

The Relative Influence of Advice From Human Experts and Statistical Methods on Forecast Adjustments

DILEK ÖNKAL^{1*}, PAUL GOODWIN², MARY THOMSON³,
SINAN GÖNÜL⁴ and ANDREW POLLOCK⁵

¹*Faculty of Business Administration, Bilkent University, Ankara, Turkey*

²*School of Management, University of Bath, Bath, UK*

³*Division of Psychology, Glasgow Caledonian University, Glasgow, UK*

⁴*Department of Business Administration, TOBB University of Economics and Technology, Ankara, Turkey*

⁵*Division of Mathematics, Glasgow Caledonian University, Glasgow, UK*

ABSTRACT

Decision makers and forecasters often receive advice from different sources including human experts and statistical methods. This research examines, in the context of stock price forecasting, how the apparent source of the advice affects the attention that is paid to it when the mode of delivery of the advice is identical for both sources. In Study 1, two groups of participants were given the same advised point and interval forecasts. One group was told that these were the advice of a human expert and the other that they were generated by a statistical forecasting method. The participants were then asked to adjust forecasts they had previously made in light of this advice. While in both cases the advice led to improved point forecast accuracy and better calibration of the prediction intervals, the advice which apparently emanated from a statistical method was discounted much more severely. In Study 2, participants were provided with advice from two sources. When the participants were told that both sources were either human experts or both were statistical methods, the apparent statistical-based advice had the same influence on the adjusted estimates as the advice that appeared to come from a human expert. However when the apparent sources of advice were different, much greater attention was paid to the advice that apparently came from a human expert. Theories of advice utilization are used to identify why the advice of a human expert is likely to be preferred to advice from a statistical method. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS advice; advice utilization; forecasting; forecast adjustment; source framing; role of experts

* Correspondence to: Dilek Önkal, Faculty of Business Administration, Bilkent University, 06800 Ankara, Turkey.
E-mail: onkal@bilkent.edu.tr

INTRODUCTION

Forecasters and decision makers often receive advice from different sources. For example, a demand forecaster in a manufacturing company may receive advice from colleagues in marketing, sales, and logistics, and may make his or her decisions accordingly. Similarly, investors, who need to assess the potential gains or losses of buying or selling a particular stock, will probably have access to advice from a range of sources including stock market experts. However, some of the advice that is received will not be from human experts. The outputs of statistical forecasting software packages also represent a source of advice. In practice, the forms of the advice received from these two sources may differ and this may lead to different levels of attention being paid to the advice. For example, the human expert may also provide a rationale for the suggested forecast while the statistical method may merely provide a number (or vice versa). Alternatively, the human expert might deliver advice orally, while the statistical software merely provides a number on a crowded computer screen. However, if all of these differences can be eliminated this raises the interesting question: Will people make greater use of a given piece of advice if it emanates from a human expert or if it comes from a statistical method, or will they treat both forms of advice equally? The current paper attempts to answer this question in the context of stock price forecasting, a domain where a universal, multi-billion dollar industry has emerged comprising of advisors and investors.

LITERATURE REVIEW

The literature on advice-taking has tended to demonstrate that the use of advice leads to increased accuracy when forecasts or other judgments need to be made (e.g. Sniezek, Schrah, & Dalal, 2004; Yaniv, 2004a). One reason for this is that the mere act of combining multiple recommendations increases accuracy because it decreases the random error associated with each separate opinion (Yaniv, 2004a, b). However, another widespread finding has been that people typically do not take sufficient account of the advice given. They overweigh their own estimation relative to that of their advisor(s), a phenomenon referred to as 'advice discounting' (e.g. Yaniv, 2004b; Yaniv, & Kleinberger, 2000). Harvey and Fischer (1997) found that decision makers typically only moved their initial estimates by a 'token' 20–30% towards their advisor's recommendation. In fact, individuals have been shown to excessively discount the value of advice even when they are shown how good the advice is and informed that they are unlikely to improve upon it (Goodwin, 2000; Goodwin & Fildes, 1999; Lawrence & Sim, 1999; Lim & O'Connor, 1995).

Various explanations have been put forward to account for advice discounting. One explanation is that people tend to anchor too closely to their initial estimate so that the adjustments they make to take into account the advice are insufficient (e.g. Lawrence & O'Connor, 1995; Lim & O'Connor, 1995; Tversky & Kahneman, 1974). It has also been argued by Yaniv and others (Yaniv, 2004a,b; Yaniv & Kleinberger, 2000) that advice discounting partly takes place because decision makers have access to their own justifications and reasoning behind their particular viewpoints, but not usually to those of their advisors. It might also be the case that an 'egocentric bias' is at work in that individuals believe that their own opinions are superior to those of other people (Krueger, 1999). Much research in the psychology literature is consistent with this possibility. For one thing, individuals have been found to have a strong self-focus. People tend to transform their absolute sense of success or failure into comparative judgements with others (Klar & Giladi, 1999). Self-related information is, therefore, paramount to all aspects of their lives (Clement & Krueger, 2000; Dunning & Hayes, 1997; Krueger & Stanke, 2001); and, in order to boost self-image and self-esteem, people tend to exaggerate their abilities, skills and traits, perceiving themselves as above average. As suggested by Krueger (1999), 'When people compare themselves with their peers, they focus egocentrically on their own skills and insufficiently take into account the skills of the comparison group' (p. 221). People appear to anchor on self-assessments and do not adjust sufficiently to take others' skills into account. Accordingly, cognitive conceit

and overconfidence can arise because people are not realistic in their self-assessments (Block & Harper, 1991). The actual quality of the advice is also significant, with judges being found to discount inaccurate advice to a greater extent than accurate advice (Lim & O'Connor, 1995; Yaniv & Kleinberger, 2000). Also, negative information about an advisor is perceived to be more diagnostic than positive information (Yaniv & Kleinberger, 2000).

It has also been suggested that task complexity influences the extent of advice discounting, with judges appearing to discount advice less when the task is difficult (Schrah, Dalal, & Sniezek, 2006). In a similar vein, the width of the confidence interval expressed by the advisor has been argued to indicate the advisor's perception of his or her knowledge, which, in turn, usually results in more confident advisors receiving less advice discounting (Yaniv, 1997; Yaniv & Foster, 1997). However, advisor confidence levels have been found to be less influential than agreement between advisors (Sniezek & Buckley, 1995). When the level of agreement between advisors is low, the judges' post-advice confidence is correspondingly low (Budescu, Rantilla, Yu, & Karelitz, 2003), particularly if the judges think that the conflicting advisors had access to the same information (Budescu & Rantilla, 2000).

Why might advice discounting vary depending on whether the source of the advice is a human expert or a statistical forecasting method? Armstrong's seer-sucker theory (Armstrong, 1980) suggests that people's faith in experts' judgments and advice persists even when there is overwhelming evidence that these experts are no more accurate than people with only basic knowledge. This has been found in areas as diverse as political forecasting (Tetlock, 2005), forecasting the outcome of conflicts (Green & Armstrong, 2007), and stock market forecasting (e.g. Armstrong, 1980; Cowles, 1933) where experts sometimes perform no better in forecasting than people with only basic knowledge. Armstrong argues that one reason for this reliance on experts is the shifting of responsibility—the expert can be blamed if the forecasts are inaccurate. As pointed out by Bonaccio and Dalal (2006), motives such as sharing or shifting responsibility and avoiding the appearance of rejecting help (Harvey & Fischer, 1997) only come into play when the advisor is human. When the advice comes from a statistical model, such motives may be less relevant.

