Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour?

Fotios Petropoulos[a,*], Robert Fildes[b], Paul Goodwin[c]

[a]*Logistics & Operations Management Section, Cardiff Business School, Cardiff University, UK*
[b]*Department of Management Science, Lancaster University Management School, Lancaster University, UK*
[c]*School of Management, University of Bath, UK*

Abstract

The behaviour of poker players and sports gamblers has been shown to change after winning or losing a significant amount of money on a single hand. In this paper, we explore whether there are changes in experts' behaviour when performing judgmental adjustments to statistical forecasts and, in particular, examine the impact of 'big losses'. We define a big loss as a judgmental adjustment that significantly decreases the forecasting accuracy compared to the baseline statistical forecast. In essence, big losses are directly linked with wrong direction or highly overshooting judgmental overrides. Using relevant behavioural theories, we empirically examine the effect of such big losses on subsequent judgmental adjustments exploiting a large multinational data set containing statistical forecasts of demand for pharmaceutical products, expert adjustments and actual sales. We then discuss the implications of our findings for the effective design of forecasting support systems, focusing on the aspects of guidance and restrictiveness.

Keywords: forecasting, judgment, behavioural analytics, decision support systems

1. Introduction

Accurate product demand forecasting is important to companies as forecasts are used in decisions relating to inventory control, production planning, purchasing, logistics, cash flow planning and other aspects of the business. A typical forecasting process includes pre-processing and analysis the data, which are usually in the form of time series, extrapolating the series with a suitable statistical method (Petropoulos et al., 2014), post-processing the statistical forecasts and monitoring and evaluating the outputs. The latter acts as feedback to inform the calculation of subsequent sets of forecasts. Often, the forecasting process is implemented within a specialised forecasting software. This paper focuses on the third, post-processing, stage of the forecasting process, and more specifically on the judgmental

*Correspondance: Fotios Petropoulos, Cardiff Business School, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK.

Email address: PetropoulosF@cardiff.ac.uk (Fotios Petropoulos)

interventions on statistical forecasts that are typically performed by demand planners and managers (Fildes and Goodwin, 2007b). Such interventions are common, with Fildes et al. (2009) reporting that 91% of the forecasts examined in one organisation were subject to judgmental adjustments. Franses and Legerstee (2009) reported similar findings.

Human adjustments of the outputs of standard forecasting methods, like exponential smoothing or ARIMA models, are primarily made for four reasons. First, managers attempt to incorporate into the forecasts the expected impact of forthcoming special events, such as promotional activities, strikes, or the launch of a competiting new product (Fildes and Goodwin, 2007a). Arguably, more formal methods that include external regressors could sometimes be used for this purpose (Huang et al., 2014). However, limitations in the available quantitative data and the complexity of the models often renders judgmental adjustment as the only practical approach. Second, demand planners may tend to change statistical forecasts in order to be in-line with budgeting or politically-related targets set by senior managers (Fildes and Goodwin, 2007b). For example, in a field study Lawrence et al. (2000) questioned whether forecast accuracy was the primary objective of their company-based forecasters and suggested that their forecasts were heavily influenced by political choices within the company framework. Despite this, a recent survey by Fildes and Petropoulos (2015) showed that accuracy is generally the most important driver in the forecasting process, confirming earlier studies (for example see: McCarthy et al., 2006). Third, managers may adjust in order to gain a sense of ownership of the forecasts, possibly as a result of a lack of trust in the statistical methods, which they may regard as "black-boxes" (Önkal and Gönül, 2005). Lastly, humans are liable to confuse the signal with the noise (Harvey, 1995) and introduce unnecessary judgmental adjustments as the result of perceived systematic changes that were not captured by the statistical methods (Goodwin and Fildes, 1999).

Previous studies of demand forecasting have focused on the efficiency of judgmental adjustments and the circumstances under which judgmental manipulation of statistical forecasts might be useful. Some studies proposed actions and strategies to prevent unnecessary interventions or to optimally combine statistics with judgment. See Lawrence et al. (2006) and Leitner and Leopold-Wildburger (2011) for reviews of progress in judgmental forecasting. Recently, researchers in behavioural operational research (Hämäläinen et al., 2013) have focused on finding links between forecasting performance and experts' behaviour (for example see: de Bruijn and Franses, 2012). Similarly, research on corporate earnings forecasting has examined the behavioural determinants of observed biases (Ramnath et al., 2008). However, while demand forecasters usually have many products to forecast and obtain rapid feedback on accuracy, earnings analysts in contrast tend to focus more intensively on particular companies and observe the outcomes of their forecasts less frequently.

While a number of factors may affect forecasters' behaviour, the occurrence of a significant event or outcome in the previous period may be particularly influential. In an interesting study by Smith et al. (2009), poker players were found to change their strategy after significant wins or losses. Big losses were followed by playing less cautiously, with players tending to be more aggressive compared to their behaviour after big wins. Similar behaviours have been found in sports gambling (Xu and Harvey, 2014) and in financial markets (Coval and Shumway, 2005; Garvey et al., 2007). Here we investigate whether forecasters' behaviour in relation to judgmental adjustments is affected by the experience of previous poor interventions and, if the effect is damaging to accuracy, how this might be mitigated. Specifically, we address the following research questions:

RQ1 How do adjustments to statistical forecasts that lead to large errors affect experts' behaviour in performing interventions for the very next period?

RQ2 If judgmental adjustments are unduly influenced by large errors in the previous period what corrective actions would be likely to result in improved forecasting performance?

In order to deal with these two questions, we have to define what a big loss in a judgmentally adjusted forecast is. So, after a review of the background literature in section 2, in section 3 we propose a new way for classifying and measuring the quality of judgmental adjustments. Sections 4 and 5 attempt to answer the research questions by analysing a large empirical data set of judgmental adjustments to demand forecasts made by managers in a multinational company. Finally, the last section summarises the findings, offering conclusions as to their managerial implications.

2. Background literature

A few decades ago, most researchers discouraged managers from making judgmental adjustments to statistical forecasts because it was believed that they would generally damage accuracy (Carbone et al., 1983; Armstrong, 1985). These researchers found evidence that judgment was associated with a wide range of biases including over-optimism, anchoring (Eroglu and Croxton, 2010), overconfidence (Blattberg and Hoch, 1990; Kottemann et al., 1994), inconsistency, and confusion of the signal with the noise (Eggleton, 1982; O'Connor et al., 1993).

Mathews and Diamantopoulos were the first to show empirically through a series of company-based studies (Mathews and Diamantopoulos, 1986, 1989, 1990) that "forecast manipulation" can lead to improvements in accuracy. Interestingly, they showed that forecasters are more likely to adjust the forecasts that would have produced the largest forecast errors had the statistical forecasts remained unrevised. Other researchers have provided further evidence on the efficacy of judgmentally adjusted forecasts in economic (McNees, 1990; Turner, 1990; Donihue, 1993), accounting earnings (Brown, 1988) and business forecasting (Vere and Griffiths, 1995; Wolfe and Flores, 1990). Syntetos et al. (2010) showed that in addition to the improvements in performance as measured by traditional error metrics, judgmental adjustments of demand forecasts also result in significant reductions in inventory costs. The common factor in these studies is that when important domain knowledge is missing from the statistical forecasts, this can be integrated efficiently into the operational forecasts by applying judgmental adjustments to improve performance. However, two key elements affecting the success of an intervention are the reliability and importance of the missing information (Goodwin and Fildes, 1999) and the requirement that humans should not discount reliable statistical forecasts (Donihue, 1993).

