Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting from time series subject to sporadic perturbations: Effectiveness of different types of forecasting support

Shari De Baets [a,b], Nigel Harvey [c,*]

[a] *Ghent University, Belgium*
[b] *Vlerick Business School, Belgium*
[c] *University College London, United Kingdom*

A B S T R A C T

How effective are different approaches for the provision of forecasting support? Forecasts may be either unaided or made with the help of statistical forecasts. In practice, the latter are often crude forecasts that do not take sporadic perturbations into account. Most research considers forecasts based on series that have been cleansed of perturbation effects. This paper considers an experiment in which people made forecasts from time series that were disturbed by promotions. In all conditions, under-forecasting occurred during promotional periods and over-forecasting during normal ones. The relative sizes of these effects depended on the proportions of periods in the data series that contained promotions. The statistical forecasts improved the forecasting accuracy, not because they reduced these biases, but because they decreased the random error (scatter). The performance improvement did not depend on whether the forecasts were based on cleansed series. Thus, the effort invested in producing cleansed time series from which to forecast may not be warranted: companies may benefit from giving their forecasters even crude statistical forecasts. In a second experiment, forecasters received optimal statistical forecasts that took the effects of promotions into account fully. This increased the accuracy because the biases were almost eliminated and the random error was reduced by 20%. Thus, the additional effort required to produce forecasts that take promotional effects into account is worthwhile.

1. Introduction

Business forecasters use both unaided judgmental forecasting and forecasting aided by formal statistical forecasts (Sanders & Manrodt, 2003). The latter approach may become increasingly common as users become more familiar with the sorts of software that provide forecasting support. As a result, forecast support systems have great potential for improving forecast performances. However, there are various factors that prevent this potential from being realised fully. Forecasters tend to ignore the 'advice' provided by a formal forecast, or take it into account too little (Goodwin, Fildes, Lawrence, & Nikolopoulos, 2007; Lim & O'Connor, 1996; Önkal, Goodwin, Thomson, Gönul, & Pollock, 2009). That is, even when they do take it into account, they do not assign enough weight to it. Consequently, the improvement in accuracy that it produces is generally small, albeit somewhat greater when the series are complex and the formal forecasts are of a higher quality

* Correspondence to: University College London, Department of Experimental Psychology, Gower Street, London WC1E 6BT, United Kingdom.

E-mail address: n.harvey@ucl.ac.uk (N. Harvey).

(Goodwin & Fildes, 1999; Goodwin, Fildes, Lawrence, & Stephens, 2011; Lim & O'Connor, 1995; Trapero, Pedregal, Fildes, & Kourentzes, 2013).

The picture is more complex in the case of series with sporadic perturbations, such as those associated with promotions. Goodwin and Fildes (1999) showed that, in this situation, statistical forecasts tend to be helpful in normal periods, but not in those that are subject to promotions. However, the statistical forecasts they used did not take the effects of promotions into account, but were based on the baseline time series cleansed of the effects of promotions. Recently, forecasting models that do allow for the effects of promotions have been developed (Huang, Fildes, & Soopramanien, 2014; Kourentzes & Petropoulos, 2016; Trapero et al., 2013). However, given that there is a considerable lag between the development of more sophisticated statistical models and their implementation by practitioners (Lawrence, 2000; Sanders & Manrodt, 2003), it is likely to be some time before they have any impact on business practice.

Even in the case of relatively simple models, there appears to be a gap between the formal forecasts used in experimental studies and those used in business practice. In experimental studies, formal forecasts are based on non-promotional periods only (e.g., Goodwin & Fildes, 1999); in other words, they are calculated from the baseline series cleansed of promotion effects. In non-experimental studies, on the other hand, formal forecasts do not take into account whether past periods contain promotions (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Trapero et al., 2013). Hence, if we are interested in considering the relevance of experimental results to business practice, we need to ask whether the potential advantage of using judgmentally adjusted statistical forecasts rather than unaided judgment depends on the type of statistical forecast used.

Goodwin and Fildes (1999) argued that the benefit of providing statistical forecasts should be greater when they are based on data that have been cleansed of promotional effects. Referring to the estimated level of sales when a promotion does not run as the *baseline* value, they point out that this is because the baseline values provided by that type of statistical forecast can be accepted without any adjustment when no promotions are planned. Moreover, the past differences between promotional and non-promotional periods can be used directly as a basis for assessing the size of the adjustment that is needed when promotions are planned.

In what follows, we address the following questions. First, does the use of a judgmentally-adjusted statistical forecast provide an advantage over the use of unaided judgment? Second, is any such advantage greater when statistical forecasts are based on past data that have been cleansed of promotional effects? Third, does any benefit that may be derived from the provision of statistical forecasts depend on features of either the data series (i.e., the ratio of promotional to non-promotional periods) or the periods to be forecast (i.e., whether a promotion is planned)? Finally, can people make good use of 'ideal' statistical forecasts that make allowance for the effects of promotions (cf., Huang et al., 2014; Kourentzes and Petropoulos, 2016; Trapero et al., 2013)? In other words, if

their goal is to maximize the forecasting accuracy, do they adopt these forecasts without any adjustment?

2. Development of hypotheses

In their survey, Fildes and Goodwin (2007) found that 75% of respondents indicated that they used judgment when making forecasts, with 25% saying that they used unaided judgment and 50% saying that they used a combination of judgment and statistical forecasting (averaging, judgmental adjustment). Over recent years, the use of statistical software has become more pervasive in business settings, and therefore the proportion of forecasters using a combinatorial approach has increased: it had risen to 55% by 2014 (Fildes & Petropoulos, 2015).

Judgmental adjustment does not always improve statistical forecasts, as people tend to make *unnecessary* adjustments even when they have no additional information (Goodwin, 2000; Lawrence, Goodwin, O'Connor, & Önkal, 2006). This may be because they discern patterns in noise (Fildes et al., 2009), because they are too optimistic and place excess weight on positive signals (Bovi, 2009; Durand, 2003; Kotteman, Davis, & Remus, 1994), or because they want to feel ownership of their forecasts (Önkal & Gönul, 2005). They also tend to be overconfident in the accuracy of their forecasts (Arkes, 2001; Bovi, 2009; Lawrence et al., 2006), perhaps because a self-serving attribution bias causes them to overestimate the importance of their own judgment relative to that of the statistical forecast (Hilary & Hsu, 2011; Libby & Rennekamp, 2012).

All of these studies have focused on whether judgmentally-adjusted forecasts are better or worse than raw statistical forecasts. The underlying issue was whether forecasters should be allowed to make adjustments to statistical forecasts and, if they should, whether there is anything that can be done to ensure that their adjustments are beneficial (Goodwin et al., 2011). In contrast, our primary aim here is to investigate the value of providing a formal forecast in order to increase the forecasting accuracy. Thus, our main focus is on whether judgmentally-adjusted statistical forecasts are better or worse than unaided judgmental forecasts.[1] For us, the underlying aim is to quantify the benefit of providing forecasters with forecasting support (operationalized in this paper as the provision of a statistical forecast, including historic forecasts). Such support has been assumed to be beneficial (Alvarado-Valencia & Barrero, 2014) because it reduces the processing demands imposed on forecasters (Fildes & Goodwin, 2013). Furthermore, combining forecasts from more than one source outperforms the results of a single forecasting method (Armstrong, 2001), particularly when the two methods are independent and rely on different information. The complementary nature of judgment and statistical methods means that their combination should be especially beneficial (Blattberg & Hoch, 1990). Therefore:

---

[1] However, we will also report comparisons between judgmentally adjusted forecasts and raw statistical forecasts in Section 5.
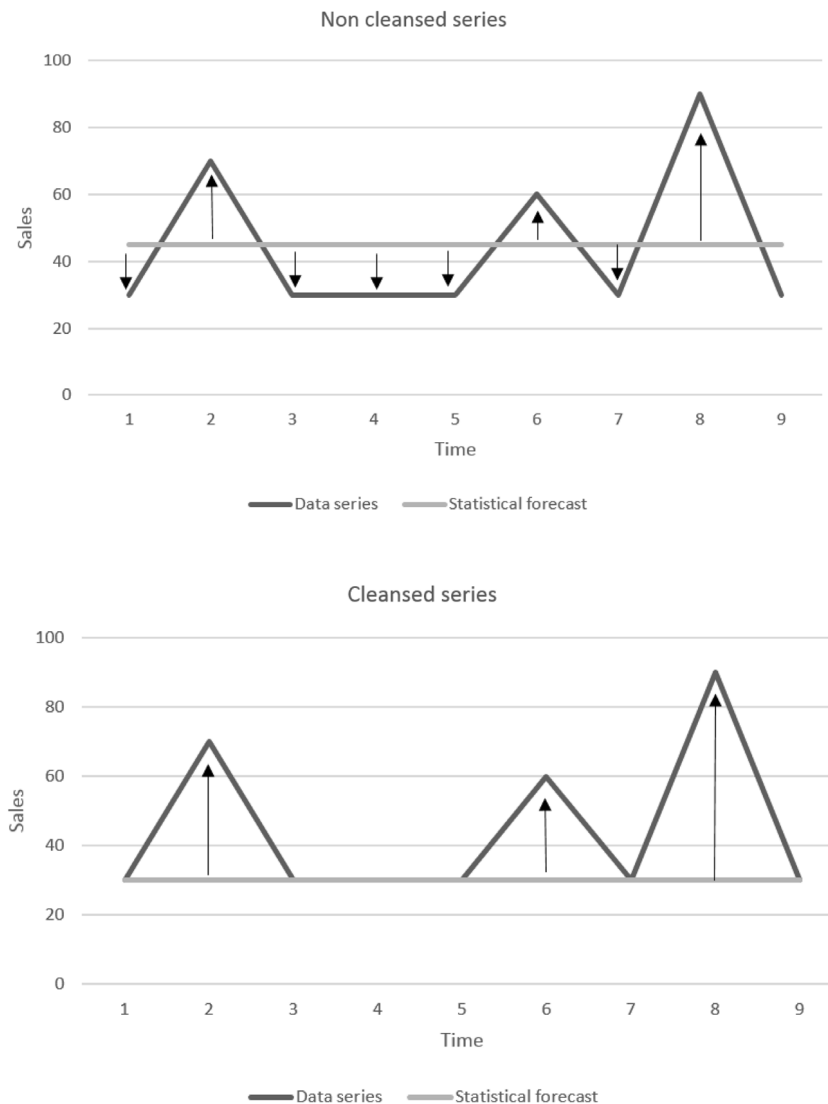
Fig. 1. Adjustments necessary for a statistical forecast based on non-cleansed series (upper panel) and cleansed series (lower panel).