There may be other reasons for a tendency to rely more on human experts (Wærn & Ramberg, 1996). Indeed the tendency may be deeply ingrained. Even de Dombal, who developed a medical diagnosis system that was far superior to the judgments of clinicians, believed that where human life was at risk decision makers should rely on their own expertise, not a computer's (Kleinmutz, 1990). Such inflated confidence in expert/clinical judgment at the expense of inferior expectations for statistical/mechanical predictions has been repeatedly documented, even in light of the better accuracy of the latter (Dawes, Faust, & Meehl, 1989; Grove & Lloyd, 2006; Grove & Meehl, 1996; Meehl, 1954). The term 'expert' itself carries connotations of great knowledge and experience and acts as an advertisement for the advice itself. The frequent use of experts in print and broadcast media carries an implied message that their advice is reliable. Moreover, the performance of expert judgment in many fields (such as weather forecasting, physics, mathematics, and medicine) is generally good (Shanteau, 1992), hence it is easy to wrongly infer that it is also good in other fields like stock market forecasting. Harvey and Fischer (1997) found that advice discounting is less when advisors are perceived by judges to be more experienced and knowledgeable than themselves. This may be because advice from an expert mitigates the egocentric component of advice discounting. When advice is unsolicited, expert advice is also seen as being less intrusive than advice from other sources (Goldsmith & Fitch, 1997, Jungermann & Fischer, 2005).

Even when experts are poor at forecasting, they may be able to provide convincing and valid explanations of past and present events (Armstrong, 1980) so that people may judge that they are also capable of accurate forecasts. Experts are also usually good at communication (Shanteau, 1992) and will generally supply a clear rationale to underpin their forecasts. In contrast, the reasoning underlying a statistical forecast may be unavailable or, even if it is supplied, it may be mysterious to those not trained in statistics. For example, Goodwin and Fildes (1999) found that computer-based explanations intended to underpin statistical forecasts and discourage adjustments to these forecasts, were only effective when time series had a relatively simple

structure so that short, easy-to-understand justifications could be delivered. This limits the ability of the statistical method to compete with the advisee's own rationale for a particular forecast.

People's experience of experts is that they manifest strong confidence in their abilities (Shanteau, 1992), and they have an ability and propensity to explain away errors in their forecasts by attributing them to causes other than their own expertise (Tetlock, 2005). This will serve to reduce the impact of negative information about the advisor and maintain confidence in their advice (e.g. see Snizek & Buckley, 1995). Experts may also be seen to have a reputation to maintain and hence a motive to strive for forecast accuracy. This means that judges will perceive the expert's goals as being congruent with their own and place greater trust in their advice (Eisenhardt, 1989). In addition, in macroeconomic and earnings forecasting, experts often manifest herding behaviour (e.g. Cote & Goodstein, 1999) so that their forecasts are similar. Such agreement may reinforce belief in each individual expert's forecast and the perception that the experts are able to draw on a common wisdom.

In contrast, most users of statistical forecasts are likely to know that these forecasts are based on extrapolations of relationships or patterns observed in the past, while experts may be perceived to have information about future events. Moreover, there may be a perception that statistical forecasts are slow to respond to a fast changing world and hence may be out of date (Blattberg & Hoch, 1990). Also, the aim of statistical forecasts may not be clear to the user. While an expert usually claims to forecast an actual outcome at some future date, the statistical method aims only to forecast the expected value of the probability distribution which will generate the outcome. If this is known by the judge, then the goal of the statistical forecast may not be congruent with the judge's and hence the advice may be discounted. Indeed, a common reason for rejecting statistical guidance is the perception that statistical methods are not applicable to unique individual cases (Dana & Thomas, 2006). If the expert's forecast is perceived to be 'case-specific', it is likely to attract far more attention than the 'base rate' forecast of the statistical method.

In summary, the literature suggests that, when making forecasts, people are likely to attach more weight to advice when it is perceived to come from a human expert than when it is perceived to come from a statistical method. The aim of the current study is to test this in a domain where there is no clear evidence that experts' forecasts have undisputable value, namely stock market forecasting. The next sections describe two studies that were designed to test the extent to which people make use of 'statistical' and 'expert' advice when the nature, delivery and presentation of the advice from the two sources is identical. This is followed by 'Discussion' and 'Conclusion' Sections which consider the implications of the findings for forecasting practice and provide suggestions for future research.

STUDY 1

Participants

The participants were 76 second and third-year business and economics students at Bilkent University who were taking statistics and forecasting courses. Participants' ages ranged between 19 and 22; gender composition was 45% female ($n = 34$) and 55% male ($n = 42$). No monetary incentives were given, but participation in the study earned the students extra credit points in their courses.

Procedure

In a pencil and paper task, participants were given time series plots for the weekly closing prices of 30 stocks. Real data from two years prior to the experiment were used to avoid biases from contextual information and expectations. Friday closing prices from stocks traded in the Istanbul Stock Exchange were selected on the basis of high volume of trade at the time period under consideration. For each of the 30 stocks, a time series plot of 30 weekly closing prices along with the numerical values for the last four closing prices appearing in

the plot were given (see Appendix A for a sample elicitation form). Participants were informed that the data reflected original stock price values from an undisclosed time period, and that the stock names were not revealed so as to prevent potential biases from extraneous information. The participants were randomly distributed to two groups receiving the same set of time-series (with a different randomized ordering for each participant). For each stock in turn, the participants were:

- (1) asked to make a one-week-ahead point forecast of the closing price, followed by a 70% prediction interval and a 95% prediction interval;
- (2) requested to examine a given set of external forecasts (a point forecast, a 70% prediction interval, and a 95% prediction interval)—where participants in Group 1 ('expert-forecast framing' group; $n = 38$) were led to believe that they were presented with forecasts given by a financial expert; whereas the students in Group 2 ('statistical-forecast framing' group; $n = 38$) were told that they were given predictions generated by a statistical model;
- (3) asked to use their judgment to adjust their original forecasts in light of the presented external forecasts (if deemed necessary or pertinent), thus providing a judgmentally adjusted point forecast, and judgmentally adjusted 70 and 95% prediction intervals.

Each participant was presented with a booklet containing the instructions, the forms, and a post-experiment questionnaire. In this questionnaire participants were asked to evaluate their level of knowledge of stock prices and probability, to evaluate their own forecasting performance and that of the advice, to indicate their level of trust in the sources of advice and to assess how meaningful and accurate they thought the point and interval forecasts were (a copy of the questionnaire is available from the lead author).

Instructions were discussed and detailed examples were presented at the beginning of experimental sessions. Participants in Group 1 ('expert-forecast framing' group) were told that the forecasts they were given came from a financial expert who makes good stock price predictions. Similarly, participants in Group 2 ('statistical-forecast framing' group) were told that their given predictions were generated by a statistical model that makes good stock price predictions. Regardless of the labelling for their groups, all participants were in fact given external forecasts that were supplied by a financial expert (i.e. Group 1 has received the correct information for the source of external forecasts).