Despite these findings, there is considerable evidence from both field and laboratory studies that relatively accurate statistical forecasts are frequently judgmentally adjusted without reference to domain knowledge or its reliability (Fildes et al., 2009; Goodwin, 2000). A particularly salient cue that the forecasters are likely to be prompted with is the latest error and this raises the question: to what extent are such adjustments a behavioural response to an error resulting from a judgmental adjustment in the previous period, and in particular, a large error, as this is likely to be especially prominent? For example, do forecasters have a propensity to make a large adjustment after a previous adjustment has led to a large error, even when they have no reliable domain knowledge to justify such an intervention?

The literature suggests a number of possible behavioural reactions to an adjustment in the previous period that has led to a large error. There are two reasons why the subsequent adjustment might be large: *a*) greater risk taking behaviour by the forecaster and *b*) over-reaction to outcome feedback. Smith et al. (2009) found that, after large losses in games of poker, players engaged in more aggressive and riskier gambles (see also Xu and Harvey, 2014). They largely attributed this to the break-even hypothesis whereby, after sustaining a large loss, the players were prepared to take risks in an attempt to cancel out the loss. In demand forecasting a large adjustment may be a sign of risk taking behaviour. To make a large adjustment following a previous damaging intervention may be a brave action that risks further compounding both financial costs and damage to the forecaster's reputation. Significantly, it involves an act of commission. While, not making an adjustment when it was warranted would risk one being guilty of an act of omission, there is evidence that an erroneous act of commission is seen as worse than an erroneous act of omission (Spranca et al., 1991).

However, decisions in poker games differ from forecasting judgments in several ways so the break-even hypothesis might not apply in the forecasting context. First the concept of losses and gains differs between the two concepts. In the context of demand forecasting we define a gain as a improvement in accuracy as a result of an adjustment, while a loss is a reduction in accuracy. Thus large losses are to be distinguished from large forecast errors. A forecaster's adjustment may actually lead to a gain in accuracy compared to the statistical forecast, but a large error may still result. Second, the outcomes of poker games are independent while observations in time series are usually dependent. Intervals between poker games are likely to be shorter than periods between successive demand forecasts so that immediate emotional reactions to poor judgments are likely to be less prevalent in demand forecasting. Also, traditional poker players are likely to be engaged in one game at a time while most demand forecasters will have the task of forecasting many series (Fildes and Goodwin, 2007b) over long periods so that a large forecast error in one series at one point in time will be less prominent than the consequences of a poor judgment in poker.

In particular, errors in forecasts differ from financial losses and gains so an adjustment that significantly improves accuracy will not necessarily compensate for a preceding intervention that reduced accuracy. For example, experts' reputations are more easily lost than gained (Bonaccio and Dalal, 2006). While prospect theory (Tversky and Kahneman, 1992) suggests that people do tend to risk even further losses to try to negate a current loss, the idea of forecasters making a large reckless adjustment to a forecast for an individual product in order to recover their reputation, because their previous adjustment had significantly damaged accuracy, seems less plausible. Hence, demand forecasting seems unlikely to be associated with the break-even hypothesis.

Nevertheless, there are still reasons to believe that forecasters may tend to make large adjustments following large errors and these reasons are related to the direction of adjustment. First, it is known that forecasters have a tendency to overreact to outcome feedback, which will, in part, reflect the noise in a time series (Lawrence et al., 2006). For example, an outcome that is significantly higher than a forecast may be interpreted as a sign that an upward movement in the signal has occurred even when much of the error can be attributed to noise. As a consequence, the subsequent forecast may be subject to considerable upwards adjustment. This causes it to be too high and a large error in one direction is followed by a large error in the opposite direction. Where the initial large error largely results from the forecaster's adjustment then the forecaster will incur a loss (as defined above). Assuming that forecasters receive outcome feedback on the success or otherwise of their adjustments then large adjustments following large losses would be manifested in a tendency for large adjustments to be made in the same direction as the previous large error so that a positive adjustment will follow a positive error (where error = actual − forecast) and vice versa.

Alternatively, forecasters may persist in making large adjustments in the opposite direction to the previous large error. For example, a significant upwards adjustment, resulting in a negative error and a large loss may still be followed by a large upwards adjustment. This behaviour is likely to incur similarly large errors in the same direction. This may in part relate to the gambler's fallacy where chance events are perceived to be self-correcting (Smith et al., 2009). For example, a forecaster might expect that a run of lower than expected sales figures that are judged to be due to random factors will be balanced in the future by higher sales because 'on average half the sales are lower and half are higher than expected'. The probable result would be a persistent tendency to over-forecast as the 'compensating' higher-than-expected sales are awaited.

However, there may be other reasons for the persistency of large errors of the same sign after big losses including an unforeseen delay in an expected special event and, when decisions are associated with asymmetric loss, a confusion of forecasts with decisions. For example, this may occur when, in order to meet customer service targets, decisions are made to hold inventory at two standard deviations above expected sales but these decisions are represented as forecasts of expected demand (Fildes et al., 2009). There is also evidence that people are reluctant to modify a previous act of commission and persist in pursuing the same action (Staw, 1976). This can occur despite evidence that continuing the action is counterproductive (Lim and O'Connor, 1995). Persistent errors of the same sign represent a rejection of outcome feedback and can result from a belief that what happened in the last period is irrelevant.

Of course, there are reasons why big losses might tend to be followed by relatively small adjustments. Following the poker analogy (Smith et al., 2009), a large loss may have negative effects on a forecaster's confidence in his or her ability to contribute to forecast accuracy for a given product. This would also lead to a propensity to avoid a large adjustment in the subsequent period. In other cases a big loss in the previous period may have been associated with an inability to forecast the effects of a special event. In the subsequent period, when no special event is anticipated, an adjustment might not be considered to be necessary.

On balance, the literature suggests the following hypothesis:

H1 Experts are more likely to make large judgmental adjustments to forecasts in periods following big losses.

H1 implies that forecasters are likely to pay attention to the latest loss and this is also fairly supported by responses to a questionnaire administered by Boulaksil and Franses (2009). Here the experts of the pharmaceutical company (on which our later analysis is based) indicated that they review their past forecasting performance when making new forecasts. Also, there is some evidence that these experts compare the performance of the statistical forecasts with that of their own forecasts. As we have seen, outcome feedback will be expected to cause forecasters to adjust in the direction suggested by their previous error. Hence, we have:

H2 Following a big loss, experts are more likely to make adjustments in same direction as the previous large error.

The preceding hypotheses suggest that, following a large loss, large adjustments will be made based largely on the basis of outcome feedback. Since this feedback relates only to the latest period and is contaminated by noise it is an unreliable basis for these large adjustments which are therefore likely to be seriously detrimental to forecast accuracy. Hence we hypothesise:

H3 Big losses are more likely in the period following a big loss.

3. A new measure for understanding judgmental adjustments

3.1. Types of judgmental adjustments and their effects on accuracy

Generally, the effects of judgmental adjustments can be divided into three types, graphically depicted in figure 1. Wrong direction adjustments are interventions in the opposite direction compared to the sign of the deviation between real outcome and statistical forecast. These adjustments always lead to inferior accuracy in the final forecast compared to the statistical forecast. Undershoots refer to revisions that are to the correct direction, but not enough to fully explain the true outcome. Despite that, undershoots always improve forecast accuracy, as adjustments of this type decrease the difference between statistics and reality. Lastly, overshoots are interventions to the correct direction, but of magnitude larger than the 'optimal'. Overshoots may lead to either improvements or deterioration in forecasting performance, depending on the magnitude of the adjustment.