Hypothesis 1. Providing forecasters with statistical forecasts improves the forecasting accuracy compared to unaided judgment.

Önkal, Sayim, and Lawrence (2012) noted that some differences exist between the forecast characteristics examined in experimental research and those that are prevalent in business practice. As was mentioned above, one such difference lies in the nature of the statistical forecasts that are provided when the series are subject to perturbations of the sort that are typically produced by promotions: the series used for producing statistical forecasts in experimental work have generally already been cleansed of promotional effects (Goodwin & Fildes, 1999; Goodwin et al., 2011), whereas the business data analysed by researchers generally have not been cleansed (Fildes et al., 2009; Trapero et al., 2013). As was mentioned above, Goodwin and Fildes (1999) expected the former approach to produce better results. Specifically, they argued that "This has the benefit of clearly separating the underlying time series from the promotion effects. Moreover, some commercial forecasting packages like *Forecast Pro* now allow observations for special periods to be separated out so that they cannot contaminate forecasts for normal periods. ...With access to a statistical time series forecast of the 'baseline value' the judge has only to estimate the effect of the cue and make an appropriate adjustment to the statistical forecast" (p. 41).

As an example, consider a promotion of a given size that has previously elevated sales by 100 units above the baseline. If a promotion of the same size is planned for the future, one could simply add 100 units to the statistical forecast of the baseline (Fig. 1, lower panel). On the other hand, if no promotion is planned, the baseline forecast can be adopted without adjustment. In contrast, statistical forecasts based on non-cleansed data always need to be adjusted: downwards when no promotion is planned, and upwards when one is planned (Fig. 1, upper panel).

However, forecasters need to know how much the statistical forecast has been influenced by the presence of past promotions in the data series. Without that knowledge, it is difficult for them to know how much to adjust upwards when promotions are planned and downwards when they are not. Thus, the forecasting process is more complex than when the statistical forecast is based on cleansed data series.

In fact, few studies have compared the effects of different types of statistical forecasts on the accuracy of judgmental forecasters who are provided with those forecasts. One exception is Lim and O'Connor's (1995) experimental study of forecasting from time series without disturbances. They manipulated the accuracy of the statistical forecasts, varying from low (naïve forecast) to medium (forecast produced by damped exponential smoothing) to high (average of the actual value and the forecast produced by damped exponential smoothing), then asked the participants to make an initial forecast based on their own judgment and one of the three types of statistical forecast. After each trial, the participants were able to see their final forecast, the statistical forecast and the actual value, thus facilitating learning over trials. Overall, the provision of statistical forecasts was beneficial, which is consistent with our first hypothesis. In addition, more accurate statistical forecasts provided greater improvements in accuracy.

Thus, based on Goodwin and Fildes' (1999) reasoning and Lim and O'Connor's (1995) findings, we have:

Hypothesis 2. Formal forecasts based on cleansed series are more beneficial than those based on non-cleansed series.

Judgmental forecasting from time series appears to depend on the use of anchoring heuristics (Lawrence & O'Connor, 1992). Given an un-trended data series that includes both normal and promotional periods, unaided forecasters are likely to anchor on the mean of that series, then adjust either upwards to allow for the presence of a planned promotion in the forecast period or downwards to allow for the absence of a promotion (Fig. 1, upper panel). Given that the adjustments made when anchoring heuristics are used are typically insufficient (Tversky & Kahneman, 1974), we expect under-forecasting in promotional periods but over-forecasting in normal ones. As the statistical forecasts based on non-cleansed series follow the mean of the data series, we expect the mental anchor to be used to be the same as for unaided forecasting. Thus, taking directional error to be given by the outcome minus the forecast, we have:

Hypothesis 3a. For forecasting that is unaided or aided by statistical forecasts based on non-cleansed data series, directional error will be positive for normal periods and negative for promotional ones.

However, when the statistical forecasts are based on cleansed data series, the mean of the statistical forecast history will approximate the mean of the non-promotional periods. Hence, no adjustment is needed for predicting sales in a period when no promotion is planned. However, an upward adjustment is still needed for promotional periods (Fig. 1, lower panel), and this adjustment will be insufficient. Hence:

Hypothesis 3b. For forecasting that is aided by statistical forecasts based on cleansed data series, the directional error will be zero for normal periods and negative for promotional periods.

The statistical forecasts based on non-cleansed series tend to lie between the sales level associated with non-promotional periods and the average sales level associated with promotional periods. The historical mean of the statistical forecasts will be much closer to the actual baseline of the series when the ratio of promotional to non-promotional periods is low (e.g., 10%) than when it is high (e.g., 40%). This should benefit forecasting for periods without promotions, as the adjustment required will be minimal. However, when this ratio is low, there is less information from which to estimate the relationship between the promotional size and its effect, which is likely to impair forecasting for promotional periods. Thus, when statistical forecasts are based on non-cleansed series:

Hypothesis 4. A lower proportion of promotions in the data series will benefit forecasts for non-promotional periods but impair those for promotional periods.

When there are relatively few promotional periods in the data, statistical forecasts based on non-cleansed data series will be closer to the baseline, and will therefore approximate statistical forecasts based on cleansed data. In contrast, when the proportion of promotional periods is high, the statistical forecasts based on non-cleansed series will be well above the baseline, and the difference between them and statistical forecasts based on cleansed series will be larger. Hence:

Hypothesis 5. Any difference between the benefits derived from the two types of statistical forecasts will be greater when the proportion of promotional periods in the data series is higher.

3. Experiment 1: Use of statistical forecasts derived from series that have or have not been cleansed of promotional effects

A mixed design was used to test these hypotheses. The type of task (unaided judgmental forecasting/forecasting aided by statistical forecasts based on non-cleansed series/forecasting aided by statistical forecasts based on cleansed series) was varied between participants, while the proportion of promotions in the data presented (40% versus 10%) and the forecasting for promotional versus non-promotional periods were varied within participants.

3.1. Method

3.1.1. Participants
A total of 153 students from University College London participated in the study. Their mean age was 18.56 years (SD = 1.03 years), and 127 of them were female.

3.1.2. Design and stimulus materials
Forty series, each consisting of 50 data points, were generated using the R statistical software. Half of the series were independent (mean = 300, error = 7%) and half

Table 1
Experiment 1: Mean values of participants' mean absolute errors (MAE) for each level of promotion frequency and each type of promotion period in the three conditions of the experiment.

Independent variables		Statistical forecast			
		None	Cleansed	Not cleansed	Mean
40% promotions	Promotion	33.48	30.13	30.24	31.28
	No promotion	35.06	28.66	31.67	31.79
10% promotions	Promotion	34.54	31.11	29.43	31.69
	No promotion	29.94	24.44	23.73	26.03

were autoregressive (mean $= 300$, $\rho = 0.7$, error $= 7\%$). The series were displayed as grey lines and were labelled 'sales'. Each graph also contained vertical blue bars[2] that represented promotional expenditure on either five or 20 of the 50 periods. Both the locations and sizes of these promotions were assigned randomly. The promotion size was selected at random without replacement from a list of every tenth value between 50 and 200. The size of the promotion had a same-week effect on the sales number according to the following formula:

$$PI_t = \frac{Pt}{5} * S_t.$$

This indicates a same-week percentage increase PI at time t (over the regular sales S at that time) that is equal to one-fifth of the promotional expenditure P.

The participants were asked to forecast both one step ahead (period 51) and two steps ahead (period 52). A promotion was present in either time period 51 or time period 52. The size of this promotion was randomized for every participant across the trials. Over the experimental session, it included every tenth value between 30 and 220. Thus, the participants were required to forecast four promotion sizes (30, 40, 210, 220) that were not included in the range presented in the data series (i.e., 50–200).

The presence and type of statistical forecasts were manipulated between participants. The first group (A) did not receive statistical forecasts (Fig. 2, panel (a)). The other two groups each received statistical forecasts calculated using the Holt-Winters exponential smoothing method, where a line graph was used to represent the statistical forecast history for weeks 2 to 52. One of these groups (B) received a statistical forecast based on the baseline data series, cleansed of promotional effects (Fig. 2, panel (b)), while the other (C) received a statistical forecast based on the total sales, where no distinction was made between normal and promotional periods in calculating it (Fig. 2, panel (c)).

3.1.3. Procedure

The participants were given instruction sheets that differed only on the explanation of the statistical forecast (see the Appendix). In addition, they were orally instructed to pay close attention to the explanation of the graphical components on their instruction sheet, and were given a short demonstration of two example trials.

3.2. Results

We present analyses of three error scores[3]: the mean absolute error (MAE), mean error (ME), and variable error (VE).

3.2.1. Mean absolute error

The MAE was used to measure the overall error level. The errors were calculated relative to the ideal forecast, which was provided by the *signal* (excluding the noise) of the time series in non-promotional periods and by the signal plus the promotion effect in promotional periods. For the independent time series, the ideal forecast for a non-promotional period was 300 (the mean). For the autoregressive series, it was 0.2 $(1 - \rho)$ of the distance between the last data point and the mean. An outlier analysis indicated that two participants had scored more than two standard deviations away from the mean on at least half of the trials, so these two were excluded from the analyses.

Table 1 shows MAE values for each combination of the three independent variables. The overall mean value of MAE was 30.20 ($SD = 9.65$).

An analysis of the variance with the statistical forecast as a between-participants variable and the promotion frequency and promotion presence as within-participant variables revealed a main effect of the statistical forecast (F (2,150) $= 3.99$, $p = 0.021$, $\eta_p^2 = 0.050$). Hypothesis 1 stated that the provision of a statistical forecast would be beneficial for the forecasting accuracy, such that unaided judgment would result in higher errors than aided judgment. One-tailed t-tests confirm that the MAE of the unaided judgment group ($MAE = 33.26$, $SD = 9.84$) was significantly higher than those of both the cleansed ($MAE = 28.59$, $SD = 9.29$; t (100) $= 2.47$, $p = 0.008$) and non-cleansed ($MAE = 28.77$, $SD = 9.26$; t (100) $= 2.37$, $p = 0.010$) forecast groups.