RESULTS OF STUDY 1

Influence of advice

There are a number of measures designed to assess the extent to which advice has influenced an individual's judgment and all have advantages and disadvantages (Bonaccio & Dalal, 2006). On balance, Bonaccio and Dalal (2006) favour regression-based approaches where the criterion, in our case, would be the adjusted forecast of the judge. In particular, they recommend the use of dominance weights (e.g. see Azen & Budescu, 2003). These weights are designed to compare the relative importance of predictors in a multiple regression model and are based on the average increase in the criterion variance that is explained when a given predictor is added to models containing subsets of the other predictors. However, our main interest here is not in the relative importance of the initial forecast and the advice for a single situation, rather, it is to compare the extent of advice utilization between two situations. This would involve comparing dominance indices between variables in different regressions models, rather than for variables within the same model. For our purposes, we therefore think that the positional measure of Harvey and Fischer (1997) allows a clearer comparison to be made of the utilization of advice from the two apparent sources. This measure, which is set out below, is problematical where the initial forecast and the advised forecast are the same, or when the forecast is adjusted in the opposite direction to that suggested by the advice. As recommended by Bonaccio and Dalal (2006), we have therefore reported the number of 'problem cases' and assessed the extent to which

our conclusions are sensitive to these cases. The positional measure (labeled SHIFT) simply indicates the extent to which participants' adjusted forecasts shifted between their initial forecasts and the external forecasts given to them. It is computed as follows:

$$\text{SHIFT} = (\text{Adjusted forecast} - \text{Initial forecast}) / (\text{External forecast} - \text{Initial forecast})$$

Excluding those cases where the initial and the external forecasts are identical, $\text{SHIFT} > 0.5$ means that the adjusted prediction is closer to the external forecast while $\text{SHIFT} < 0.5$ means that the adjusted prediction is closer to the initial forecast. If $\text{SHIFT} = 0.5$ then the adjusted forecast is halfway between the external and the initial predictions. Note that if $\text{SHIFT} > 1.0$, the adjusted forecast is closer to the external forecast (but outside the interval of initial and external forecasts); while $\text{SHIFT} < 0.0$ means that the adjusted prediction is closer to the initial forecast (but outside the interval of initial and external forecasts). SHIFT scores can be used to examine the positioning of point forecasts, as well as the positioning of the lower and upper bounds of prediction intervals.

Table 1A presents the mean SHIFT scores for all these forecasts, while Table 1B shows the percentage of cases in each category where the participant's initial forecast was the same as the advice that was subsequently provided. *t*-tests were carried for each of the five estimates to compare the mean percentage of forecasts per participant where the initial forecast and the advice were the same between the expert and statistical framing groups. The *t*-tests had *p*-values ranging from 0.094 to 0.803 and therefore were not significant. The similarity of the percentages for the two groups suggests that our comparisons between the position measures were not seriously affected by the exclusion of these cases. Moreover, the forecast or the limits of the prediction intervals were adjusted in the opposite direction to that suggested by the advice in only 1.4% of all cases.

As shown in Table 1A, on average, both groups gave adjusted forecasts that were positioned closer to their initial forecasts (i.e. all mean scores < 0.5). However, a multivariate analysis of variance (MANOVA) on the five SHIFT scores revealed that there was a significant main effect due to group ($F_{1,74} = 12.53$, $p = .001$), with participants in the 'expert-forecast framing' group adjusting their forecasts closer to the external forecast provided (as compared to 'statistical-forecast framing' group participants). Indeed, in the 'expert-forecast framing' group, the mean SHIFT towards the advised values exceeded the 'token 20–30%' observed by Harvey and Fischer (1997). Thus, although the advised values received by both groups were identical, they had a much greater influence when participants thought that these values came from a human expert rather than from a statistical method.

Table 1A. Study 1: SHIFT scores for point forecasts and for lower bounds (LB) and upper bounds (UB) of 70 and 95% prediction intervals (PI)

	SHIFT: Point forecasts	SHIFT: LB of 70%PI	SHIFT: UB of 70%PI	SHIFT: LB of 95%PI	SHIFT: UB of 95%PI
Expert forecast framing group	0.390	0.404	0.403	0.455	0.454
Statistical forecast framing group	0.280	0.286	0.306	0.343	0.326

Table 1B. Study 1: Percentage of cases where the initial forecast was the same as the advice

% of cases where participants' initial forecast agreed with advice	Point forecasts	LB of 70% PI	UB of 70% PI	LB of 95% PI	UB of 95% PI
Expert forecast framing group	13.3	11.8	11.1	7.9	12.0
Statistical forecast framing group	15.5	10.4	11.5	9.6	10.4

Was the advice worth having?

A secondary question relates to the usefulness of the advice provided. As shown in Table 2, the advised point forecasts, which had a mean absolute percentage error (MAPE) of only 5.46%, were significantly more accurate than the initial point forecasts of the expert-framing group ($t_{37} = 16.8$, $p < .001$) and those of the statistical-framing group ($t_{37} = 18.2$, $p < .001$). This was still the case after the point forecasts had been adjusted (for the expert-framing group: $t_{37} = 7.1$, $p < .001$ and the statistical-framing group: $t_{37} = 10.4$, $p < .001$). However, because those in the 'expert-forecast framing' group made adjustments that set their revised point forecasts closer to the advice, they gained the greatest reduction in their MAPE ($t_{71} = 2.5$, $p = .016$). One possibility is that the expert framing group simply had the least accurate initial forecasts and hence had the most to gain from the advice, but Table 2 shows that the MAPEs of the initial point forecasts of the two groups were very similar and the difference is not significant.

A number of measures are available for assessing performance of the 70 and 95% interval forecasts (see Yaniv & Foster, 1995). We report only hit rates here (other measures tell a similar story). Hit rates address the percentage of intervals containing the realized values, and thus provide a global measure of interval forecasting performance. Once again the 'expert-forecast framing' group benefited from their greater use of the advice. Table 3 summarizes the hit rates of the participants in the two experimental groups. Both groups started with overconfident intervals (i.e. hit rates much lower than the corresponding confidence percentages of 70 and 95%), and adjusted their intervals to attain improved calibration (i.e. hit rates closer to the corresponding confidence percentages). The improvements in the hit rates were significantly greater than zero for both groups and both prediction intervals (all p -values $< .001$ and all values of $t_{37} > 5.0$). However, the participants in the 'expert-forecast framing' group showed a bigger improvement in their hit rates after adjustment for both the 70% interval ($t_{73} = 2.2$, $p = .035$) and the 95% interval ($t_{73} = 2.5$, $p = .016$).

There is an indication in Table 3 that the expert framing group had the worst initial hit rates for both the 70% ($t_{73} = -2.0$, $p = .046$) and 95% intervals ($t_{73} = -2.0$, $p = .049$) though the significance tests should be treated with care given that the p -values only just suggest significance and the two sets of hit rates are likely to be dependent. Thus the expert framing group appeared initially to be more overconfident and hence have the most to gain from the advice. We address this further in the 'Discussion' Section.