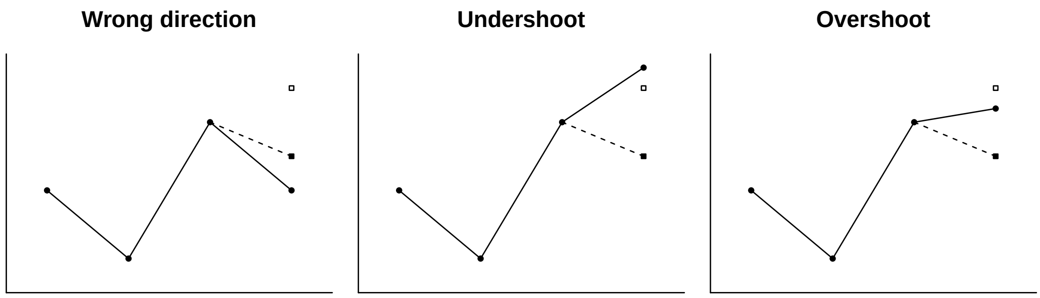

Figure 1: Types of adjustments in judgmental forecasting (solid line: actual outcome; dash line, black square: statistical forecast; unfilled square: expert forecast).

3.2. The β coefficient in judgmental adjustments

In this study we are interested in analysing the behaviour of experts in performing adjustments directly after revisions that resulted in big losses. So, we have to first answer the question 'what is a big loss?'. However, to the best of our knowledge the definition of 'big losses' is absent from the literature. Arguably, it could be linked to wrong direction and overshoot adjustments, but a non-arbitrary quantitative measure of the type, quality and magnitude of a single judgmental adjustment is needed. In this section, we define a new measure for understanding judgmental revisions of statistical forecasts. This new measure enables us to analyse the behaviour of experts when performing judgmental adjustments, focusing on the cases after big losses.

Let us assume that the statistical output of a forecasting method is unbiased. This means that the cumulative signed forecast error over a large number of periods is zero. In other words, any optimistic statistical point forecasts are balanced off by other pessimistic ones, and vice versa. Let us also assume that there is a deviation between the true outcome and the statistical prediction. In essence, this deviation, or statistical forecast error, is to be reconciled by an ideal judgmental adjustment. In other words, the aim of a judgmental adjustment is to alter the statistical output by the construction of an expert forecast which will be closer or even equal to the actual value. Given the aforementioned assumptions, we can regard the quality of a judgmental adjustment as a percentage of the deviation between the statistical forecast and the actual outcome.

We define a scale-free measure for identifying the type, quality, and magnitude of a judgmental adjustment:

$$\beta_t = \frac{EF_t - SF_t}{Y_t - SF_t} = \frac{FD_t}{RD_t} \tag{1}$$

where:

- Y_t: actual value at time t.
- SF_t: statistical forecast at time t.

- EF_t: expert forecast at time t produced given the statistical baseline (SF_t). It may be equal to SF_t. Most usually, this is used as the final (operational) forecast.
- RD_t: *real difference* that needs to be reconciled or difference between actual and statistical forecast ($Y_t - SF_t$).
- FD_t: *forecasts' difference* or difference between expert forecast and statistical forecast ($EF_t - SF_t$). This is the actual judgmental adjustment.

This measure gives the signed ratio of the judgmental adjustment to the error in the statistical forecast. A positive sign denotes an adjustment in the correct direction, while negative β values refer to wrong direction adjustments. The interpretation of this measure is very intuitive. For example, $\beta = 0.5$ means that only 50% of the statistical forecast error has been removed by the judgmental adjustment, $\beta = 1$ refers to a perfect adjustment (100% of the statistical forecast error is removed by judgment), while $\beta = 1.5$ indicates that judgment is over-compensating (by 50%) for the statistical forecast's error. A linkage of β values with different types of adjustments is provided in table 1. We also provide a translation of the different values of β with the effect of the adjustment on forecast accuracy, when compared to no adjustment.

While some of the critical values derive directly from the definition of this new measure, the values of β for which an adjustment is translated to a 'big loss' rather than just a loss may differ in various applications. However, we opt for retaining the symmetry of the critical values and therefore propose that a big loss may be regarded as the result of an adjustment that deviates by more than 200% from a perfect adjustment (i.e. has a β value of less than -1 or greater than 3).

Table 1: Linkage of β values with different types of adjustments.

Type of adjustment	Value of β	Effect on accuracy, compared to no adjustment
XL overshoot	$\beta > 3$	big loss
L overshoot	$2 < \beta \leq 3$	loss
overshoot	$\beta = 2$	no gain nor loss
overshoot	$1 < \beta < 2$	gain
spot-on	$\beta = 1$	maximum gain
undershoot	$0 < \beta < 1$	gain
no adjustment	$\beta = 0$	no gain nor loss
wrong direction	$-1 \leq \beta < 0$	loss
L wrong direction	$\beta < -1$	big loss

A limitation of this measure is that is does not distinguish between upwards and downwards adjustments, which proved to be of some importance in other studies (Fildes et al., 2009; Trapero et al., 2013). A further limitation is that β is undefined in the extreme case that the statistical forecast coincides with the actual value (the error of the statistical forecast is zero). In this case, we argue that a no-adjustment is the optimal behaviour (maximum gain), resulting in a β coefficient of 1. If an adjustment has been made, then the β coefficient will be infinite, denoting that this adjustment significantly deteriorates the accuracy (compared to the statistical forecast), so it is a big loss.

3.3. Links to the literature

Let us now see how this new measure links to the existing literature on judgmental adjustments. Franses and Legerstee (2011a) examine the effectiveness of linearly combining the statistical and expert forecasts. So, they suggest that a final forecast at time t (FF_t) may be derived as:

$$\begin{aligned} FF_t &= \alpha_t EF_t + (1-\alpha_t) SF_t \Leftrightarrow & (2)\\ FF_t &= SF_t + \alpha_t (EF_t - SF_t) \Leftrightarrow & (3)\\ FF_t &= SF_t + \alpha_t FD_t \Leftrightarrow & (4)\\ Y_t - FF_t &= Y_t - SF_t - \alpha_t FD_t & (5)\end{aligned}$$

where a is the weight to be assigned on the expert forecast at time t. Obviously, the weight to be assigned on the statistical forecast should be $1 - a$, so that the summation of the two weights is unity.

Letting $Y_t - FF_t$ being the forecast error at time t (e_t), equation 5 gives:

$$e_t = RD_t - \alpha_t FD_t \tag{6}$$

From equation 6, for $e_t = 0$:

$$\alpha_t = \frac{RD_t}{FD_t} = \frac{1}{\beta_t} \tag{7}$$

So, the optimal weights for combining the statistical forecast (SF_t) and the expert forecast (EF_t) in order to end up with a zero forecast error are $1 - \frac{1}{\beta_t}$ and $\frac{1}{\beta_t}$ respectively. For example, if β_t reflects a situation where only 0.4 of a required upwards adjustment has been made, a weighted average of the statistical and expert forecast using respective weights of -1.5 and 2.5 would yield a perfectly accurate forecast. It is worth mentioning that the optimal weights, along with the β values, are most likely to change over time.