The MAE scores of the cleansed and non-cleansed forecast groups were not significantly different from one another (t (100) $= -0.10$, $p = 0.921$), meaning that we failed to find support for Hypothesis 2, which stated that participants who were given a statistical forecast based on cleansed series would be more accurate than those who were given a statistical forecast based on non-cleansed series.

---

[2] The colours in this paper have been converted to greyscale. Thus, the screenshots that we provide show the sales series as a dark grey line, the statistical forecast as a light grey line, and the promotional expenditure as light grey bars.

---

[3] Promotional increases and sales were on the same scale in all graphs that the participants saw. Thus, it is appropriate to use scale-dependent measures because their meaning is immediately transparent (Hyndman & Koehler, 2006). Scale-independent measures often suffer from asymmetry and do not cope well with values that are close to zero (Hyndman & Koehler, 2006). Given these considerations, we opted to use scale-dependent error measures.
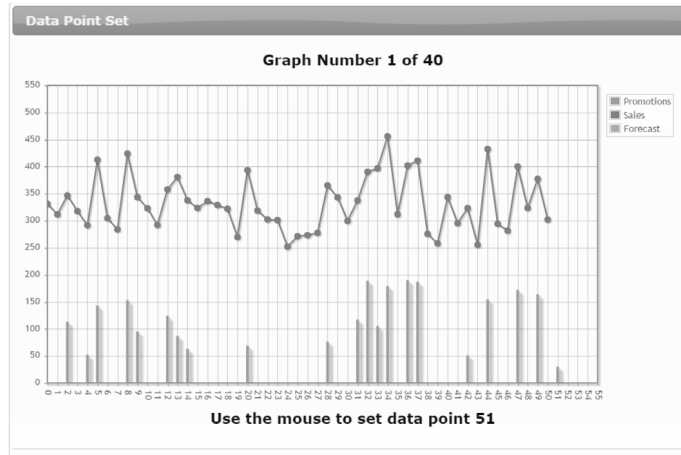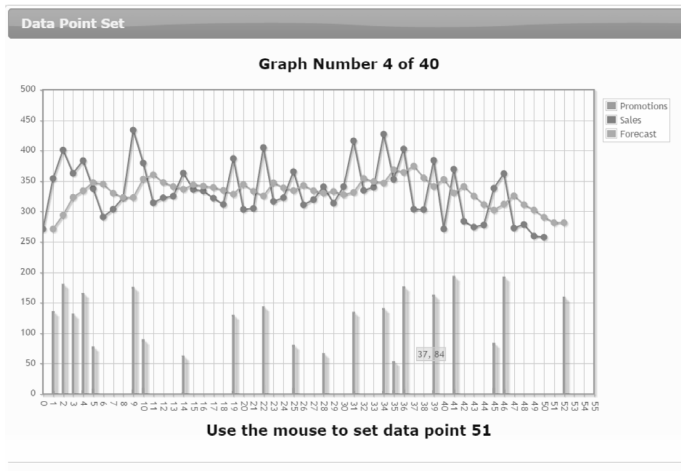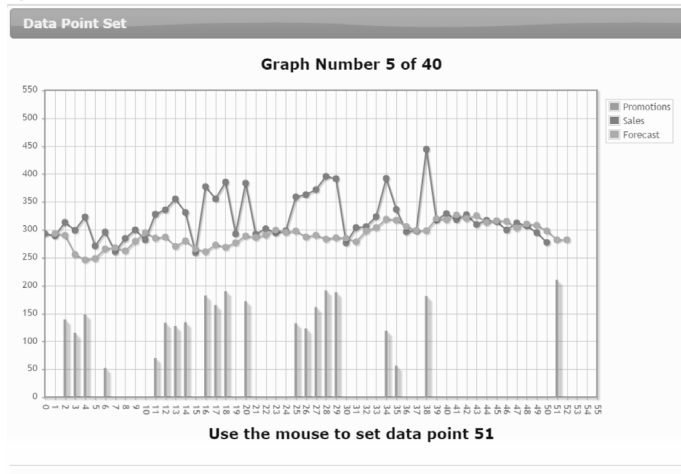
Fig. 2. Experiment 1: (a) Example screenshot in the unaided group; (b) Example screenshot in the group aided by statistical forecasts based on non-cleansed data series; (c) Example screenshot in the group aided by statistical forecasts based on cleansed data series.

There was a main effect of the frequency of promotions in the data series (F (2,150) = 28.21, $p < 0.001$, $\eta_p^2 = 0.158$), a main effect of the presence of a promotion in the period to be forecast (F (2,150) = 7.35, $p = 0.008$, $\eta_p^2 = 0.047$), and an interaction between these two variables (F (2,150) = 743.82, $p < 0.001$, $\eta_p^2 = 0.226$). An analysis

Table 2
Experiment 1: Mean values of the mean error (ME) and variable error (VE) for each level of promotion frequency and each type of promotion period in the three conditions of the experiment.

Independent variables		ME				VE			
		No SF	Cleansed	Not cleansed	Means	No SF	Cleansed	Not cleansed	Means
40% promotions	Promotion	−8.07	−6.47	−3.85	−6.13	32.75	30.41	29.50	30.89
	No promotion	18.21	14.08	18.38	16.89	29.92	28.01	28.51	28.81
10% promotions	Promotion	−10.93	−11.39	−7.04	−9.79	32.29	29.35	28.50	30.04
	No promotion	12.63	8.07	9.84	10.18	29.14	25.58	23.40	26.04

of simple effects showed that these effects arose because lower errors were found for less frequent promotions when forecasting non-promotional periods, but not when forecasting promotional ones (F (1, 150) = 61.66, $p < 0.001$).

We failed to obtain support for Hypothesis 5: there was no significant interaction between the type of statistical forecast provided and the frequency of promotions in the data series.

3.2.2. Mean error

The MAE score discussed above is a measure of the overall error. Following Thurstone (1926), the overall error can be decomposed into directional error or bias (ME) and scatter or variable error (VE). Taking D as (Forecast – Actual), ME is defined as $\Sigma D/n$ and VE as $\sqrt{([\Sigma (D - ME)^2]/n)}$. Thus, the overall error could theoretically comprise (a) bias but no scatter (all forecasts are a fixed distance from the optimal forecast with no distribution around that point), (b) scatter but no bias (forecasts are distributed around a central point, but that central point is the optimal forecast), or (c) bias and scatter (forecasts are distributed around a central point that is a fixed distance from the optimal forecast). In practice, both bias and scatter contribute to the overall error, but their relative contributions depend on contextual factors.

We investigate the reasons for the differences in MAE reported above and test Hypotheses 3–5 by reporting analyses of ME in this section and VE in the following one (see Table 2).

No effects involving the statistical forecast variable were significant. There was a main effect of whether the forecast was for a period with promotions (F (2,150) = 126.69, $p < 0.001$, $\eta_p^2 = 0.458$): the mean error was negative when the forecasts were for periods in which promotions were planned, but positive when they were for periods with no promotions planned. There was also a main effect of the proportion of promotions in the data series (F (2,150) = 67.17, $p < 0.001$, $\eta_p^2 = 0.309$): overall, ME was lower when the proportion of promotions in the data series was low than when it was high. There was also a significant interaction between these two variables (F (2,150) = 6.49, $p = 0.012$, $\eta_p^2 = 0.041$). An analysis of simple effects showed that this arose because a lower proportion of promotions in the data series decreased the positive ME of forecasts for non-promotional periods (F (1, 150) = 63.77, $p < 0.001$), but increased the negative ME of forecasts for promotional periods (F (1, 150) = 16.54, $p < 0.001$). This result is consistent with Hypothesis 4, which states that having fewer promotions will benefit the forecasts for non-promotional periods but impair those for promotional ones.

Hypothesis 3a predicted that the use of statistical forecasts based on *non-cleansed* series would lead to under-forecasting for promotional periods but over-forecasting for normal periods. One-sample t-tests confirm that the ME was significantly below zero in promotional periods (t (50) = −2.27, $p = 0.028$) and significantly above zero (t (50) = 8.20, $p < 0.001$) in normal ones.

Hypothesis 3b predicted that the use of forecasts based on *cleansed* series would lead to an ME of zero for normal periods and a negative ME for promotional periods (i.e., under-forecasting). While the ME for promotional periods in the cleansed series condition was indeed significantly below zero (t (50) = −3.51, $p = 0.001$), that for normal periods was positive and significantly different from zero (t (50) = 7.37, $p < 0.001$). This unexpected over-forecasting on normal periods was greater when there were 40% of promotions in the data series than when there were only 10% of promotions in the data series (t (50) = 4.25, $p < 0.001$).

3.2.3. Variable error

There was a significant effect of group on VE (F (2,150) = 3.56, $p = 0.031$). We hypothesized that the error of the unaided judgment group would be higher than those of the aided judgment groups. One-tailed t-tests confirm that the VE of the unaided judgment group was indeed larger than those of both the group who received cleansed forecasts (t (100) = 2.04, $p = 0.044$) and the group who received non-cleansed forecasts (t (100) = 2.53, $p = 0.013$). The VE scores in the latter two groups were not significantly different from one another.

The forecasts from the data series with 40% of promotions had higher VE scores than those from the series with 10% of promotions (F (1,150) = 10.23, $p = 0.002$, $\eta_p^2 = 0.064$). In addition, the forecasts for promotional periods had higher VE scores than those for non-promotional ones (F (1,150) = 24.66, $p < 0.001$, $\eta_p^2 = 0.141$). In addition, there was also an interaction effect between these two variables (F (1,150) = 4.84, $p = 0.029$, $\eta_p^2 = 0.031$). An analysis of simple effects indicated that the error difference between normal and promotional periods was more pronounced in the low promotion frequency trials (F (1,150) = 29.26, $p < 0.001$, $\eta_p^2 = 0.163$) than in the high promotion frequency trials (F (1,150) = 7.37, $p = 0.007$, $\eta_p^2 = 0.047$).

3.3. Discussion

The experiment produced two separate groups of effects. The first group is the effects on MAE and VE of whether the participants made unaided forecasts, made

forecasts after being given non-cleansed statistical forecasts, or made forecasts after being given cleansed statistical forecasts. The second is the effects on MAE, ME, and VE of the proportion of promotional periods in the data series and of whether the forecasts were being made for promotional or normal periods. As there were no interactions between these two groups of effects, we will discuss them separately, after which we will provide a summary of the cognitive processes underlying the performance that explain both types of effects.