Post-experiment questionnaire

The wrap-up questionnaire given after the experiment focused on the participants' trust in the advice provided and their perceptions of their own performance. When asked to evaluate their overall knowledge

Table 2. Study 1: MAPE values for the initial versus adjusted point forecasts

	Initial forecast MAPE (%)	Adjusted forecast MAPE (%)
Expert forecast framing group	8.14	6.55
Statistical forecast framing group	7.95	6.86
External forecast MAPE		5.46

Note: Mean absolute percentage error (MAPE) = $\{ \sum |(\text{forecast error}/\text{realized value}) \times 100| \} / \{ \text{number of forecasts made} \}$.

Table 3. Study 1: Hit rates for the initial versus adjusted 70 and 95% prediction intervals (PI)

	Initial 70% PI hit rate	Adjusted 70% PI hit rate	Initial 95% PI hit rate	Adjusted 95% PI hit rate
Expert forecast framing group	.51	.60	.76	.87
Statistical forecast framing group	.57	.63	.81	.88
External forecast hit rate		.60		.93

about stock prices they gave mean response of 3.4 (on a scale from 1 = none to 7 = excellent) with a standard deviation (SD) of 1.3. When asked to evaluate their practical understanding of probability they gave a mean response of 4.4 on the same scale with a SD of 1.2.

O'Neill (2002) has argued that people's stated trust in sources of advice is sometimes inconsistent with the level of trust they reveal through their behaviour. When the participants in this experiment were asked to rate their overall trust in the forecasts given by the statistical techniques and the experts on a scale of 1 (no trust at all) to 7 (complete trust), they indicated a significantly higher level of trust in the experts (mean scores were 5.0 for the experts and 4.7 for the statistical techniques, $t_{75} = 2.3$, $p = .024$), and this corresponded with their forecasting adjustment behaviour. However, it is worth noting that despite its statistical significance, the difference between the mean ratings is small. Moreover, when asked to evaluate the forecasting performance of the experts and statistical techniques, the respective mean scores of 5.1 and 4.8 were not significantly different ($t_{74} = 1.9$, $p = .07$). This suggests that participants may have exhibited a greater preference for experts' forecasts than they were prepared to state. Within the two groups there were no significant correlations between the participants' mean SHIFT and the level of expressed trust in the source of the forecasts received (for expert-framing group: $r = 0.11$; for statistics-framing group: $r = 0.25$).

To what extent did the responses to the questionnaire provide explanations for the differences in the way the advice was used between participants? For each group we examined the correlations between the mean extent to which the participants adjusted towards the advice for the point forecasts and their responses to the questions. In almost all cases these correlations were very low. For example, for point forecasts the correlation of participants' self-rating of their knowledge of stock prices and their mean SHIFT were low for both groups ($r < 0.06$). Thus there was no evidence that participants took more notice of the advice when they considered themselves to be less expert in stock prices. Moreover, participants who perceived the point forecasting task to be more difficult did not shift towards the advice to a greater or lesser extent (for expert-framing group: $r = 0.13$; for statistics-framing group: $r = 0.03$). There was also no significant correlation between the accuracy of the initial forecasts and the use of advice. For the point forecasts the correlation between participants' MAPEs and their mean SHIFT was only 0.43 ($p = .218$). Thus participants whose point forecasts were initially less accurate did not tend to shift more towards the advice. Overall, this suggests that people's actual use of advice did not reflect their explicit questionnaire responses relating to the nature of the task, their trust in the advice and their own *perceived* level of expertise. This contrasts with Yaniv (2004b) who found that advice utilization could be predicted from participants' expertise. However, in Yaniv's study expertise was formally assessed by posing a series of general knowledge questions. In the current study it was self-assessed.

STUDY 2

The first study involved providing advice from a single source. In many practical forecasting contexts, advice is received from multiple sources. In these situations forecasters will have to compare the reliability of the advice from the various sources and combine these multiple pieces of advice with their own judgment. Compared to situations where there is a single source of advice, this will increase the cognitive load placed on the forecaster (Bonaccio & Dalal, 2006) particularly when the advice varies between the sources or when prediction intervals have to be estimated. When there are disagreements between the advisors, these cognitive demands might cause the forecaster to simplify the task by focussing predominantly on the favoured source of advice. The second study was designed to investigate this possibility by assessing the relative attention that is paid to 'statistical' and 'expert' advice when both types of advice are available simultaneously.

Participants

The participants were 54 second and third-year business and economics students who were taking statistics and forecasting courses at Bilkent University. Participants' ages ranged between 19 and 22 with a 52% female

($n = 28$) and 48% male ($n = 26$) gender breakdown. There were no monetary incentives; but participation led to extra credit points in courses.

Procedure

The procedure was similar to that of Study 1. Again, a pencil and paper task was used in which the participants were given time series plots for the weekly closing prices of 30 stocks (the same stock price series as in Study 1). Participants were informed that these were stock price series with undisclosed stock names and concealed time periods. They were randomly distributed to three groups receiving the same set of time-series (with a different randomized ordering for each participant). For each stock in turn, the participants were

- (1) asked to make a one-week-ahead point forecast, followed by a 70% prediction interval and a 95% prediction interval;
- (2) requested to examine two given sets of external forecasts which were presented simultaneously (two sets of point forecasts, 70% prediction intervals, and 95% prediction intervals)—where participants in Group 1 ('statistical-forecasts framing' group; $n = 18$) were led to believe that they were presented with forecasts given by two statistical models; whereas those in Group 2 ('expert-forecasts framing' group; $n = 18$) were told that they were given predictions generated by two financial experts, and those in Group 3 ('expert forecast-statistical forecast framing' group; $n = 18$) were told that the first set of forecasts emanated from a financial expert and the second set was obtained from a statistical model;
- (3) asked to use their judgment to adjust the forecasts (if deemed necessary or pertinent), thus providing a judgmentally adjusted point forecast, 70 and 95% prediction intervals in light of the presented external forecasts.

Each participant was presented with a booklet containing the instructions, the forms (see Appendix B for a typical elicitation form), and a wrap-up questionnaire. Instructions were discussed and detailed examples were given in the beginning of experimental sessions. Regardless of the labelling for their groups, all participants were in fact given external forecasts whereby the first forecast set was given by a financial expert and the second set was given by a statistical model (via Holt's exponential smoothing). Accordingly, Group 3 was the only group receiving the correct information on the source of external forecasts.

RESULTS OF STUDY 2

Use of advice

In order to assess the weights the participants placed on each source of advice, the following regression model was estimated for each individual participant for each of the five tasks (i.e. adjusting the initial point forecast and adjusting the initial upper and lower limits of the two prediction intervals). This meant that 270 (i.e. 54×5) individual models were estimated (since each participant made estimates for 30 series the experimental design involved repeated measures (see Lorch & Myers, 1990)).

$$\begin{aligned} \text{Adjusted estimate} - \text{Initial estimate} = & \alpha + \beta_1(\text{First advised value} - \text{Initial estimate}) \\ & + \beta_2(\text{Second advised value} - \text{Initial estimate}) + \varepsilon \end{aligned}$$

where ε is a random error.

The coefficients β_1 and β_2 are intended to show the extent to which an individual's adjustment to their initial estimate is dependent on the discrepancy between a given piece of advice and this initial estimate. Positive values for the β_1 or β_2 would indicate that, on average, an individual's adjustments were in the direction suggested by the advised value. The reliability of the estimates of the β_1 and β_2 would be reduced if they were subject to multicollinearity (Bonaccio & Dalal, 2006). There was no evidence that this was the case—typically correlations between these variables were low and the mean level across all regressions was

0.21. The models also generally explained a significant percentage of the variation in the dependent variable: their median R^2 value was 73%.