Hyndman and Koehler (2006) define the relative error as the ratio of the forecast error deriving from a method whose performance is to be measured divided by the error of a benchmark method. For example, the relative absolute error (RAE) incurred for a statistical forecast, SF_t, for time t can be defined as:

$$RAE_t = \frac{|Y_t - SF_t|}{\left|Y_t - SF_t^b\right|} = \frac{|e_t|}{\left|e_t^b\right|} \tag{8}$$

where SF_t^b and e_t^b refer to the statistical forecast and the respective forecast error of the benchmark method. Following Davydenko and Fildes (2013), we can replace the benchmark method in equation 8 with the pure statistical forecast and the method in the numerator with the expert forecast (the one containing the judgmental adjustment). By doing this, we can directly compare the performance of the expert forecast relatively to the statistical forecast:

$$RAE_t = \frac{|e_t^{EF}|}{|e_t^{SF}|} = \left|\frac{Y_t - EF_t}{Y_t - SF_t}\right| \tag{9}$$

However, equation 9 can be rewritten as:

$$RAE_t = \left|\frac{Y_t - SF_t - (EF_t - SF_t)}{Y_t - SF_t}\right| = \left|1 - \frac{FD_t}{RD_t}\right| = |1 - \beta_t| \tag{10}$$

So, the β is also linked with the relative absolute error of the expert forecast, using the statistical forecast as the benchmark. This is an important property which provides a direct relationship of the β with a recently introduced error measure (Average Relative Mean Absolute Error or AvgRelMAE, Davydenko and Fildes, 2013) for evaluating judgmentally adjusted forecasts across different periods and multiple time series.

4. Analysing expert forecasts

4.1. The data

In order to examine the behaviour of experts in performing judgmental adjustments after big losses, we consider a database that was initially introduced in a study by Franses and Legerstee (2009) and was afterwards used in other studies by the same researchers (for example see Franses and Legerstee, 2010, 2011a,b, 2013; Legerstee and Franses, 2014). This database contains the monthly sales of 1,101 pharmaceutical SKUs. The SKUs come from 37 countries and were the responsibility of 50 different managers.

The length of each series is 25 months, spanning from October 2004 to October 2006. Besides the actual sales (Y), in each period the database also contains the statistical (SF) and expert forecast (EF). However, missing values exist for some periods in specific SKUs. We focus on the 774 time series where the triplet Y, SF, and EF is available for all periods. SF is automatically provided by some forecasting software which utilises historical information (lagged sales) and individually (per series) select an optimal method from a set of alternatives (such as Box-Jenkins or Holt-Winters). The method itself and the optimised parameters may change across origins. For more details, please see Franses and Legerstee (2009, 2013).

A typical time series from the database is presented in figure 2. This shows all different types of judgmental adjustments. For example, undershoots occurred at periods 7 and 17, overshoots are observed at periods 4 and 6, while wrong-direction adjustments are recorded for periods 3 and 5.

4.2. Analysis of all judgmental adjustments

As the target is to identify the effect of big losses in the very next judgmental adjustment, we first calculate the percentage of judgmental adjustments of each type identified in table 1. The analysis is performed for the periods $t = 2, 3, ...25$, leaving out of the observations

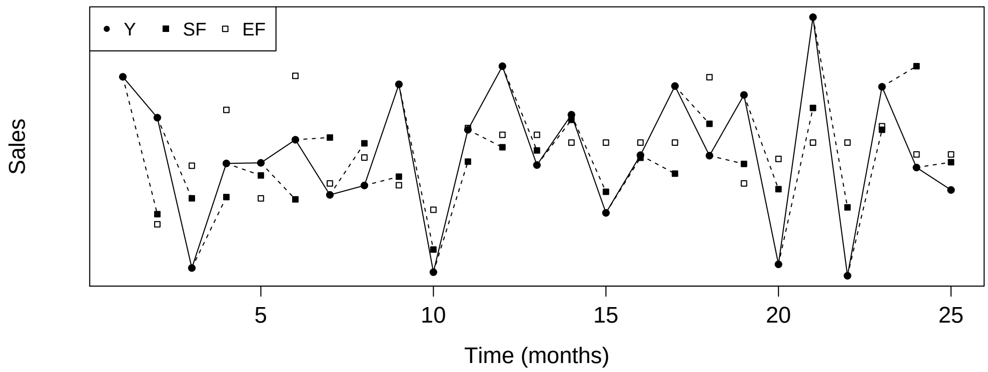

Figure 2: Example of a time series and the respective information contained in the database.

for the very first period. This is because periods with lag one will be used later to identify periods where judgmental adjustments occurred after big losses. To simplify the analysis, we excluded the limited number of cases where $\beta = 0$ (2% of the total cases), indicating that no adjustments were made despite any deviations between the actual values and statistical forecasts. So, the effective sample for the current analysis contains more than 18,000 judgmental adjustments (774 time series $\times$ 24 periods $-$ 384 cases where $\beta = 0$). The cases where $\beta \in (0, 1]$ are pooled together, creating the "undershoot or spot-on" group. Similarly, for $\beta \in (1, 2]$ in the case of small overshoots.

The relative frequency (percentage of cases) of each type of adjustment is depicted in figure 3. Undershoots (and spot-on) are the most common type of adjustments (36% of the cases), with small wrong direction adjustments being the second most common. These two categories together constitute 57% of the adjustments. Thus, managers have a tendency to perform relatively small adjustments. So, in the majority of cases, the absolute magnitude of the interventions is not enough to remove the difference between the actual outcome and the statistical forecast. This is a result that is consistent with findings in other studies (Fildes et al., 2009).

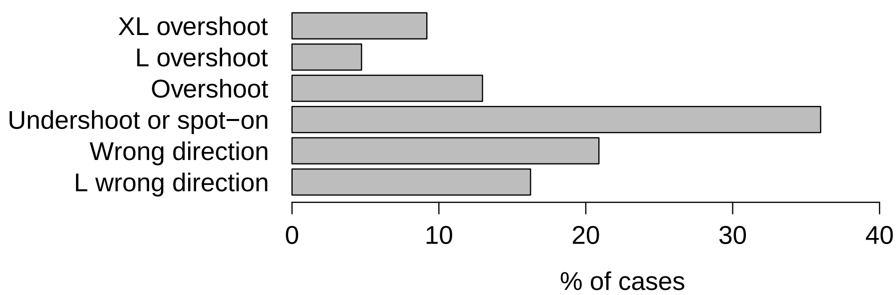

Figure 3: Analysis of all judgmental adjustments.

Only 49% of the adjustments lead to improvements in accuracy (i.e. they are undershoots, spot-on adjustments or small overshoots). This is in line with previous studies on the same database, where it was found that, when averaged across countries and categories of products, only in 43% of the cases expert forecasts were better than statistical ones (Franses and Legerstee, 2010). Hence, for more than half of the cases adjustments to the statistical forecasts result in deterioration in accuracy. On top of that, in 25.4% of cases the adjustments led to big losses (i.e. βs had values lower than -1 or greater than 3).