3.3.1. Effects of providing forecast support

Providing statistical forecasts reduced the overall error (MAE). However, further analysis showed that this was not because these forecasts reduced the directional error or bias (ME) in forecasts, but because they reduced the random error or scatter (VE); that is, they made the forecasts more consistent.

We anticipated that the cleansed statistical forecasts would improve forecasting more than the non-cleansed ones. However, our rationale for this relied on our expectation that the cleansed forecasts would lower the bias by reducing the under-adjustment from the mean of the series – the salient anchor in the unaided and non-cleansed statistical forecast conditions. It was on this basis that we generated Hypotheses 2, 3a, and 5. However, no differences in the effectiveness of the cleansed and non-cleansed statistical forecasts were evident, either as main effects or as interactions in our analyses of MAE, ME and VE. Thus, cleansing the data did not affect the degree of under-adjustment from the mean of the series.

The provision of statistical forecasts improved the overall accuracy regardless of whether the data were cleansed or not, but there was no difference in the degree to which they did so, because there was no difference in the extent to which they reduced VE.

3.3.2. Effects of promotions in the data series and in the periods to be forecast

The proportion of promotions in the data series and whether the forecast was for a normal or promotional period interacted in their effects on the overall forecast accuracy (Table 1): a greater proportion of non-promotional periods in the data specifically helped the forecasts for non-promotional periods. To understand the reason for this, we need to consider the separate analyses of ME and VE.

Forecasters are likely to anchor on the overall mean of the data series (Lawrence & O'Connor, 1992). Fewer promotions meant that the overall mean was closer to the mean value of the non-promotional periods, but further from the mean value of the promotional periods. Thus, having fewer promotions in the data series meant that a larger adjustment from the overall mean of the series was needed for forecasting promotional periods, but a smaller adjustment was needed for forecasting non-promotional periods. The data show that the degree of under-adjustment was greater when a larger adjustment was needed. This is to be expected. In psychophysics, the Weber-Fechner Law (Baird & Noma, 1978; Fechner, 1860; Weber, 1834) summarizes many findings showing that errors in discrimination tend to be proportional to the overall size of the stimulus being judged. Hence, because the extent of under-adjustment was proportional to the size of the required adjustment, ME became less positive in normal periods but more negative in promotional ones as the proportion of promotions in the data series decreased (Table 2).

A higher proportion of promotions in the data series increases the variability of the series. If people used their estimate of the overall mean of the series as a judgment anchor, this estimate would have been more variable when the proportion of promotions in the data series was higher. As a result, VE is also higher (Table 2).

Forecasters must adjust away from this initial judgment anchor to allow for the absence or presence of a promotion in the period to be forecast. When there is no promotion planned, forecasters merely need to use the data series to estimate the mean value of sales when there is no promotion (and to move their judgment away from the initial anchor towards that mean value). However, when a promotion is planned, they need to do more than just estimate the mean value of sales when a promotion occurred: they also have to take into account the relationship between the size of a promotion and its elevating effect on sales. This can be done in various ways (e.g., via some kind of mental regression). However, it is reasonable to assume that this additional process will be imperfect, and will therefore add some random error to the forecasts. As a result, VE was higher in forecasts for promotional periods (Table 2).

The reasons for the relatively low value of MAE when forecasts for normal (rather than promotional) series were made from series with 10% (rather than 40%) of promotions are now clear. VE is lower when forecasting normal periods rather than promotional periods, as well as when there are fewer promotions in the data series. Finally, having fewer promotions in the data also results in a reduction in the size of the positive ME associated with forecasts that are made for normal periods. This combination of two factors reducing VE (normal periods, fewer promotions) and one factor reducing ME (fewer promotions) results in a particularly low MAE value. The MAE is higher in all other cases because the other factors that lower either VE or ME do not combine in the same felicitous manner. For example, consider the case where forecasts are made for a promotional period from data series containing 40% promotions. Here, the higher proportion of promotions in the data series reduces the size of the negative ME associated with making forecasts for promotional periods. However, the beneficial effects of this are counteracted by the fact that VE is higher both when forecasts are made for promotional periods and when the proportion of promotions in the data series is higher.

Why did the presence of a statistical forecast lower VE, and hence, MAE? We have argued that forecasters begin by estimating the overall mean of the data series, and that this then acts as an initial judgment anchor. Furthermore, this is an error-prone process: VE is higher when the data series is more variable. Both types of statistical forecast act to make it less error-prone. Forecasters could reduce the random error in their estimate of the series mean simply by averaging it with either the non-cleansed statistical forecast or the cleansed statistical forecast plus some increment that is specific to the proportion of promotions in the data series.

3.3.3. Summary

We can explain the patterns in the data by assuming that people follow a two-step process when producing their forecasts. They begin by estimating the overall mean of the data series in order to use it as an initial judgment anchor. The size of the random error associated with this estimate is higher when the data series are more variable, but can be reduced by providing a statistical forecast.

They then adjust away from this initial anchor based on whether a promotion is planned or not. Under-adjustment results in under-forecasting in promotional periods and over-forecasting in normal ones. The size of the under-adjustment is greater when a larger adjustment is required; hence, a larger proportion of promotions in the data series results in larger (positive) ME values in normal periods but smaller (negative) ME values in promotional periods. When normal periods are being forecast, adjustments are based on just the mean value of sales in non-promotional periods, but in promotional periods they must take into account the relationship between the size of promotions and the size of their effects. This additional process is error-prone, leading to higher VE results in promotional periods.

4. Experiment 2: Use of statistical forecasts that take into account the effects of promotions

Unexpectedly, Experiment 1 failed to reveal any difference in forecast accuracy between the participants who received the cleansed statistical forecasts and those who received the non-cleansed ones. Non-cleansed statistical forecasts are cruder: they require less processing of the data series, but always require some adjustment. In contrast, cleansed forecasts provide a clearly defined baseline series, and therefore can be accepted without adjustment for non-promotional periods. However, the participants still made large upward adjustments in these periods (Table 2).

It is possible that people see that the cleansed statistical forecasts do not account for promotions and falsely infer that they cannot be 'trusted' for normal periods either, and therefore make adjustments for both types of period. In order for forecasts to be acceptable, they need to be relevant, justifiable and valuable for dealing with future uncertainties (Gönül, Önkal, & Lawrence, 2006). The clear unacceptability of the cleansed statistical forecasts for promotional periods may have been generalized inappropriately to represent the acceptability of those forecasts for both types of period (promotional and normal).

This possibility prompted us to carry out Experiment 2, in which we provided the participants with 'optimal' forecasts. Each forecast for a promotional period was based on the cleansed statistical forecasts, but with the appropriate increase in sales produced by the promotion in the promotional period added to it. While the idea of obtaining such forecasts in business practice is not completely realistic, it is an approach that can now be approximated by various recently developed forecasting methods that include promotional modelling (e.g., Huang et al., 2014; Kourentzes & Petropoulos, 2016; Trapero et al., 2013).

We suggested above that the reason why cleansed forecasts for non-promotional periods are not accepted without adjustment is that it is clear to forecasters that the cleansed forecasts for promotional periods are unacceptable without adjustment, leading to a lack of trust in all forecasts. As a result, all forecasts are adjusted. In the present experiment, it was made clear to the forecasters that the forecasts for both promotional and normal periods were acceptable without adjustment. If our suggestion is correct, then the forecasters would now have no reason not to trust the statistical forecasts, and therefore they should be judged acceptable and adopted without adjustment. More formally,

Hypothesis 6. Optimal forecasts will be accepted without adjustment.

4.1. Method

The experiment was identical to Experiment 1, except that the statistical forecasts for promotional periods were elevated by an amount that was appropriate to the size of the planned promotion.

4.1.1. Participants

Fifty students from University College participated in the study. Their mean age was 17.77 years ($SD = 0.87$ years) and 40 of them were female.

4.1.2. Stimulus materials, design and procedure

In both the instructions and the experiment, the statistical forecasts were presented as shown in Fig. 3. The instructions with regard to the statistical forecast were adapted as follows: "*The orange line is a forecast from a statistical model[2]. The model is based on the cleaned sales data: the promotion effects have been taken out of the data until only the baseline remained. The model uses these baseline data to produce the statistical forecasts and then adds the promotion effects on top of this forecast. You can see the predictions it made in the past and what it predicts for time periods 51 and 52. You can choose whether or not to follow the statistical forecast*". In all other respects, the experiment was identical to the first one.

4.2. Results

The MAE, ME, and VE results of forecasts for periods with and without promotions and from data series with low and high frequencies of promotions are shown in Table 3. A repeated-measures ANOVA of the MAE revealed a main effect of the frequency of promotions ($F (1, 49) = 30.15$, $p < 0.001$), indicating that forecasts were more accurate when there were fewer promotions in the data series, and a main effect of the presence of a promotion ($F (1, 49) = 18.62$, $p < 0.001$), showing that forecasts were more accurate for normal periods than for promotional periods. An analysis of the ME revealed only a main effect of the frequency of promotions ($F (1, 49) = 26.03$, $p < 0.001$), indicating slight over-forecasting when the data series contained 40% promotional periods but slight under-forecasting when they contained 10% promotional periods.
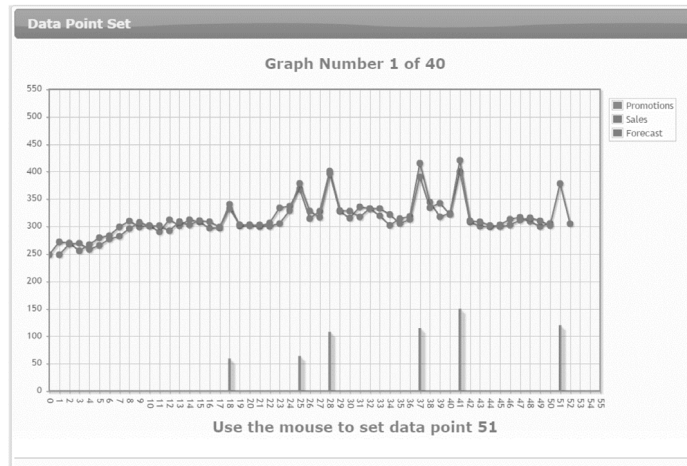
Fig. 3. Experiment 2: Example screenshot.