Table 4 shows the mean values of the estimates of β_1 and β_2 for the three treatments. The estimated values of β_1 for the different treatments and tasks were compared using a 3 between participants (treatments) \times 5 within participants (tasks) ANOVA model. The mean values of β_1 differed significantly between the treatments ($F_{2,51} = 34.0, p = .024$) but there were no significant differences between the five tasks and there was no significant [task \times treatment] interaction. An ANOVA applied to the estimates of the β_2 found no significant main effects or interactions (for the main effect of treatment: $F_{2,51} = 0.61; p = .55$).

It can be seen that the weight placed on the discrepancy between the first advised value and the initial estimate was significantly higher when participants believed that this advice came from an expert and the second advice came from a statistical model. Indeed, the mean value of β_1 is around 60% higher when this is the case compared to when the sources are perceived to be the same. Thus the simultaneous availability of statistical advice seems to have caused the participants to increase substantially the attention they paid to the advice of the expert—a reaction that did not occur when they believed the second piece of advice also came from an expert. Interestingly, as Table 4 shows, when two ostensibly statistical forecasts were provided to participants they attracted as much weight as two ostensibly expert forecasts. Table 4 also shows that for Groups 1 and 2 a higher mean weight was applied to the second set of advice. The difference is significant on a paired t test $t_{35} = -2.88, p = .007$, but it should be noted that the test was suggested by the data. We investigate this further in the ‘Discussion’ Section.

Accuracy

As was found in Study 1, the adjusted point forecasts were more accurate (i.e. had lower MAPE scores) than the initial forecasts for all three groups ($F_{1,51} = 73.41, p < .001$), hence the advice was beneficial (see Table 5). However, no significant performance differences could be found between the three groups despite the fact that the statistical model forecasts performed worse than the expert forecasts for this set of real stock-price series ($t_{29} = -2.95, p = .006$).

The estimates of the prediction intervals produced similar results to those for the point forecasts (see Table 6). Hit rates improved considerably after the adjustments ($F_{1,51} = 14.74, p < .001$ for 70% PI; $F_{1,51} = 26.94, p < .001$ for 95% PI), but for both the 70 and the 95% intervals, no significant differences could be found between the hit rates of the three groups (for the 70% interval: $F_{2,51} = 0.03; p = .98$, for the 95% interval: $F_{2,51} = 0.48; p = .62$).

Post-experiment questionnaire

The post-experiment questionnaire indicated that the participants’ typical trust in the forecasts of experts and statistical models was almost identical to that of the participants in Study 1 (e.g. their mean levels of trust in experts and statistical methods were 5.1 and 4.7, respectively, as opposed to 5.0 and 5.1 in Study 1). They also indicated almost the same mean levels of knowledge about stock prices (3.3, with a SD of 1.3) and probability (4.5, with a SD of 1.1) as the participants in Study 1. (Study 1 participants had mean responses of 3.4, and 4.4, respectively.)

Analysis of the questionnaire responses revealed that participants who perceived that they had received *both* sets of advice from experts rated their forecasting performance as being significantly higher than

Table 4. Study 2: Mean weight placed on discrepancy between initial estimate and first and second advised values

	β_1 Mean estimated values	β_2 Mean estimated values
Statistical forecasts framing group	0.192	0.278
Expert forecasts framing group	0.183	0.241
Expert forecasts-statistical forecasts framing group	0.307	0.229

Table 5. Study 2: MAPE values for the initial versus adjusted point forecasts

	Initial forecast MAPE (%)	Adjusted forecast MAPE (%)
Statistical forecasts framing group	8.65	7.55
Expert forecasts framing group	8.85	7.66
Expert forecast-statistical forecast framing group	8.19	6.86
External statistical forecast MAPE		9.72
External expert forecast MAPE		5.46

Table 6. Study 2: Hit rates for the initial versus adjusted 70 and 95% prediction intervals (PI)

	Initial 70% PI hit rate	Adjusted 70% PI hit rate	Initial 95% PI hit rate	Adjusted 95% PI hit rate
Statistical forecasts framing group	.41	.49	.71	.79
Expert forecasts framing group	.52	.59	.77	.85
Expert forecast-statistical forecast framing group	.51	.55	.73	.83
External statistical forecast hit rate		.23		.60
External expert forecast hit rate		.60		.93

participants in the other two groups ($F_{2,48} = 3.88$, $p = .027$). However, self-assessment of forecasting performance had no significant correlations with their actual performance (all correlations had an absolute value of less than 0.28 and none were significant). It may be that a perception that one has access to multiple experts boosts unwarranted confidence in one's forecasts. Apart from that no significant differences were found between the mean responses of the participants in the different groups.

Participants' self-assessed knowledge of stock prices was significantly inversely correlated with their point forecast accuracy adjustment (the correlation between self-assessed knowledge and the MAPEs of the adjusted point forecasts = 0.273, $p = .045$). Self-assessed knowledge was also inversely correlated with both the initial and adjusted hit rates (e.g. the correlation between knowledge and initial hit rate for the 70% interval was -0.343 , $p = .011$). Thus those who thought themselves to be more knowledgeable tended to be less accurate.

DISCUSSION

The main finding of Study 1 was that forecasters receiving the same piece of advice from a single source discounted it less when they thought that it came from a human expert rather than a statistical model. Because the advice was beneficial, those who thought they were receiving expert advice produced the most accurate forecasts and the best calibrated prediction intervals. This finding that people treat identical advice in different ways if they perceive its source to be different, even when it is delivered in an identical manner, is important. In many practical situations a preference for human advice may be explained by (i) the rationale that accompanies the human advice, (ii) the opportunity to question the human advisor, or (iii) the fact such human advice is typically delivered orally and hence demands attention. This study has shown that, even when these advantages of the human expert are removed, people still prefer advice from this source. Furthermore, the allure of the expert applies even in a domain (i.e. stock market forecasting) where there is no unequivocal evidence that experts perform consistently well.

Study 2 showed that, when advice came from two sources, the attention paid to the expert forecast was increased when participants were told that the second forecast came from a statistical model. This was in direct contrast to the weights observed for the groups who thought that both forecasts came from the same

type of source. When this was the case, Table 4 shows that participants tended to place a higher weight on the second set of advice. Why was this? Recall that both sets of advice were provided simultaneously so this was not a recency effect. It is possible that, when the two sources of advice were perceived to be the same, people tended to attach a greater weight to the advice that was closest to their initial value. This would be consistent with other studies (Yaniv, 2004b; Yaniv & Milyavsky, 2007) which have shown that the weighting attached to advice is greater, the shorter its distance from a person's initial judgment. For Groups 1 and 2, the second set of advice was closest to the initial value for 51.5% of the estimates. In addition, while the initial estimates of Groups 1 and 2 were on average (using the median) 2.1% higher than the first source of advice, they were on average only 1.4% above the second source so there is some, albeit relatively weak, evidence to support an explanation based on relative distance. There was, however, no evidence to support other explanations, such as a preference for the symmetrical prediction intervals of the statistical method over the asymmetrical prediction intervals of the expert. In addition, the first and second sets of advice were also almost evenly balanced in the widths of their confidence intervals, in the proximity of their point forecasts to the last observation and in the extent to which they predicted an increase or decrease from this observation. Similarly, a re-run of the regressions which allowed the weights to be estimated while the absolute percentage error of the points forecasts was controlled, led to only very small changes in the estimated weights. One possibility suggested by a referee was that the participants did not believe the sources that were presented to them and made their own assumptions about the sources of the forecasts based on their characteristics. We have no evidence for this and it would not explain the switching of the weights when heterogeneous sources were presented.