4.3. Analysis of judgmental adjustments after big losses

We first test H1 and investigate whether there is an association between the size of adjustment in a given period and whether or not a big loss occurred in the previous period. To control for different levels of volatility in the series we divided each absolute adjustment by the standard deviation of the statistical forecasts. We then categorised these normalised adjustments as being small if they were below the median of all adjustments in the database, large if they were between the median and 75$^{\text{th}}$ percentile and very large if they exceeded this percentile. Given that a few of the adjustments were extremely large, this categorisation led to a more robust analysis and reduced the influence of these extreme observations. Table 2 presents the observed frequencies, where previous moderate losses and gains ($-1 \leq \beta_{t-1} \leq 3$) are pooled, so that big losses are kept distinct. Also, the percentage differences of the realised frequencies compared to the expected ones (assuming independence) divided by the realised numbers are given in brackets. For example, far more very large adjustments are observed in periods following an extra large overshoot than would be expected if adjustment behaviour in the second period was independent of what happened in the first.

Table 2: Size of adjustment following different values of β.

Adjustment size at t	β_{t-1} (type of adjustment at $t-1$) L wrong direction $\beta_{t-1} < -1$	Moderate loss or gain $-1 \leq \beta_{t-1} \leq 3$	XL overshoot $(\beta_{t-1} > 3)$
Small	1331 (-10.3%)	7075 (3.9%)	690[1] (-20.1%)
Large	736 (0.2%)	3357 (-1.3%)	455 (8.9%)
Very large	870[1] (15.6)	3165 (-7.4%)	513[1] (19.2%)

Percentage differences of the realised frequencies compared to the expected ones, assuming independence, are in brackets.

[1] Largest contributions to chi-squared statistic.

When the chi-squared test of independence was applied to table 2 $\chi^2 = 116.6$ with $p < 0.0001$ suggesting that β_{t-1} and the size of adjustment at t are dependent. Table 2 indicates that very large adjustments are more probable particularly after a very large overshoot in the previous period and also after a large wrong direction adjustment. Also, it is less likely that a small adjustment will occur after a big loss. This provides support for H1.

We next test H2 to see whether, following a big loss, the experts adjusted in the same direction as the previous forecast error. To investigate this, the following mixed effects logistic regression equation was fitted to the 18192 observations in the database. The estimation of the model took into account that we have repeated measures for each SKU. The two tailed p-values assume that $Z = \frac{m}{se(m)}$ follows a standard normal distribution, where m is the estimated coefficient and $se(m)$ is an estimate of its standard error.

$$\begin{array}{lccccc} \ln\left(\frac{\Pi}{1-\Pi}\right) = & -0.01 & -0.84\beta_{t-1}^{LW} & -0.84\beta_{t-1}^{XL} & -0.10L_t & -0.33VL_t \\ \text{p-values:} & & (0.000) & (0.000) & (0.007) & (0.000) \end{array}$$

where:

- Π: the probability that the adjustment at t has the same direction as the error at $t-1$.
- $\beta_{t-1}^{LW} = 1$ if the loss at $t-1$ resulted from a large wrong direction adjustment (i.e. $\beta < -1$), 0 otherwise.
- $\beta_{t-1}^{XL} = 1$ if the loss at $t-1$ resulted from a very large overshoot (i.e. $\beta > 3$), 0 otherwise.
- $L_t = 1$ if the adjustment at t was large, 0 otherwise.
- $VL_t = 1$ if the adjustment at t was very large, 0 otherwise.

This logistic regression shows that the probability that the adjustment is in the same direction as the previous error is significantly reduced following large wrong direction adjustments, large overshoots and where the adjustment is large or very large. It suggests that H2 should be rejected and indicates that after a large loss forecasters are more likely to persist in adjusting forecasts in the opposite direction to that suggested by the error.

We next examine the consequences of this behaviour on losses by examining the association between β_t and β_{t-1}. When the chi-squared test of independence was applied to table 3 $\chi^2 = 183.7$ with $p < 0.0001$ so there appeared to be a dependence between losses in consecutive periods providing support for H3. It can be seen that there is a higher probability of large wrong direction adjustments when a wrong direction adjustment has been made in the previous period. Similarly, very large overshoots tend be more probable following very large overshoots.

Table 4 provides further insights. It shows the association between β_{t-1} and instances of adjustments at t that are very large in size, contrary to the direction of the previous error and result in another big loss (i.e. a large wrong direction adjustment or a very large overshoot). When the chi-squared test of independence was applied to table 4 $\chi^2 = 346.4$ with $p < 0.0001$ indicating that such adjustments are much more probable following a large wrong direction adjustment or a very large overshoot. Taken together these results support the notion that, following a large loss, forecasters are more likely persist in making large adjustments in a direction contrary to that suggested by their previous error and this behaviour is likely to lead to a serious deterioration in forecast accuracy.

4.4. Discussion

The results presented in the previous subsection indicate that after a big loss forecasters have a tendency to make an adjustment for the following period that is large and in a direction that is opposite to that suggested by their previous error. For example, if they

Table 3: Transition matrix for previous and current β values.

β_t (type of adjustment at t)	β_{t-1} (type of adjustment at $t-1$)		
	L wrong direction	Moderate loss or gain	XL overshoot
	$\beta_{t-1} < -1$	$-1 \leq \beta_{t-1} \leq 3$	$\beta_{t-1} > 3$
L wrong direction $\beta_t < -1$	622[1] (23.3%)	2018 (-9.4%)	314 (14.3%)
Moderate loss or gain $\beta_t \in [-1, 0) \cup (0, 3]$	1978 (-10.7%)	10475 (3.2%)	1115 (-10.9%)
XL overshoot $\beta_t > 3$	337 (20.0%)	1104 (-13.1%)	229[1] (33.5%)

Percentage differences of the realised frequencies compared to the expected ones, assuming independence, are in brackets.

[1] Largest contributions to chi-squared statistic.

Table 4: Association of previous β value with subsequent adjustment and consequences.

Adjustment at t	β_{t-1} (type of adjustment at $t-1$)		
	L wrong direction	Moderate loss or gain	XL overshoot
	$\beta_{t-1} < -1$	$-1 \leq \beta_{t-1} \leq 3$	$\beta_{t-1} > 3$
Very large, contrary direction adjustment resulting in a big loss	368[1] (46.2%)	643[1] (-42.5%)	215[1] (48.0%)
Does not meet all 3 conditions above	2569 (-6.6%)	12954 (2.1%)	1443 (-7.2%)

Percentage differences of the realised frequencies compared to the expected ones, assuming independence, are in brackets.

[1] Largest contributions to chi-squared statistic.

have incurred a big loss by forecasting too high in the previous period they are still more likely to make a large upwards adjustment in the following period, even though this behaviour increases the probability of a second big loss. Thus, rather than overreacting to outcome feedback, the experts appear to be ignoring it. There are a number of possible explanations for this.

The first possibility is that the forecasts are subject to asymmetric loss. In the pharmaceutical industry being out-of-stock is more serious than having surplus stocks, as holding costs of drugs are relatively low so the forecasters may have an incentive to persist in making upwards adjustments to statistical forecasts. An analysis of this data set by Franses et al. (2011) concluded that there was evidence that the experts' forecasts were subject to asymmetric loss. However, overall 57.8% of the adjustments to the statistical forecasts were in the upwards direction so forecasts were frequently lowered. Moreover, it seems unlikely that relatively rare very large changes to the statistical forecasts would be made because of asymmetric loss. Given that the same loss function would be likely to apply to a series for long periods, asymmetric loss would be more likely to lead to a pattern of consistent changes rather than the occasional very large adjustment. Indeed, the loss functions identified by Franses et al. (2011) would be consistent with moderate adjustments.