An analysis of VE indicated a main effect of the frequency of promotions ($F (1, 49) = 17.15$, $p < 0.001$) and a main effect of the presence of a promotion ($F (1, 49) = 13.03$, $p = 0.001$). The directions of these effects mirrored those obtained for MAE.

4.2.1. Comparison of the performances with those obtained in Experiment 1

Did the enhanced statistical forecasts provided in this experiment lead to better final forecasts than those obtained using unaided judgment or judgment aided by the provision of other types of statistical forecasts? To find out, we compared the forecast accuracies with those obtained in the three conditions of Experiment 1 (Fig. 4). With regard to the overall error (MAE), the performance was significantly better than unaided judgment ($F (1, 99) = 27.77$, $p < 0.001$), aided judgment with a non-cleansed-series statistical forecast ($F (1, 99) = 14.71$, $p < 0.001$) and aided judgment with a cleansed-series forecast ($F (1, 99) = 23.13$, $p < 0.001$).

We compared the sizes of ME scores across experiments by analyzing their absolute values. As hypothesized, those in the present experiment were lower than those in all conditions of the previous experiment: unaided judgment ($t (58) = −6.12$, $p < 0.001$), judgment aided with non-cleansed statistical forecasts ($t (63) = −5.46$, $p < 0.001$), and judgment aided with cleansed statistical forecasts ($t (60) = −4.56$, $p < 0.001$).[4] Similarly, the VE was significantly lower in the current experiment than in all conditions of the previous experiment: unaided judgment ($t (76) = −7$, $p < 0.001$), judgment aided with non-cleansed statistical forecasts ($t (68) = −3.02$, $p = 0.004$), and judgment aided with cleansed statistical forecasts ($t (73) = −4.11$, $p < 0.001$)[4].

---

[4] In cross-experimental comparisons of ME and VE, Levene's test indicated unequal variances, and therefore the degrees of freedom were adjusted accordingly.

Table 3
Experiment 2: Mean values of the mean absolute error (MAE), mean error (ME) and variable error (VE) for each level of promotion frequency and each type of promotion period.

Independent variables		MAE	ME	VE
40% promotions	Promotion	25.61	1.99	26.43
	No promotion	23.09	2.54	24.22
10% promotions	Promotion	23.02	−2.8	23.98
	No promotion	19.50	−0.59	21.09

4.3. Discussion

The provision of optimal statistical forecasts reduced all types of error significantly relative to the corresponding error levels observed in all conditions of Experiment 1. In particular, the absolute size of the directional error reduced very considerably. This implies that the under-adjustment from the initial anchor was reduced strongly. However, the MAE scores show that a fair amount of error still persisted (Fig. 4), driven primarily by VE. Although this type of error was significantly lower in Experiment 2 than in any of the conditions in Experiment 1, it remained high at 83% of its size in that experiment. As before, it was larger when there were more promotions in the data series and when promotions were planned for a forecast period. These influences on VE are likely to explain their re-appearance in the analyses of MAE. These results indicate that Hypothesis 6 should be rejected.

Lim and O'Connor (1995, Experiment 3) obtained similar findings to ours. They found that forecasters made insufficient use of near-perfect statistical forecasts that were generated by taking the average of a highly reliable statistical forecast and the actual outcome. Forecasters put too much weight on their own views and not enough on the statistical forecast. Similarly, Gardner and Berry (1995) found that people performing a control task who were freely offered perfectly correct advice decided to obtain it on only 44% of occasions. Furthermore, those who obtained
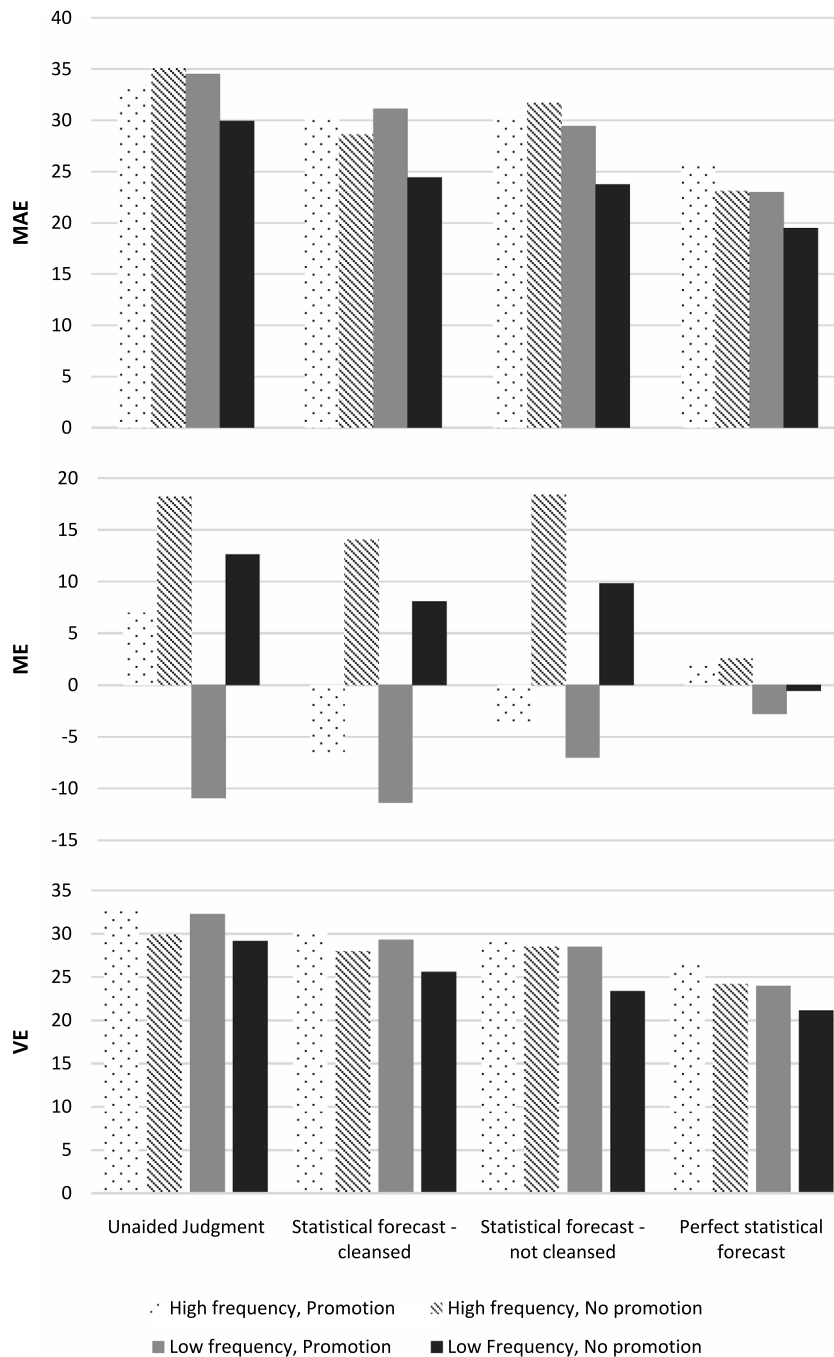
Fig. 4. Error scores associated with the different forecasting conditions studied in the two experiments: MAE (upper panel); ME (central panel); VE (lower panel).

it acted in accordance with it on only 73% of occasions. One interpretation of both of these results is that people tend to be overconfident in their own abilities, and therefore do not take sufficient account of good advice.

According to our description of the results from Experiment 1, forecasts are produced in two stages. First, forecasters (even those who are provided with statistical forecasts) make their own assessment of the mean of the data series to use as an initial judgment anchor. This

assessment is subject to random error that is reflected in the VE scores. The random error is greater when the data series are more variable, and they are more variable when they contain a higher proportion of promotions; hence, VE is greater when the proportion of promotions in the data series is higher. As the same effect was found in the present experiment, it seems reasonable to assume that forecasters initially processed the series in a similar way in the present experiment.

The statistical forecasts examined in Experiment 1 were beneficial because they reduced VE. The statistical forecasts used in the present experiment also reduced VE. We suggested that this reduction occurs because forecasters can obtain estimates of the series mean from both the raw data series and the series of past statistical forecasts, whereas unaided judgmental forecasters can use only the data series. A weighted average of these two estimates then provides the initial judgment anchor. If people are less confident in the statistical forecasts, they may not put enough weight on the estimate obtained from them. This may lead to VE being reduced, but not by as much as it could be. In the present experiment, the reduction in VE was greater than that produced by the statistical forecasts provided in Experiment 1. This may have been because the description in the instructions of the way in which the statistical forecasts were generated gave forecasters greater confidence in them, leading them to place more weight on them, and therefore generate more accurate estimates of the series mean to use as an initial judgment anchor.

In the second stage, forecasters adjust away from the initial judgment anchor to take into account the presence or absence of a promotion in the period to be forecast. We saw in Experiment 1 that such adjustments are typically insufficient (Tversky & Kahneman, 1974), meaning that promotional periods are under-forecast whereas normal periods are over-forecast. The adjustments for normal periods are based only on the mean value of normal periods in the data series, but those for promotional periods have to take into account the relationship between the sizes of promotions and the elevations in sales that they produce. This additional process is error-prone and therefore increases the VE of forecasts for promotional periods relative to those for normal periods.

This same effect (higher VE values in promotional periods) was found in the present experiment. However, in contrast to Experiment 1, analyses of ME showed that there was no evidence of either under-forecasting in promotional periods or over-forecasting in normal ones. Thus, including an element that allows for promotions in statistical forecasts is beneficial not just because it reduces VE, but also because it reduces the absolute size of ME. However, the VE of forecasts for promotional periods was still higher than those for normal ones. This implies that people do not merely accept the statistical forecast. Their low ME scores show that, on average, the mean values of their forecasts for both normal and promotional periods are very close to those provided by the statistical forecasts. However, there is a considerable degree of scatter around these mean values, and this scatter is greater for forecasts of promotional periods. We attribute this greater scatter to the additional error-prone cognitive processing that is needed to allow for the promotion function (i.e., the relationship between the promotion size and its effect).

