasts - a Cognitive Perspective on Risk-Taking. *Management Science*. **39**: 17-31.

Kahneman, D.; Tversky, A. 1973. On the psychology of prediction. *Psychological Review* **80**: 237–251.

Kahneman, D., Tversky, A. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica*, **47:** 263

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

Kremer, M., B. Moritz., E. Siemsen. 2011. Demand forecasting behavior: system neglect and change detection. *Management Science*. **57**: 1827-1843.

Kremer, M., E. Siemsen, D.J. Thomas. forthcoming. The sum and its parts: judgmental hierarchical forecasting. *Management Science*.

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

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

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.

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

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

Lerner, J. S., P.E. Tetlock. 1999. Accounting for the effects of accountability. *Psychological Bulletin*. **125(**2): 255-275.

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

Massey, C., Wu, G. 2005. Detecting regime shifts: The causes of under-and overreaction. *Management Science*, **51**: 932-947.

Moritz, B., E. Siemsen, M. Kremer, M. 2014. Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management*. **23**: 1146–1160.

Oliva, R., N. Watson. 2009. Managing Functional Biases in Organizational Forecasts: A Case Study of Consensus Forecasting in Supply Chain Planning. *Production and Operations Management*. **18**: 138-151.

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.

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

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

Özer, Ö., Raz, G. 2011. Supply chain sourcing under asymmetric information
Production and Operations Management, **20***:* 92 – 115.

Özer, Ö., Zheng, Y., Chen, K.-Y. 2011. Trust in forecast information sharing, *Management Science,* **57***:* 1111 – 1137.

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

Peeters, G., & Czapinski, J. (1990). Positive–negative asymmetry inevaluations: The distinction between affective and informational effects. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 1, pp. 33, ds. New York: Wiley.

Qiu, L., Pang, J., & Lim, K. H. 2012. Effects of conflicting aggregated rating on eWOM review credibility and diagnosticity: The moderating role of review valence. *Decision Support Systems*, **54**: 631-643.

Parikh, M., Fazlollahi, B., & Verma, S. 2001. The effectiveness of decisional guidance: an empirical evaluation. *Decision Sciences*, **32**: 303-332.

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

Rogers, G., D. Soopramanien,. 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., E.B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*. **5**: 296-320.

Seifert, M., E. Siemsen, A.L. Hadida, A.B. Eisingerich. 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*. **473**: 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.

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

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

Tversky, A., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science*. **85**: 1124-1131.

Verbeke, G., G. Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Webby, R., M. O'Connor, B. Edmundson. 2005. Forecasting support systems for the incorporation of event information: An empirical investigation. *International Journal of Forecasting*. **21**: 411-423.

Windschitl, P.D., A.R. Smith, J.P. Rose, Z. Krizan. 2010. The desirability bias in predictions: Going optimistic without leaving realism, *Organizational Behavior and Human Decision Processes*. **111**: 33-47.

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