A second possibility is that the forecasters were suffering from the gambler's fallacy and expecting that a chance event that produced unforeseen and exceptional sales in the previous period would be 'balanced out' by an exceptional sales movement in the opposite direction in the next period. If this is the case we would expect that the statistical forecast error in the period preceding a big loss would be exceptionally high. In fact, the statistical forecast errors tended to be slightly lower than average in the period preceding a big loss. The mean absolute statistical forecast error, normalised by dividing by the standard deviation of the statistical forecasts for each series, was 1.31 for all periods and 1.23 in the periods immediately preceding a big loss.

A third possibility is that a special event that would have a large impact on sales was known be occurring in the future but the timing of its effects was misjudged. If the effect of the event failed to materialise in one period then it might be expected to occur in the subsequent period instead or in the period after that. If this was the case then in the period after two consecutive periods of big losses we might expect the statistical forecast error to be higher as it failed to forecast the special event when it finally occurred. Again there was no support for this. In the periods after two big losses the normalised mean absolute statistical forecast error was lower than the norm at 1.11.

This leaves the possibility that the forecasters were prepared to make bold interventions on the basis of unreliable information or a misinterpretation of information and that they were prepared to persist in making large adjustments on this basis even when there was evidence from the previous period that this had reduced forecast accuracy and they had incurred a big loss. Their persistence in making a large wrong direction adjustment adjustment following an earlier such adjustment suggests a resistance to recognising a step change in sales. This could indicate that information pointing to such a change was either not available or was discounted. Such behaviour might occur when the initial large loss is simply attributed to a transient shock to the system. Their persistence in making an adjustment that resulted in a large overshoot, following an earlier such adjustment, is consistent with an expectation of a step change in sales that is not forthcoming. Again, this may reflect a lack of availability of reliable information or the misinterpretation of such information.

If the forecasters' judgments were being distorted by unreliable information this would not be consistent with the results of a laboratory study by Remus et al. (1998) which found that judgmental forecasters were not necessarily misled by incorrect information. However, in the Remus study people were supplied with rumours that suggested particular future movements in time series, but no reasons or arguments to support these possible movements were provided. In the field there is likely to be a richer variety of information and misinformation available to forecasters who have the difficult task of assessing its reliability, relevance and importance. Motivational and political factors may contribute to the misinterpretation of information. For example, wishful thinking may lead to the discounting of negative information if an increase in sales is desired (Tyebjee, 1987). A desire to produce forecasts that are politically acceptable to senior managers may have a similar effect on how information is interpreted and whether it is discounted (Fildes and Hastings, 1994).

The findings in section 4.3 are insensitive to the values of the thresholds for defining big losses. A replication of the analysis using as a threshold for big losses an adjustment that deviated by more than 150% or 250% (instead of 200%) from a perfect adjustment produced practically the same results. However, some of the results differed when the threshold is set to 100%, which is equivalent to no separation between moderate losses and big losses. In this case, there is no change in experts' behaviour with regards to the size of adjustment after a wrong direction adjustment (H1). This indicates that differentiating between moderate losses and big losses has enhanced our understanding of forecaster behaviour.

In addition, directional analysis has been performed in order to check if there are any differences in experts' behaviour after a big loss that was the result of a positive or a negative adjustment. While generally big losses are observed more frequently after positive adjustments (a result that corroborates with Fildes et al. (2009)), there are no significant differences in subsequent experts' behaviour linked with the direction of the adjustment at the previous period. In both cases, positive and negative adjustments that resulted in big losses at period $t-1$ are more likely to be followed by another big loss at period t, as the result of a large or very large adjustment, contrary to the direction suggested by the previous error.

So, to address the first research question, following big losses, experts are more likely to make large judgmental adjustments in the opposite direction to the previous large error. Also, it is more likely that these adjustments will lead once again to another big loss.

4.5. Forecasting performance

Table 5 presents the performance of the experts forecasts (judgmentally adjusted statistical forecasts). This is provided for the different values of β at period $t-1$ that refer to the various types of judgmental adjustments occurred in the previous period. Moreover, the results are presented separately for each group with regards to the size of adjustments at time t. The forecasting performance is measured in terms of accuracy by the Average Relative Mean Absolute Error (AvgRelMAE). As mentioned in section 3, this error measure benchmarks the performance of the expert forecasts comparing it directly to that of the statistical forecasts. Values lower than unity denote improvement in performance compared to the benchmark, while values greater than one indicate deterioration in forecast accuracy. The AvgRelMAE is applied after considering a symmetric trimmed mean so that any extreme values are eliminated, as suggested by Davydenko and Fildes (2013). We opt for a 2% trimming level.

Table 5: Average forecasting performance for various previous β values and current adjustments' sizes.

Adjustment size at t		β_{t-1} (type of adjustment at $t-1$)			
		L wrong direction	Moderate loss or gain	XL overshoot	Any type
		$\beta_{t-1} < -1$	$-1 \leq \beta_{t-1} \leq 3$	$\beta_{t-1} > 3$	$\beta_{t-1} \in \mathbb{R}$
Small	AvgRelMAE	1.052	0.989	1.009	0.999
	n	1331	7075	690	9096
Large	AvgRelMAE	1.144	0.981	1.119	1.011
	n	736	3357	455	4548
Very large	AvgRelMAE	1.331	0.997	1.331	1.070
	n	870	3165	513	4548
Any size	AvgRelMAE	1.156	0.983	1.118	1.021
	n	2937	13597	1658	18192

The first observation is that for this specific data set, judgmental adjustments are overall worse by 2.1% compared to the statistical benchmark. This result agrees with previous research by Franses and Legerstee (2010). A possible explanation for this is that managers may not always have domain knowledge that needs to be incorporated in the system forecasts but just interfere to take control and ownership of the forecasts. The "illusion of control" is a well known disadvantage in management judgment (Kottemann et al., 1994). In order to have a more clear view about this statistic, let us focus on the different groups with regards to the type of adjustment made at time $t-1$ as measured by the β coefficient.

We observe that after a moderate gain or loss ($-1 \leq \beta_{t-1} \leq 3$) experts' interventions lead to improving the statistical forecasts by 1.7% on average. This improvement is even larger (5.1%) if we further analyse the data focusing on the cases where adjustments at time t follow a gain at $t-1$ ($0 \leq \beta_{t-1} \leq 2$). On the other hand, the average value of AvgRelMAE is significantly higher than unity for the judgmental adjustments that follow big losses. In fact, these adjustments are on average 14.2% worse than the statistical forecasts. Also, while the deterioration occurs for any sizes of adjustments, the value of the AvgRelMAE increases together with the size of adjustment after a big loss. This indicates that very large adjustments after big losses resulted in significant losses in forecasting performance.

So, experts' large-sized adjustments after big losses may significantly hurt forecast accuracy. As such, it is of critical importance that we try to reduce, eliminate or adjust the interventions made after big losses. We will try to address this in the next section by controlling and correcting the experts' behaviour.

5. Supporting forecasters' behaviour

So far we have shown through a very large database that big losses in judgmental adjustments negatively affect forecasters' behaviour. The obvious next step is how can we limit these negative effects or even take advantage of them. In this section we introduce some strategies as to support the forecasters' behaviour with regards to judgmental interventions occuring after big losses.