Statistical forecasts that include an element that allows for the effects of promotions are beneficial because they reduce both the bias and the random error in forecasts. However, forecasters do not accept them automatically. In fact, of the 4000 forecasts that participants made in Experiment 2, only 333 (8.33%) were equal to the statistical forecast that they had been given. Thus, adjustments were still made and must have been responsible for the high levels of VE that persisted in this experiment.

The levels of VE were also affected by the same variables as in Experiment 1: the nature of both the data series (proportion of promotions) and the periods to be forecast (normal or promotional). Because these variables affect VE levels in the same way when statistical forecasts (of whatever type) are provided as in unaided judgmental forecasting, we consider that the provision of statistical forecasts does not alter the cognitive processes that forecasters employ to perform their task in any fundamental way. Instead, they facilitate these processes, though more for some (e.g., the 'de-biasing' observed in Experiment 2) than for others (e.g., extracting an initial mental anchor from the data series). In other words, forecasters still use an anchoring-and-adjustment heuristic when they are given optimal statistical forecasts, but their estimate of the appropriate anchor is somewhat more consistent and their adjustment from that anchor is almost free of bias.

5. Comparison of participants' performances with those achieved by raw statistical forecasts

Thus far, we have compared the accuracy of unaided judgmental forecasts with those of judgmental forecasts made following the provision of statistical forecasts of various types. This type of comparison addressed the primary question that motivated the work reported here: is it worthwhile to provide judgmental forecasters with statistical forecasts? This section addresses a different question that is also of interest but was not a primary motivator of our work: does the judgmental adjustment of raw statistical forecasts improve the forecast accuracy?

We answered this question by comparing the accuracy of judgmental forecasts (made with or without access to statistical forecasts) with that of raw statistical forecasts. Previous non-experimental research that has analyzed company forecast records has indicated that forecasters tend to make too many adjustments (Frances & Legerstee, 2009), but that, on the whole, these adjustments still tend to produce final forecasts that are better than the original raw statistical ones (Syntetos, Nikolopoulos, & Boylan, 2010; Syntetos, Nikolopoulos, Boylan, Fildes, & Goodwin, 2009). However, these studies did not examine the effect of the proportion of promotions in the data series.

Table 4 shows the MAE scores of participants' forecasts and of raw statistical forecasts for series with 40% and 10% of promotions in the four conditions of the two experiments. Begin by considering the three conditions in which statistical forecasts were provided. When those forecasts were based on non-cleansed data, the participants' forecasts were *more* accurate than the raw statistical forecasts when there was a large proportion (40%) of promotions in the data series ($t(50) = 4.27$; $p < 0.001$), but *less* accurate than the raw statistical forecasts when there was a small proportion (10%) of promotions in the data series ($t(50) = 7.54$; $p < 0.001$). The pattern was similar when the statistical forecasts were based on cleansed data: the participants' forecasts were *more* accurate than the raw statistical forecasts when there was a large proportion

Table 4
Mean absolute error (MAE) scores of participants' forecasts and of raw statistical forecasts for series with 40% and 10% promotions in the four conditions of the two experiments.

Independent variables		Statistical forecast condition			
		None	Not cleansed	Cleansed	Optimal
40% promotions	Participants' forecast	34.43	31.10	29.25	24.10
	Statistical forecast	N.A.	36.56	36.86	12.87
10% promotions	Participants' forecast	30.40	24.30	25.11	19.85
	Statistical forecast	N.A.	13.66	15.47	8.88

(40%) of promotions in the data series (t (50) = 5.71; $p <$ 0.001), but *less* accurate than the raw statistical forecasts when there was a small proportion (10%) of promotions in the data series (t (50) = 7.21; $p <$ 0.001). However, in Experiment 2, when the statistical forecasts were optimal, the participants' forecasts were *less* accurate than the raw statistical forecasts regardless of whether the proportion of promotions in the data series was large (40%; t (49) = 19.07; $p <$ 0.001) or small (10%; t (49) = 17.37; $p <$ 0.001).

Now consider the condition in Experiment 1 where no statistical forecasts were provided. As the participants in this condition received the same set of data series as those in the other conditions, we can compare their performances with those of the statistical forecasts in the other three conditions. These analyses revealed that, when there was a high proportion (40%) of promotions in the data series, the participants' performances in the unaided condition were not significantly different from that achieved by statistical forecasts based on non-cleansed (t (50) = 1.49; NS) or cleansed (t (50) = 1.70; NS) data, but were worse than the performance of optimal forecasts (t (50) = 15.04; $p <$ 0.001). In contrast, when there was a low proportion (10%) of promotions in the data series, the performances of the participants in the unaided condition were worse than those achieved by all three types of statistical forecast: those based on non-cleansed data (t (50) = 9.54; $p <$ 0.001), those based on cleansed data (t (50) = 8.51; $p <$ 0.001) and the optimal forecasts (t (50) = 12.26; $p <$ 0.001).

In summary, judgmental adjustment was beneficial only when (a) the statistical forecasts did not take promotions into account, *and* (b) a high proportion of periods in the data series were affected by promotions. When the statistical forecasts did take promotions into account, the forecasters made unnecessary adjustments (c.f. Frances & Legerstee, 2009), but when the forecasts did not take promotions into account, the adjustments improved the forecasts if there was a high proportion of promotion periods in the data series (c.f. Syntetos et al., 2010, 2009).

6. General discussion

We provided forecasters with different types of statistical forecasts in order to investigate how effective such forecasts are at improving the forecasters' accuracies. We also varied both the type of period (normal versus promotional) to be forecast and the proportion of promotional periods in the data series, because we expected these factors to influence the benefits that statistical forecasts bestow

on the forecasting performance.[5] Finally, we developed an account of the way in which forecasts are produced from time series that are perturbed by sporadic events (i.e., promotions) and of the effects on those forecasts when forecasters have access to statistical forecasts. We now discuss each of these aspects of our work in turn.

6.1. Effects of statistical forecasts on the forecast accuracy

Statistical forecasts that do not take into account whether periods in the data series were affected by sporadic events, such as promotions, form the most common form of forecasting support provided to practitioners (e.g., Fildes et al., 2009; Trapero et al., 2013). However, in experimental research (e.g., Goodwin & Fildes, 1999; Goodwin et al., 2011), researchers have also investigated the usefulness of statistical forecasts based only on normal periods that are not subject to promotions. We expected the latter approach to be more effective in improving the forecasting accuracy (Hypothesis 2).

While both of these types of statistical forecast improved the accuracy relative to that observed for unaided judgmental forecasting (Hypothesis 1), there was no difference in the degree to which they did so. Given previous work by Lim and O'Connor (1995) and the persuasiveness of the arguments in favour of using statistical forecasts based on cleansed data series, this finding was unexpected. However, the rationale for Hypothesis 2 was based on the assumption that statistical forecasts reduce the bias: we anticipated that the anchoring bias for normal periods would be removed when the statistical forecasts were based on cleansed rather than uncleansed series. However, our data actually show that the statistical forecasts were effective because they reduced the scatter (VE), not the bias (ME), and there is no reason to expect the scatter to be reduced more by statistical forecasts based on cleansed series than by statistical forecasts based on non-cleansed series.

It appears that statistical forecasts that are clearly inadequate for promotional periods affect the degree to which

---

[5] We also varied the type of series (independent versus autoregressive) and the forecast horizon (one step ahead versus two steps ahead). As these variables were not germane to our hypotheses, they were included only to increase the generality of our conclusions, and were not part of our main analyses. However, across all conditions and both experiments, the MAE increased with the forecast horizon (F (1, 204) = 104.57; $p <$ 0.001), as would be expected on the basis of decreased predictability and error accumulation (Harvey, 1995; Theocharis & Harvey, 2016). It was also higher for independent series than for autoregressive series (F (1, 204) = 27.10; $p <$ 0.001), as we would expect on the basis of previous work (Reimers & Harvey, 2011).

forecasters feel able to trust them for normal periods (even when they are actually optimal for those periods). We reasoned that statistical forecasts that are optimal for both promotional and normal periods should be seen as more trustworthy, and therefore be capable of reducing the anchoring bias. Experiment 2 demonstrated that such was the case: the ME values were very close to zero, showing that the anchoring biases were virtually eliminated. However, the VE values still remained high, at 83% of the level observed in the aided conditions of Experiment 1. Nevertheless, the marked drop in overall error (MAE) levels indicates that attempts to incorporate promotional effects into statistical forecasts (e.g., Huang et al., 2014; Kourentzes and Petropoulos, 2016; Trapero et al., 2013) hold great promise for increasing the effectiveness of forecasting support systems.

6.2. Effects of promotions in the periods to be forecast

We expected participants to anchor on the mean level of the data series and to adjust upwards/downwards from this to take into account the presence/absence of a promotion planned for the forecast period. As such adjustments are typically not sufficient (Tversky & Kahneman, 1974), we expected under-forecasting on promotional periods but over-forecasting on normal ones when forecasting was unaided or supported by a statistical forecast based on non-cleansed data series (Hypothesis 3a). This is indeed what we found, which confirmed forecasters' use of the anchoring heuristic. At the same time, though, we expected that this anchoring bias would be absent in normal periods when the statistical forecasts were based on cleansed data series, due to the forecasters realising that the statistical forecast could be accepted without adjustment (Hypothesis 3b). However, as we discussed in the previous section, these forecasts appear not to have been trusted (perhaps because those for promotional periods obviously needed adjustment). Instead, the forecasters continued to use the mean of the series as a judgment anchor and to adjust down from it (insufficiently) when forecasting for normal periods, and hence, the over-forecasting for those periods persisted.

6.3. Effects of the proportion of promotions in the data series

We expected that a lower proportion of promotional periods in the data series would reduce the overall forecasting error in normal periods but increase it in promotional ones (Hypothesis 4). In fact, lowering the proportion of promotions resulted in lower MAEs in normal periods but made no difference in promotional periods. Decomposing the overall error showed why this was the case. In promotional periods, the absolute size of the under-forecasting bias increased when the proportion of promotions in the data series was reduced, but the scatter decreased. As these two effects cancelled one another out, there was no resultant effect on the overall error. (For normal periods, reducing the proportion of promotions in the data series decreased both the over-forecasting bias and the scatter, and hence, the predicted effect occurred.)