Whatever the cause, our results suggest that the participants, faced with the cognitively difficult task of attempting to integrate 10 advised values from two heterogeneous sources with their five initial values, used a lexicographic heuristic. If the sources were different they favoured the expert advice over the statistical. If the sources were the same, they favoured the second set of forecasts for some reason. This reason was clearly less important than the preference for expert forecast over statistical and was discarded when the forecast sources were mixed.

Study 1 provided evidence that the perceived source of advice can have an influence before an initial forecast is made. In all cases, participants were informed in advance of the apparent source of the advice they would be receiving, a situation which is likely to apply in many practical contexts (Gönül, Önkal, & Goodwin, 2009). This may account for the inferior calibration of the initial prediction intervals of the expert-framing group in Study 1. Participants anticipating expert advice may have taken less care with their initial interval estimates because of a sense of shared responsibility for forecasting with the expert (e.g. see Önkal, Gönül, & Lawrence, 2008; Promberger & Baron, 2006). Alternatively, they may have simply taken less care, anticipating that good advice was on the way, which would probably lead to a modification of their initial estimates anyway.

Given the results of the first study, it was surprising that the results of Study 2 did not find more advice discounting when both advised values ostensibly came from statistical methods as opposed to experts. There are several possible explanations here. It may be that multiple advice from similar sources increases the salience of the advice (relative to the individual's initial forecasts), thereby escalating the attention that is paid to the advised values, so that relatively less attention is paid to the source of such advice. Alternatively, two advised values that are not too dissimilar may increase confidence in the advice so that discrepancies in the relative trust associated with different sources are negated. Also the forecasting tasks in Study 2, involving two sources of advice was clearly more complex than the tasks in the first study. Schrah et al. (2006) found that people discount advice less when the task is more complex, and it may be that in such situations people are equally prepared to use expert or statistical model-based advice. Given these possibilities, future research could systematically encourage different levels of trust and/or systematically control levels of task difficulty to examine their specific effects on taking advice from multiple sources.

In addition, there was some evidence in our studies that the effect of advice can operate in two stages when the source of future advice is known. First, the anticipation of the advice may influence the effort that is put

into the initial judgment with less effort being made if it is thought that reliable advice is due to be received. Second, the source of the advice will clearly influence the adjustments that are made after the advice has been offered. An experiment which randomly mixed sources of advice, which were known to the judge in advance, with those that were only revealed after the initial judgment had been made would allow the relative size of these effects to be determined.

Other findings of these two studies are consistent with earlier work and demonstrate that these apply in the context of stock price point forecasting and in the estimation of stock price prediction intervals. In both studies, the use of advice led to increases in accuracy and, for the prediction intervals, improved calibration by reducing overconfidence. Also, consistent with the findings of a large number of studies in other contexts (e.g. Harvey & Fischer, 1997; Lim & O'Connor, 1995; Yaniv, 2004b; Yaniv & Kleinberger, 2000), substantial advice discounting was evident despite the fact that the participants indicated that their knowledge of stock prices was only moderate (recall that their mean self-ratings were only 3.3 and 3.4, respectively on a 1 (no knowledge) to 7 (excellent knowledge scale). This would have been expected to have reduced the extent of egocentric advice discounting (Harvey & Fischer, 1997).

CONCLUSIONS

In forecasting tasks people often receive advice from human experts or statistical methods. When offered advice from a single source, they take greater account of the advice when they think it has been provided by a human expert rather than a statistical method. This applies even when the different attributes of human and statistical advice, such as mode of delivery and opportunity for interaction, are removed so that both forms of advice are identical. It even applies in a domain where there is no clear evidence that experts' forecasts are any better than chance. When advice is available from two similar sources, the differential treatment of expert and statistical advice seems to disappear. However, when one of the sources is thought to be a human expert while the other is believed to be a statistical method, much greater attention is devoted to advice from the expert source.

The findings from this research have important practical implications for those practitioners who provide advice in stock markets and in wider forecasting and decision-making contexts. They also have implications for the role of forecasting support systems in organizations because such systems are generally used to produce forecasts based on statistical methods. In particular, the results may, at least in part, account for the tendency in many companies to override statistical forecasts and replace them with forecasts based on management judgment even when such interventions are shown to reduce accuracy (Fildes, Goodwin, Lawrence, & Nikolopoulos, in press). Our belief in the judgment of human experts appears to be deeply rooted and may be motivated by different considerations that moderate our personal decisions like justifiability (Yates, 1990) or our individual psychological needs, such as a need to believe that the world is predictable and controllable (Tetlock, 2005). Whatever the cause, the task of persuading people to give an equal or greater weight to the output of statistical methods is likely to be a difficult one.

ACKNOWLEDGEMENTS

This research was conducted when Dilek Önköl was an academic visitor at the Department of Economics, Oxford University with financial support provided from The Scientific and Technological Research Council of Turkey (TÜBİTAK) research grant 2219. Earlier versions of this work were presented at the *London Judgment and Decision Making Group Seminar Series*, *Lancaster Centre for Forecasting Seminars*, *Durham Business School Seminar Series* and *Oxford Economics Department Econometrics Workshops*. The authors thank all the participants for their constructive comments and suggestions.

**APPENDIX A
A SAMPLE FORECAST ELICITATION FORM FOR STUDY 1**

SHEET 1 FOR STOCK XX

Stock price

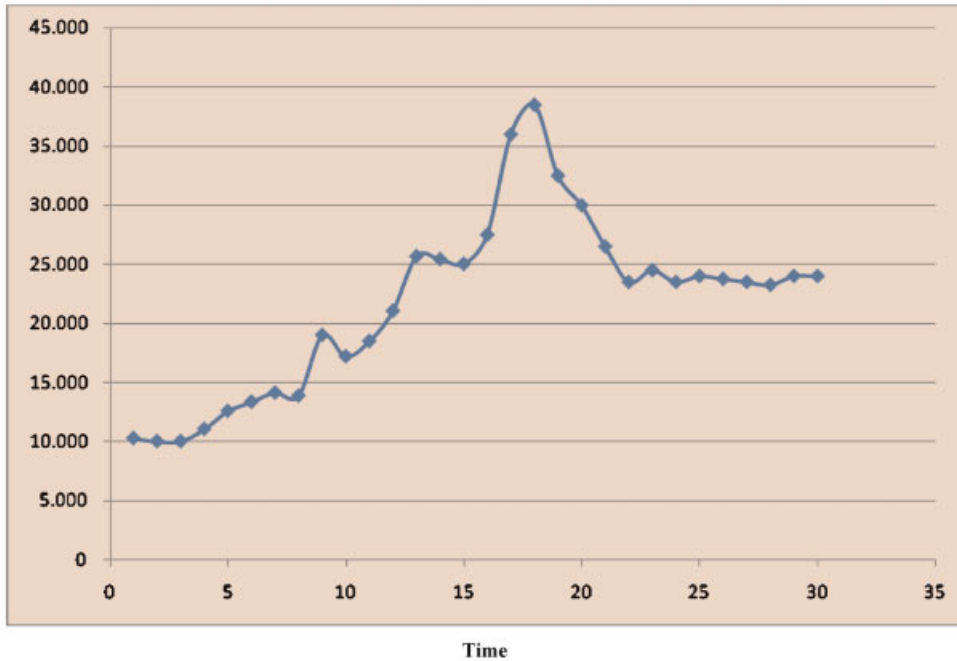


FIGURE 1: Time series plot for Stock XX

OBSERVED VALUES FOR THE LAST FOUR FRIDAY'S CLOSING PRICES:

<u>TIME PERIOD</u>	<u>CLOSING PRICE</u>
t=27	23.500
t=28	23.250
t=29	24.000
t=30	24.000

*MY POINT FORECAST FOR NEXT FRIDAY'S CLOSING PRICE (t=31): _____

*I AM 70% CONFIDENT THAT THE NEXT FRIDAY'S CLOSING PRICE WILL BE BETWEEN
 _____ (LOWER BOUND) AND _____ (UPPER BOUND)

*I AM 95% CONFIDENT THAT THE NEXT FRIDAY'S CLOSING PRICE WILL BE BETWEEN
 _____ (LOWER BOUND) AND _____ (UPPER BOUND)

SHEET 2 FOR STOCK XX

HERE ARE THE FORECASTS GIVEN BY A STATISTICAL TECHNIQUE REGARDING THIS STOCK'S CLOSING VALUE FOR NEXT FRIDAY ($t=31$):

POINT FORECAST: _____

70% PREDICTION INTERVAL: _____

95% PREDICTION INTERVAL: _____

WE WOULD LIKE FOR YOU TO THINK ABOUT YOUR PREDICTIONS FOR THIS STOCK'S CLOSING VALUE AND ALSO TO CAREFULLY EXAMINE THE FORECASTS GIVEN ABOVE. THEN, WE REQUEST THAT YOU ADJUST/REVISE YOUR PREDICTIONS (IF YOU THINK YOU NEED TO) CONSIDERING THE FORECASTS GIVEN ABOVE USING EACH OF THE THREE FORMATS (i.e., point forecast, 70% prediction interval, and 95% prediction interval).

PLEASE WRITE DOWN YOUR FINAL FORECAST ADJUSTED/REVISED IN LIGHT OF THE STATISTICAL TECHNIQUE'S FORECASTS FOR EACH OF THE THREE FORMATS BELOW.

MY ADJUSTED FORECASTS IN LIGHT OF THE FORECASTS PRESENTED ABOVE:

ADJUSTED POINT FORECAST : _____

ADJUSTED 70% PREDICTION INTERVAL : _____

ADJUSTED 95% PREDICTION INTERVAL : _____

**APPENDIX B
A SAMPLE FORECAST ELICITATION FORM FOR STUDY 2**

SHEET 1 FOR STOCK YY

Stock price

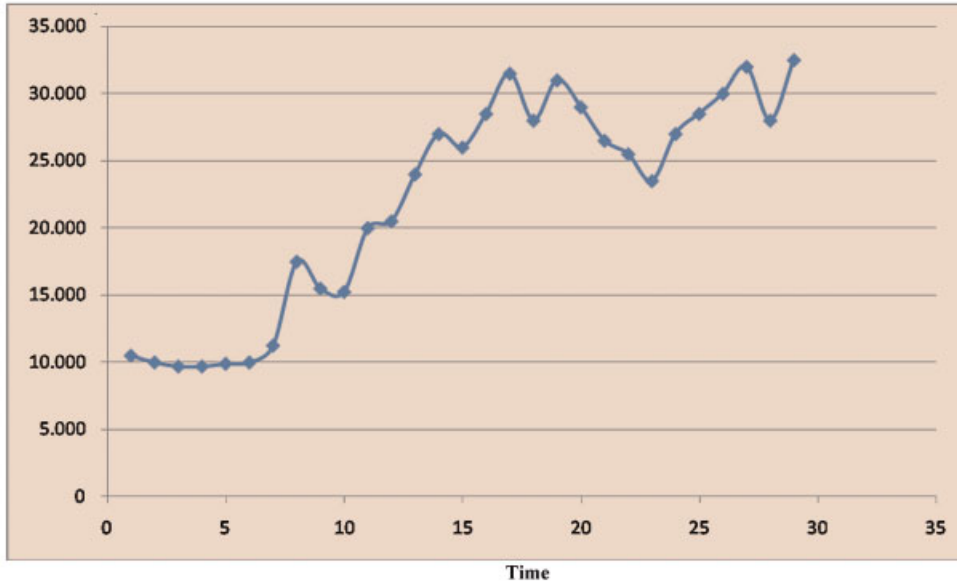


FIGURE 2: Time series plot for Stock YY

OBSERVED VALUES FOR THE LAST FOUR FRIDAY'S CLOSING PRICES:

<u>TIME PERIOD</u>	<u>CLOSING PRICE</u>
t=27	30.000
t=28	32.000
t=29	28.000
t=30	32.500

*MY POINT FORECAST FOR NEXT FRIDAY'S CLOSING PRICE (t=31): _____

*I AM 70% CONFIDENT THAT THE NEXT FRIDAY'S CLOSING PRICE WILL BE BETWEEN
 _____ (LOWER BOUND) AND _____ (UPPER BOUND)

*I AM 95% CONFIDENT THAT THE NEXT FRIDAY'S CLOSING PRICE WILL BE BETWEEN
 _____ (LOWER BOUND) AND _____ (UPPER BOUND)

SHEET 2 FOR STOCK YY

HERE ARE THE FORECASTS GIVEN BY TWO EXPERTS REGARDING THIS STOCK'S CLOSING VALUE FOR NEXT FRIDAY (t=31):

	<u>EXPERT A</u>	<u>EXPERT B</u>
POINT FORECAST:	_____	_____
70% PREDICTION INTERVAL:	_____	_____
95% PREDICTION INTERVAL:	_____	_____

WE WOULD LIKE FOR YOU TO THINK ABOUT YOUR PREDICTIONS FOR THIS STOCK'S CLOSING VALUE AND ALSO TO CAREFULLY EXAMINE THE FORECASTS GIVEN ABOVE. THEN, WE REQUEST THAT YOU ADJUST/REVISE YOUR PREDICTIONS (IF YOU THINK YOU NEED TO) CONSIDERING THE FORECASTS GIVEN ABOVE USING EACH OF THE THREE FORMATS (i.e., point forecast, 70% prediction interval, and 95% prediction interval).

PLEASE WRITE DOWN YOUR FINAL FORECAST ADJUSTED/REVISED IN LIGHT OF THE EXPERTS' FORECASTS FOR EACH OF THE THREE FORMATS BELOW.