We identify three simple strategies that could be potentially applied in such cases:

- Guidance. Given that the judgmental adjustments are performed within a specialised computer software (Forecasting Support System or FSS), users could be exposed to information with regards to their past performance. In addition, they could be advised not to perform any correcting actions (interventions) when producing forecasts for periods that follow big losses as a result from experts' adjustments, as these have empirically shown to lead to decreased forecasting performance. Decisional guidance for support systems has been previously suggested in the literature (for example see Silver, 1991; Sauter, 1997; Goodwin et al., 2007).
- Restrictiveness. This strategy suggests that any judgmental interventions should be simply disregarded. In other words, the final forecast should be equal to the statistical forecast, $\alpha_t = 0$ in equation 2. This suggests that we assume that the expert forecast does not provide any additional insights ($\beta_t = \pm\infty$). When implemented into a FSS, managers should (by default) not be able to intervene at all on the statistical outputs after big-losses periods.
- 50% statistics + 50% judgment. This strategy is also known as the Blattberg-Hoch approach and is based on the argument that "any combinations of forecasts proves more accurate than the single inputs" (Blattberg and Hoch, 1990). As such, this strategy suggests the use of unconditional 50-50% weights applied for linearly combining statistical and judgmental inputs. From equation 2, α should be simply replaced with 0.5 for all periods t; equivalently, as suggested by equation 7, the ex-ante prediction for β_t is 2 for any future period. The Blattberg-Hoch approach by definition works on sets of forecasts that have been independently produced. In the case of judgmental adjustments, though, it acts as a dampener on the adjustments (Fildes et al., 2009). As such, it might be beneficial when managers tend to over-adjust, as it is the case after big losses. In any case, it has shown promising performance when applied for combining statistical and expert forecasts (Fildes et al., 2009; Franses and Legerstee, 2011a). While Franses and Legerstee (2011a) applied it unconditionally on the same database examined in this study, Fildes et al. (2009) identified positive adjustments as those benefiting from the 50-50% strategy. Here we focus on the application of this strategy strictly after big losses, which are found to be linked with large adjustments at the next period.

Table 6 presents the forecasting performance when each one of the aforementioned strategies (column 1) is applied to the judgmental adjustments following big losses. The performance of the current practice (no action in correcting adjustments after big losses) is provided as well. The forecasting performance is measured in terms of AvgRelMAE. Apart from providing the results generally after a big loss independently of its direction (column 2), we also distinguish between big losses as a result of a large wrong direction adjustment and extra large overshoots (columns 3 and 4 respectively). Finally, the last column provides the overall accuracy, measured as the AvgRelMAE across all periods, when the proposed strategy is applied only to the periods after a big loss. In other words, it presents the improvements in the overall forecasting performance when taking actions only after an expert adjustment that resulted in a very large forecast error.

Table 6: Average forecasting performance for different correction strategies.

Correction strategy	After a big loss	After L wrong direction	After XL overshoot	Overall
Current practice	1.142	1.156	1.118	1.021
Guidance*	1.077	1.083	1.064	1.006
Restrictiveness	1.000	1.000	1.000	0.989
50% statistics + 50% judgment	0.965	0.979	0.938	0.981

* Assuming that the adjustment was prevented in 50% of the cases.

Assuming that the judgmental adjustment was prevented by appropriate guidance and advice in 50% of the cases, the first strategy essentially halves the difference in the performance between expert forecasts and the benchmark (statistical forecasts). Generally, the improvement in the forecasting performance as a result of the guidance strategy can be regarded as a function of the percentage of managers that essentially follow the provided advice. At the same time, the guidance (after big losses) strategy leads to 1.5% improvement of the judgmental adjustments overall compared to the current practice.

Restrictiveness, as expected, results in values of 1 for the AvgRelMAE after a big loss. It is very interesting, however, that when this strategy is applied, the overall performance of experts' interventions is for the first time positive (AvgRelMAE = 0.989), improving the relative performance of statistical forecasts by 1.1% overall.

Last but not least, the Blattberg-Hoch 50-50% approach seems to work best for this data set. By applying an equal weight combination of statistics and experts, we end up with final forecasts that are up to 16% better than the expert forecasts The approach also delivers improvements over the statistical forecasts demonstrating there is value in the judgmental adjustments. Increases in accuracy are more substantial after an extra-large overshoot, where, as shown in section 4.3, it is more probable that another extra-large overshoot will occur in the next period. The dampening of these overshoots most likely leads to final forecasts with $\beta < 2$, meaning gains in forecast accuracy. It is worth mentioning that this strategy improves almost 2/3 of the adjustments after big losses.

So, to address the second research question, by taking simple corrective actions for just the 1/4 of the judgmental interventions (the ones made after big losses), not only are we able to tackle the poor performance following such adjustments, but also to improve the overall forecasting performance by 4% compared to the current practice (from 1.021 to 0.981, according to table 6). This translates into an improvement in the performance relative to statistical forecasts of 1.9%.

6. Concluding remarks

Judgmental adjustments of statistical forecasts are very common in companies and other organisations. However, the integration of field knowledge and soft data through judgment is not always performed in the most effective way. As a result, judgmental interventions do not always lead to improved accuracy and in some cases such adjustments lead to significant performance losses. Thus, it is important for organisations to understand the conditions under which judgmental adjustments are more likely to fail.

This paper focuses on the cases where judgmental adjustments are made after significant accuracy losses have resulted from adjustments made in the previous period. After defining a new measure for interpreting the type, magnitude and quality of judgmental adjustments, we examined, through a large empirical data set, the behaviour of forecasters after performing adjustments that led to big losses. We showed that the probability of performing an adjustment that leads to big loss increases in a period following one where a big loss has already occurred. Despite the earlier loss the probability of making a large adjustment in the opposite direction to the one suggested by the previous forecasts error also increases following a big loss. After a big loss forecasters have a propensity to persist in their belief that similar large changes to the statistical forecast are needed despite evidence from outcome feedback that the previous change led to lower accuracy. This appears to be because forecasters are prepared to make repeated bold adjustments based on inaccurate information. Explanations based on asymmetric loss, the gambler's fallacy or misjudging the timing of the effects to special events were not supported by the data. In terms of forecast accuracy, these adjustments produced forecasts that were on average 14% worse than that of statistical methods alone.

Simple correction strategies, such as guidance, restrictiveness and unweighted combination of statistical and expert forecasts, can be applied to improve the forecasting performance after big losses. In fact, we recorded improvements of up to 16% for these periods. The overall gain in accuracy is 4% and this is coming from adjusting further just a quarter of the total number of adjustments. Given the simplicity of the proposed strategies, this result is of practical importance.

The phenomenon identified here potentially applies to other operational situations. Its essential feature is of experts misunderstanding the outcome feedback they receive and repeating the same mistakes. For example, a review of the behavioural newsvendor type experiments in the context of examining medical operations room scheduling shows repetitive error prone behaviour despite feedback (Wachtel and Dexter, 2010). There is also evidence of experimental participants rejecting models in favour of their own (mis)judgments even when given evidence on the superior performance of models (Dietvorst et al., 2015). As Hämäläinen et al. (2013) remark, a key issue of behavioural OR is "how to help people find better strategies" in problem-solving situations, overcoming the biases that typically intrude. In the light of our results we would add that this should incorporate how to design

Desicion Support Systems (or, in the context of this article, Forecasting Support Systems) to provide that effective support so that feedback from systems is used to its best advantage.