When there were fewer promotional periods in the data series, the statistical forecasts derived from non-cleansed series were closer to the baseline forecasts provided by the statistical forecasts derived from cleansed data series. Hence, we expected any accuracy advantage of the statistical forecasts based on cleansed series (over those based on non-cleansed series) to be greater when the proportion of promotions in the data series was higher (Hypothesis 5). However, there was no evidence of any interaction between the proportion of promotions in the data series and the type of statistical forecast. As we have seen, the forecasters in Experiment 1 all appear to have made their judgments in similar ways, regardless of the type of aid they received, if any. The only help that the statistical forecasts provided was to enable them to make these judgments more consistently.

6.4. Forecasting from time series that are subject to sporadic perturbation

We have suggested that the cognitive processes that underlie forecasting from time series that are subject to sporadic perturbation are broadly the same whether or not forecasting is aided by the provision of statistical forecasts. This is particularly true for the two types of statistical forecasts that are in current use: those that ignore whether the periods in the data series are normal or promotional and those that are based only on the normal periods. As Experiment 1 showed, both the anchoring effects and the effects of the proportion of promotions in the data series remained unaffected by either the presence of a statistical forecast or its type when present. This implies that the way in which the judgments were made was the same across all conditions of Experiment 1. Although the provision of statistical forecasts did improve the accuracy, this was because they made the judgment process more consistent, rather than because they changed the nature of the process.

The optimal statistical forecasts provided in Experiment 2 virtually eliminated under-adjustment, but the VE values remained high. Furthermore, they were still affected by the variables that affected VE in Experiment 1. We suspect that similar cognitive processes were responsible for the performances in the two experiments. That is, a mental anchor based on the mean of the data series was first extracted. This process was based on noisier data when the series contained more promotions, which explains the effect of that variable on VE. The optimal forecasts ensured that, on average, the adjustments from this anchor were appropriate. However, VE was still higher when forecasts had to be made for promotional periods, which seems to us to imply that the adjustment process was more complex in promotional periods than in normal ones (because of the additional processing stage involved in extracting and using the relationship between the size of a promotion and its effect). Clearly, optimal statistical forecasts are not accepted automatically: they influence judgment but do not supersede it.

Why did there remain a considerable level of variable error regardless of the presence and type of statistical forecast? Human forecasters introduce inconsistency or random error into forecasts. This error is likely to arise at least in part from the noise that is inherent in cognitive processing. Since the work of Thurstone (1926), it has been

known that judgment contains a random element. When forecasters make a series of judgments about a criterion variable (e.g., the salary levels of a number of different people) from information that they are given about cue variables which are correlated imperfectly with the criterion (e.g., the weight, age, or nationality of those people), the relationship between their judgment and the cues contains a random element (Brehmer, 1978) that decreases with practice and feedback but does not disappear. There are many hypotheses as to why this occurs (Harvey, 1995). For example, Hammond and Summers (1972) referred to a failure of cognitive control: just as hand tremor causes inconsistency in the execution of fine motor skills, so some analogous process affects judgment. The modern computational modelling of cognition is based on the notion that each component process contributes some random error to the total observed in the data (Lewandowsky & Farrell, 2011).

The noise that is inherent in cognitive processing is unlikely to be the only reason for high VE levels. Lawrence et al. (2006, p. 501) suggest that small damaging adjustments of the sort reported by Fildes et al. (2009) may reflect "a tendency to tinker at the edges". In other words, forecasters *intend* to introduce these small changes that do not, overall, lead to (greater) over- or under-forecasting, but do increase the scatter. But why would forecasters do this?

There are various possibilities. One is that the changes that they make form a way of asserting their 'ownership' of the forecasts (Önkal & Gönul, 2005). Another concerns people's responses to automation. Whenever tasks become partially automated, concerns tend to arise among those responsible for performing them that they risk becoming de-skilled (Bainbridge, 1983). In the absence of feedback about the effects of their actions, they will not be able to acquire or maintain the abilities that they need in order to be able to perform their tasks autonomously (something that may be necessary if the automated system suddenly becomes unavailable). Hence, operators may occasionally over-rule or interfere with the output produced by the automatic system in order to ensure that they receive such feedback for forecasters, receiving feedback about statistical forecasts is no substitute for receiving it about forecasts that they have generated themselves: only in the latter case is the rationale for the forecasts known.

6.5. Practical implications

These results have a number of practical implications. First, our main message for practitioners is that the provision of statistical forecasts reduces the forecast error, but whether the data on which those statistical forecasts are based has been cleansed of promotional effects does not matter. This finding could save both time and money, because it implies that cleansing the data, a process that is itself subject to biases (Webby, O'Connor, & Edmundson, 2005), is unnecessary. Even a relatively simple statistical forecast can be of value to a company. Hence, companies that wish to improve their forecasting accuracy but do not currently have a large budget or manpower to spare can still benefit from a simple approach that requires minimal effort.

The second experiment indicates that the forecasting accuracy can benefit greatly from statistical forecasts that incorporate promotion effects. Importantly, this means that practitioners should come to grips with the new developments in forecasting research, as the resulting improved forecasting accuracy can give early adopters a significant competitive advantage. This also has implications for forecast support system developers: regression-based models of promotions (e.g., Huang et al., 2014; Kourentzes & Petropoulos, 2016; Trapero et al., 2013) should be incorporated in future statistical forecasting software.

6.6. Limitations and paths for future research

Our study is subject to various limitations, some of which suggest avenues for future research.

6.6.1. More complex series and other approaches to forecasting

We used relatively simple series and forecast them statistically using an exponential smoothing approach. It is possible that the results might have been different had we used more complex (e.g., seasonal) series or other statistical approaches to forecasting (e.g., ARIMA). While we agree that more appropriate forecasting methods can reduce the bias in the final forecast (as Experiment 2 demonstrated), we have seen that they have little effect on the variable error in the final forecast (again, as Experiment 2 showed). Hence, we expect that our overall conclusions will be generalizable to a range of combinations of different series and forecasting methods that vary in their appropriateness.

Our use of exponential smoothing should ensure the relevance of our findings for practitioners. Many surveys have shown that it is by far the most dominant approach in use by business practitioners. Mentzer and Kahn's (1995) survey of 478 organizations revealed that 92% of them used exponential smoothing. In Sanders and Manrodt's (1994) survey of 96 companies, it was the second most commonly used quantitative technique, after a simple moving average. As Goodwin (2010, p. 33), pointed out: "Fifty years on, researchers are still finding ways to improve the Holt-Winters method and to extend the conditions where it can be applied. This continued interest is a testament to the method's ability to produce reliable forecasts without sacrificing simplicity or transparency".

6.6.2. Promotions with other characteristics

In the future, it would be useful to examine forecasts for promotional periods that are subject to promotion functions with other properties: noise, post-promotion effects, and non-linear promotion functions. First, the promotion function that we used was noise-free, whereas in practice, promotion functions are likely to be noisy. Noisy promotion functions are likely to increase the complexity of the cognitive processes that are needed to identify them, thus increasing the variable error associated with those processes. This would be likely to impair the forecast accuracy further in promotional periods relative to normal ones. Second, in our experiments, promotions had effects only on the periods in which they were applied, whereas in practice, the effects of promotions may extend beyond that. For

example, if they bring forward people's one-off purchases to the promotion period, they will increase sales for that period but decrease them for the following one(s). Again, this is likely to increase the complexity of the cognitive processes that are needed to identify the effects of promotions, making those processes more error prone, and therefore adding to the variable error of forecasts for promotional periods. Finally, the promotion functions considered in the experiments reported here were linear whereas in practice they may be non-linear (e.g., sigmoid in shape). Given that human judges tend to linearize non-linear functions (e.g., Brehmer and Slovic, 1980), sigmoid promotion functions are likely to be associated with an over-estimation of the effects of large promotions. This would make the mean error less negative in those particular cases.

We examined forecasts produced from data series containing either 10% or 40% promotional periods, and saw that a lower frequency of promotions produced a lower overall error on non-promotional periods, but not on promotional ones. This was largely because the over-forecasting on non-promotional periods was lower when there were fewer promotions, which we attributed to the fact that those periods required little adjustment from the anchor (i.e., the mean of the series). What would we expect if we were to examine forecasts using data series containing, say, 90% promotional periods? Would the effect be reversed because little adjustment would be needed for promotional periods but a lot would be needed for non-promotional ones? One could certainly argue that, relative to Experiment 1, ME should be larger for non-promotional periods but smaller for promotional ones, because more adjustment would be needed for the former but less for the latter. However, such a high proportion of promotions would produce a much more variable series anyway, because the promotion size itself varies. This would make extracting the mean of the series to act as the judgment anchor more difficult, and as a result, we could expect an increase in VE for both promotional and non-promotional periods. This could lead to MAE being higher than in Experiment 1 for both types of period.

6.6.3. Other types of forecasters

We used student participants. It could be argued that experts will have more insight into the way in which statistical forecasts should be used, but previous work has shown that experts are subject to similar errors in reasoning as novices. Indeed, in some cases, research has even revealed inverse expertise effects (Önkal & Muradoğlu, 1994; Yates, McDaniel, & Brown, 1991). Advice discounting (ignoring or under-weighting the 'advice' of the statistical forecast) may be even greater in experts, because they value their own opinions even more than novices do theirs. In fact, Önkal and Muradoğlu (1994) demonstrated that experts exhibited even more over-confidence in their forecasts than those who were less expert. This situation is typical of what happens when experience at a task (e.g., forecasting) fails to produce learning as quickly as people expect it to Harvey and Fischer (2005).

An experiment is a simulation or model of a task that is performed by practitioners. As with any model, some features of the real world task are excluded. Thus, we do not expect to see all characteristics of practitioner performances reflected in our experimental results. Previous analyses of data obtained from a range of organizations have revealed that forecasters are often subject to optimism effects: inappropriate upward adjustments of statistical forecasts are greater or made more often than inappropriate downward ones (e.g., Fildes et al., 2009). We did not observe any such optimism in our experiments, as they were not designed to study or reveal it. However, it could be argued that optimism would have produced less under-forecasting on promotional periods than over-forecasting on normal periods. We did not find this pattern in the data. However, this prediction does not compare like with like. As we have emphasized, the processes that underlie forecasting in promotional periods are different from those that underlie it in normal ones. To research optimism experimentally, further studies should be specifically designed with that aim in mind. One possible approach is to compare two groups who are performing exactly the same forecasting task but where the variable being forecast is labelled 'profits' in one case but 'losses' in the other. We would expect the forecasts to be systematically higher in the former case (Harvey & Reimers, 2013).