Future paths for research include the exploration of more sophisticated strategies for manipulating the adjustments made after big losses. For example, the use of error bootstrap rules (Fildes et al., 2009) or correlation with experts experience and/or behaviour (Franses and Legerstee, 2011a) could be considered as alternatives. The strategy of responding to big losses could be thought of as a simple monitoring scheme which could be compared to regular methods of monitoring such as tracking signals (Gorr and Ord, 2009). The method proposed in this paper using β_t is volatile so comparison with smoothed tracking signals would be valuable. Another critical question for future research is to examine if the automatic adjustment of judgmental adjustments (through a combination of statistical and expert forecasts, or dampening of the experts' adjustments) would lead to a long-term change of forecasters' behaviour with regards to how they perform judgmental interventions. To that end, a possibility for future research would be to examine the effectiveness of providing feedback on adjustments that led to big losses, as explicit feedback to experts on their performance has been shown to lead to more accurate forecasts (Legerstee and Franses, 2014).

References

Armstrong, J. S., 1985. Long range forecasting: From crystal ball to computer (2nd edition). Wiley, New York.

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

Bonaccio, S., Dalal, R. S., 2006. Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences. Organizational Behavior and Human Decision Processes 101, 127–151.

Boulaksil, Y., Franses, P. H., 2009. Experts' stated behavior. Interfaces 39, 168–171.

Brown, L. D., 1988. Comparing judgmental to extrapolative forecasts - it's time to ask why and when. International Journal of Forecasting 4, 171–173.

Carbone, R., Andersen, A., Corriveau, Y., Corson, P. P., 1983. Comparing for different time series methods the value of technical expertise, individualized analysis and judgmental adjustment. Management Science 29, 559–566.

Coval, J. D., Shumway, T., 2005. Do behavioral biases affect prices? Journal of Finance 60, 1–34.

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

de Bruijn, B., Franses, P. H., 2012. Managing sales forecasters. Tinbergen Institute Discussion Paper TI 2012-131/III.

Dietvorst, B. J., Simmons, J. P., Massey, C., 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General 144, 114–126.

Donihue, M. R., 1993. Evaluating the role judgement plays in forecast accuracy. Journal of Forecasting 12, 81–92.

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

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

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

Fildes, R., Goodwin, P., 2007b. Good and bad judgment in forecasting: Lessons from four companies. Foresight: The International Journal of Applied Forecasting 8, 5–10.
Fildes, R., Goodwin, P., Lawrence, M., Nikolopoulos, K., 2009. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. International Journal of Forecasting 25, 3–23.
Fildes, R., Hastings, R., 1994. The organization and improvement of market forecasting. Journal of the Operational Research Society 45, 1–16.
Fildes, R., Petropoulos, F., 2015. Improving forecast quality in practice. Foresight: The International Journal of Applied Forecasting 36, 5–12.
Franses, P. H., Legerstee, R., 2009. Properties of expert adjustments on model-based SKU-level forecasts. International Journal of Forecasting 25, 35–47.
Franses, P. H., Legerstee, R., 2010. Do experts' adjustments on model-based SKU-level forecasts improve forecast quality? Journal of Forecasting 29, 331–340.
Franses, P. H., Legerstee, R., 2011a. Combining SKU-level sales forecasts from models and experts. Expert Systems with Applications 38, 2365–2370.
Franses, P. H., Legerstee, R., 2011b. Experts' adjustment to model-based SKU-level forecasts: does the forecast horizon matter? Journal of Operational Research Society 62, 537–543.
Franses, P. H., Legerstee, R., 2013. Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? International Journal of Forecasting 29, 80–87.
Franses, P. H., Legerstee, R., Paap, R., 2011. Estimating loss functions of experts. Tinbergen Institute Discussion Paper TI 2011-177/4.
Garvey, R., Murphy, A., Wu, F., 2007. Do losses linger? Evidence from proprietary stock traders. Journal of Portfolio Management 33, 75–83.
Goodwin, P., 2000. Improving the voluntary integration of statistical forecasts and judgement. International Journal of Forecasting 16, 85–99.
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.
Gorr, W. L., Ord, J. K., 2009. Introduction to time series monitoring. International Journal of Forecasting 25, 463–466.
Hämäläinen, R. P., Luoma, J., Saarinen, E., 2013. On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. European Journal of Operational Research 228, 623–634.
Harvey, N., 1995. Why are judgments less consistent in less predictable task situations? Organizational Behavior and Human Decision Processes 63, 247–263.
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.
Kottemann, J. E., Davis, F. D., Remus, W. E., 1994. Computer assisted decision making: Performance, beliefs, and the illusion of control. Organizational Behavior and Human Decision Processes 57, 26–37.
Lawrence, M., Goodwin, P., O'Connor, M., Onkal, D., 2006. Judgmental forecasting: A review of progress over the last 25 years. International Journal of Forecasting 22, 493–518.
Lawrence, M., O'Connor, M., Edmunson, B., 2000. A field study of sales forecasting accuracy and processes. European Journal of Operational Research 122, 151–160.
Legerstee, R., Franses, P. H., 2014. Do experts' SKU forecasts improve after feedback? Journal of Forecasting 33, 69–79.
Leitner, J., Leopold-Wildburger, U., 2011. Experiments on forecasting behavior with several sources of information a review of the literature. European Journal of Operational Research 213, 459–469.

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

Mathews, B., Diamantopoulos, A., 1986. Managerial intervention in forecasting. an empirical investigation of forecast manipulation. International Journal of Research in Marketing 3, 3–10.

Mathews, B., Diamantopoulos, A., 1989. Judgmental revision of sales forecasts: A longitudinal extension. Journal of Forecasting 8, 129–140.

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

McCarthy, T. M., Davis, D. F., Golicic, S. L., Mentzer, J. T., 2006. The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. Journal of Forecasting 25, 303–324.

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

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

Önkal, D., Gönül, M. S., 2005. Judgmental adjustment: a challenge for providers and users of forecasts. Foresight: The International Journal of Applied Forecasting 1, 13–17.

Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. 'Horses for Courses' in demand forecasting. European Journal of Operational Research 237, 152–163.

Ramnath, S., Rock, S., Shane, P., 2008. The financial analyst forecasting literature: A taxonomy with suggestions for further research. International Journal of Forecasting 24, 34–75.

Remus, W., O'Connor, M., Griggs, K., 1998. The impact of information of unknown correctness on the judgmental forecasting process. International Journal of Forecasting 14, 313–322.

Sauter, V. L., 1997. Decision support systems: an applied managerial approach. Wiley, New York.

Silver, M. S., 1991. Decisional guidance for computer-based decision support. MIS Quarterly 15, 105–122.

Smith, G., Levere, M., Kurtzman, R., 2009. Poker player behavior after big wins and big losses. Management Science 55, 1547–1555.

Spranca, M., Minsk, E., Baron, J., 1991. Omission and commission in judgment and choice. Journal of Experimental Social Psychology 27, 76–105.

Staw, B. M., 1976. Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. Organizational Behavior and Human Performance 16, 27–44.

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.

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.

Turner, D. S., 1990. The role of judgement in macroeconomic forecasting. Journal of Forecasting 9, 315–345.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

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

Vere, D. T., Griffiths, G. R., 1995. Modifying quantitative forecasts of livestock production using expert judgements, an application to the australian lamb industry. Journal of Forecasting 14, 453–464.

Wachtel, R. E., Dexter, F., 2010. Review of behavioral operations experimental studies of newsvendor problems for operating room management. Anesthesia and Analgesia 110, 1698–1710.

Wolfe, C., Flores, B., 1990. Judgmental adjustment of earnings forecasts. Journal of Forecasting 9, 389–405.

Xu, J., Harvey, N., 2014. Carry on winning: The gamblers' fallacy creates hot hand effects in online gambling. Cognition 131, 173–180.