6.6.4. Increasing the acceptability of statistical forecasts

An additional avenue for further research is indicated by the results of the second experiment, which demonstrated that highly sophisticated statistical forecasts that take the effects of promotions into account explicitly benefit forecasters considerably more than those that do not. Further research efforts that aim to develop ways of producing such forecasts (e.g., Huang et al., 2014; Kourentzes & Petropoulos, 2016; Trapero et al., 2013) are clearly worthwhile. However, this second experiment also showed that even forecasters who are given optimal statistical forecasts still make adjustments that impair the accuracy. As we have seen, there are various different possible explanations for this finding, but they all imply that, for one reason or another, forecasters are not good at taking 'advice' from a statistical model. Such a discounting of advice has been reported before (e.g., Goodwin, 2000; Lim & O'Connor, 1995), and the factors that have been proposed to account for it include concerns about the credibility of a statistical model rather than a human being as a source of advice (Önkal, Gönül, & Lawrence, 2008; Önkal et al., 2009), and people's beliefs that their own opinions are better founded than those of others (Harvey & Harries, 2004).

The prevention of damaging adjustments has long been an important topic in judgmental forecasting. Goodwin et al. (2011) found that neither restrictions nor guidance improved the accuracy. Indeed, guidance was met with resistance on the part of forecasters. Such resistance is consistent with Bainbridge's (1983) views about responses to automation. However, as we have pointed out, the reasons why forecasters make damaging adjustments may not be purely volitional (i.e., arising because, for one reason or another, they *want* to make those adjustments), but may also be at least partly cognitive (i.e., noise may be inherent in the cognitive processes that underlie forecasting). Further research into facilitating the acceptance of statistical forecasts is needed.

6.7. Conclusions

The provision of statistical forecasts, even crude ones, can improve the forecasting accuracy by reducing the variable error. When forecasts are made from time series that are perturbed by sporadic exogenous events, the effort required to produce forecasts that are cleansed of their effects does not appear to be warranted. However, the current attempts to develop methods of incorporating the effects of these events into statistical forecasts are worthwhile, and are likely to result in improvements in the forecast accuracy.

Acknowledgments

Appendix. Instruction sheets for Experiment 1

Group A (unaided judgment, no statistical forecast) were given the following text on an instruction sheet: "*Please read this document carefully before you start with the first graph! In this experiment, you will receive a number of graphs such as the one depicted below. On the X-axis, you will find the time period, ranging from 0 to 55. On the Y-axis, you will find the sales number, ranging from 0 to 500. The grey line indicates the sales data of a product in the past 50 time periods. The blue bars[2] indicate the promotional investment (e.g., an advertisement campaign) made for that product. The number of promotions can vary: some graphs will have 5 promotions, others will have 20. It is your job to predict the sales number of the following two time periods (51 and 52), as accurately as possible. Pay attention, because sometimes there is a promotion present and sometimes there isn't. You can make your prediction by clicking with your mouse on the graph. An information box with your mouse's location appears next to your cursor. First click on your prediction for time period 51 and only then for time period 52. Afterwards, a box 'next graph' will appear on the bottom of the page*".

Participants in group B (statistical forecast based on cleansed series) saw the following additional text: "*The orange line[2] is a forecast from a statistical model. The model is based on the cleaned sales data: the promotion effects have been taken out of the data until only the baseline remained. The model uses these baseline data to produce the statistical forecasts. You can see the predictions it made in the past and what it predicts for time period 51 and 52. You can choose whether or not to follow the statistical forecast*".

For those in group C (statistical forecast based on non-cleansed data), the additional text was as follows: "*The orange line is a forecast from a statistical model. We have fed the sales data to a statistical model. You can see the predictions it made in the past and what it predicts for time period 51 and 52. This statistical model is a simple model that ignores whether or not a promotion took place: it is just based on the value of the sales figures. You can choose whether or not to follow the statistical forecast*".

References

Alvarado-Valencia, J., & Barrero, L. H. (2014). Reliance, trust and heuristics in judgmental forecasting. *Computers in Human Behavior*, *36*, 102–113.

Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (2001). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners* (pp. 417–439). New York: Kluwer.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*, 775–779.

Baird, J. C., & Noma, E. (1978). *Fundamentals of scaling and psychophysics*. New York: Wiley.

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

Bovi, M. (2009). Economic versus psychological forecasting. Evidence from consumer confidence surveys. *Journal of Economic Psychology*, *30*, 563–574.

Brehmer, B. (1978). Response consistency in probabilistic inference tasks. *Organizational Behavior and Human Decision Processes*, *22*, 103–115.

Brehmer, B., & Slovic, P. (1980). Information integration in multiple-cue judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 302–308.

Durand, R. (2003). Predicting a firm's forecasting ability: the roles of organizational illusion of control and organizational attention. *Strategic Management Journal*, *24*, 821–838.

Fechner, G. T. (1860). *Elemente der psychophysik [Elements of psychophysics] (Vol. 1)*. Leipzig: Breitkopf und Harterl.

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. (2013). Forecasting support systems: What we know, what we need to know. *International Journal of Forecasting*, *29*, 290–294.

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., & Petropoulos, F. (2015). Improving forecast quality in practice. *Foresight: The International Journal of Applied Forecasting*, *36*, 5–12.

Frances, P. F., & Legerstee, R. (2009). Properties of expert adjustments on model-based SKU-level forecasts. *International Journal of Forecasting*, *25*, 35–47.

Gardner, D. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology*, *9*, 555–579.

Gönül, M. S.,Önkal, D., & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems*, *42*, 1481–1493.

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

Goodwin, P. (2010). The Holt-Winters approach to exponential smoothing: 50 years old and still going strong. *Foresight*, *19*, 30–34.

Goodwin, P., & Fildes, R. (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., Fildes, R., Lawrence, M., & Nikolopoulos, K. (2007). The process of using a forecasting support system. *International Journal of Forecasting*, *23*, 391–404.

Goodwin, P., Fildes, R., Lawrence, M., & Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega: The International Journal of Management Science*, *39*, 242–253.

Hammond, K. R., & Summers, D. A. (1972). Cognitive control. *Psychological Review*, *79*, 58–67.

Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, *63*, 247–263.

Harvey, N., & Fischer, I. (2005). Development of experience-based judgment and decision-making: The role of outcome feedback. In T. Betsch, & S. Haberstroh (Eds.), *The routines of decision-making* (pp. 119–137). Mahwah, NJ: Erlbaum.

Harvey, N., & Harries, C. (2004). Effects of judges' forecasting on their later combination of forecasts for the same outcomes. *International Journal of Forecasting*, *20*, 391–409.

Harvey, N., & Reimers, S. (2013). Trend damping: under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology*, *39*, 589–607.

Hilary, G., & Hsu, C. (2011). Endogenous overconfidence in managerial forecasts. *Journal of Accounting and Economics*, *51*, 300–313.

Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, *237*, 738–748.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

Kotteman, J. E., Davis, F. D., & Remus, W. (1994). Computer-assisted decision making: performance, beliefs, and the illusion of control. *Organizational Behavior and Human Decision Processes*, *57*, 26–37.

Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*, *181*, 145–153.

Lawrence, M. (2000). Editorial: What does it take to achieve adoption in sales forecasting? *International Journal of Forecasting*, *16*, 147–148.

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

Lawrence, M., & O'Connor, M. (1992). Exploring judgmental forecasting. *International Journal of Forecasting*, *8*, 15–26.

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. London: Sage.

Libby, R., & Rennekamp, K. (2012). Self-serving attribution bias, overconfidence, and the issuance of management forecasts. *Journal of Accounting Research*, *50*, 197–231.

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

Lim, J. S., & O'Connor, M. (1996). Judgmental forecasting with time series and causal information. *International Journal of Forecasting*, *12*, 139–153.

Mentzer, J. T., & Kahn, K. (1995). Forecasting technique familiarity, satisfaction, usage and application. *Journal of Forecasting*, *14*, 465–476.

Önkal, D., & Gönul, M. S. (2005). Judgmental adjustment: A challenge for providers and users of forecasts. *Foresight: The International Journal of Applied Forecasting*, *1*, 13–17.

Önkal, D., Gönul, S., & Lawrence, M. (2008). Judgmental adjustments of previously adjusted forecasts. *Decision Sciences*, *39*, 213–238.

Önkal, D., Goodwin, P., Thomson, M., Gönul, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*, 390–409.

Önkal, D., & Muradoğlu, . (1994). Evaluating probabilistic forecasts of stock prices in a developing stock market. *European Journal of Operational Research*, *74*, 350–358.

Önkal, D., Sayim, K. Z., & Lawrence, M. (2012). Wisdom of group forecasts: Does role-playing play a role? *Omega*, *40*, 693–702.

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

Sanders, N. R., & Manrodt, K. B. (1994). Forecasting practices in US corporations: Survey results. *Interfaces*, *24*, 92–100.

Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, *31*, 511–522.

Syntetos, A. A., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting*, *26*, 134–143.

Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., & Goodwin, P. (2009). The effects of integrating management judgment into intermittent demand forecasts. *International Journal of Production Economics*, *118*, 72–81.

Theocharis, Z., & Harvey, N. (2016). Order effects in judgmental forecasting. *International Journal of Forecasting*, *32*, 44–60.

Thurstone, L. L. (1926). The scoring of individual performance. *Journal of Educational Psychology*, *17*, 446–457.

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.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

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

Weber, E. H. (1834). *De pulsu, resorptione, auditu et tactu [On stimulation, response, hearing and touch]*. Leipzig: Koehler.

Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational Behavior and Human Decision Processes*, *40*, 60–79.

Shari De Baets is attached to the Department of People and Organization at Vlerick Business School and was recently awarded her Ph.D. in the Department of Applied Economics at the University of Ghent. Her research focusses on the role of judgment in the provision of forecasting support for practitioners.

Nigel Harvey is Professor of Judgment and Decision Research in the Department of Experimental Psychology at University College London and Visiting Fellow in the Department of Statistics at the London School of Economics and Political Science. His research focusses on the role of judgment in the forecasting, monitoring and control of time series.