MY ADJUSTED FORECASTS IN LIGHT OF THE FORECASTS PRESENTED ABOVE:

ADJUSTED POINT FORECAST : _____

ADJUSTED 70% PREDICTION INTERVAL : _____

ADJUSTED 95% PREDICTION INTERVAL : _____

REFERENCES

- Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review*, 83, 18–24.
- Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, 8, 129–148.
- Blattberg, R. C., & Hoch, S. J. (1990). Database models and managerial intuition: 50% model + 50% manager. *Management Science*, 36, 887–899.
- Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes*, 49, 188–207.
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision making: An integrative literature review, and implications for organizational sciences. *Organizational Behavior and Human Decision Processes*, 101, 127–151.
- Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, 104, 371–398.
- Budescu, D. V., Rantilla, A. K., Yu, H., & Karelitz, T. K. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, 90, 178–194.
- Clement, R. W., & Krueger, J. (2000). The primacy of self-referent information in perceptions of social consensus. *British Journal of Social Psychology*, 39, 279–299.
- Cote, J., & Goodstein, J. (1999). A breed apart? Security analysts and herding behavior. *Journal of Business Ethics*, 18, 305–314.
- Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica*, 1, 309–324.
- Dana, J., & Thomas, R. (2006). In defence of clinical judgment and mechanical prediction. *Journal of Behavioral Decision Making*, 19, 413–428.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668–1774.
- Dunning, D., & Hayes, F. (1997). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology*, 71, 213–229.
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14, 54–74.
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (in press). Effective forecasting for supply-chain planning: An empirical evaluation and strategies for improvement. *International Journal of Forecasting*.
- Goldsmith, D. J., & Fitch, K. (1997). The normative context of advice and social support. *Human Communication Research*, 23, 454–476.
- Goodwin, P. (2000). Improving the voluntary integration of statistical forecasts and judgment. *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.
- Gönül, M. S., Önkal, D., & Goodwin, P. (2009). Expectations, use and judgmental adjustment of external financial and economic forecasts: An empirical investigation. *Journal of Forecasting*, 28, 19–37.
- Green, K. C., & Armstrong, J. S. (2007). The ombudsman: Value of expertise for forecasting decisions in conflicts. *Interfaces*, 37, 287–293.
- Grove, W. M., & Lloyd, M. (2006). Meehl's contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology*, 115(2), 192–194.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy and Law*, 2, 293–323.
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, 70, 117–133.
- Jungermann, H., & Fischer, K. (2005). Using expertise and experience for giving and taking advice. In T. Beetsch, & T. Haberstroh (Eds.), *The routines of decision making*. Mahwah, NJ: Lawrence Erlbaum.
- Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin*, 25, 585–594.
- Kleinmütz, B. (1990). Why we still use our heads instead of formulas: Towards an integrative approach. *Psychological Bulletin*, 107, 296–310.
- Krueger, J. (1999). Lake Wobegon be gone! The 'below-average effect' and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77, 221–232.
- Krueger, J., & Stanke, D. (2001). The role of self-referent and other-referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin*, 27, 878–888.
- Lawrence, M., & O'Connor, M. (1995). The anchoring and adjustment heuristic in time series forecasting. *Journal of Forecasting*, 14, 443–451.

- Lawrence, M., & Sim, W. (1999). Prototyping a financial DSS. *Omega: International Journal of Management Science*, 27, 445–450.
- Lim, J. S., & O'Connor, M. (1995). Judgmental adjustment of initial forecasts - its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149–168.
- Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 149–157.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- O'Neill, O. (2002). *A question of trust*. Cambridge: Cambridge University Press.
- Önkal, D., Gönül, M. S., & Lawrence, M. (2008). Judgmental adjustments of previously-adjusted forecasts. *Decision Sciences*, 39, 213–238.
- Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making*, 19, 455–468.
- Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). The adaptive decision maker is not an island: Integrative expert advice with information search. *Journal of Behavioral Decision Making*, 19, 43–60.
- Shanteau, J. (1992). The psychology of experts: an alternative view. In G. Wright, & F. Bolger (Eds.), *Expertise and decision support*. New York: Plenum.
- Sniezek, J. A., & Buckley, T. (1995). Choice accuracy and confidence in Judge-Advisor decision-making systems. *Organizational Behavior and Human Decision Processes*, 62, 159–174.
- Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgment with prepaid expert advice. *Journal of Behavioral Decision Making*, 17, 173–190.
- Tetlock, P. E. (2005). *Expert political judgment. How good is it? How can we know?* Princeton: Princeton University Press.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Wærn, Y., & Ramberg, R. (1996). People's perception of human and computer advice. *Computers in Human Behavior*, 12, 17–27.
- Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes*, 69, 237–249.
- Yaniv, I. (2004a). The benefit of additional opinions. *Current Directions in Psychological Science*, 13, 75–78.
- Yaniv, I. (2004b). Receiving other people's advice: Influence and benefit. *Organizational Behavior & Human Decision Processes*, 93, 1–13.
- Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124, 424–432.
- Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making*, 10, 21–32.
- Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, 83(2), 260–281.
- Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, 103, 104–120.
- Yates, J. F. (1990). *Judgment and decision making*. New Jersey: Prentice-Hall, Inc.

Authors' biographies:

Dilek Önkal is a professor of Decision Sciences at Bilkent University, Turkey. She received a PhD in Decision Sciences from the University of Minnesota, and is an Associate Editor of the *International Journal of Forecasting* as well as the *International Journal of Applied Management Science*. Professor Önkal's research focuses on judgmental forecasting, forecasting support systems, probabilistic financial forecasting, risk perception, and risk communication. Her work has appeared in several book chapters and journals such as *Journal of Behavioral Decision Making*, *Organizational Behavior and Human Decision Processes*, *Decision Sciences*, *Risk Analysis*, *Decision Support Systems*, *International Journal of Forecasting*, *Journal of Forecasting*, *Omega: The International Journal of Management Science*, *Foresight: The International Journal of Applied Forecasting*, *Frontiers in Finance and Economics*, *International Federation of Technical Analysts Journal*, *Journal of Business Ethics*, *Teaching Business Ethics*, *International Forum on Information and Documentation*, *Risk Management: An International Journal*, and *European Journal of Operational Research*.

Paul Goodwin is professor of Management Science at the University of Bath. His research interests are concerned with the role of management judgment in forecasting and decision making. He is a Director of the International Institute of Forecasters, an Associate Editor of the *International Journal of Forecasting* and a member of the editorial board of the

Journal of Behavioral Decision Making. He also writes a column in *Foresight: The International Journal of Applied Forecasting* that is aimed at practitioners.

Mary E. Thomson is currently a Reader in the Division of Psychology, Glasgow Caledonian University. She was awarded her PhD in 1999 and has since published in various books and journals, including *Decision Support Systems*, *European Journal of Operational Research*, *Risk Analysis*, *European Journal of Finance*, *Risk Management: An International Journal*, *International Journal of Forecasting*, *Frontiers in Finance and Economics* and *IMA Journal of Management Mathematics*.

M. Sinan Gönül joined the Department of Business Administration at TOBB University of Economics and Technology, Turkey, after receiving his PhD in Management from Bilkent University, Turkey. In 2006–2007, he held an academic visitor position at Glasgow Caledonian University, UK. He was also a participant in the International Management Program at McGill University, Canada in 2000. His research interests focus on judgmental forecasting, judgment and decision making, decision support systems, and econometric forecasting. His work has appeared in *Journal of Forecasting*, *Decision Support Systems*, and *Foresight: International Journal of Applied Forecasting*.

Andrew C. Pollock is a Reader at Glasgow Caledonian University. He completed a PhD on exchange rates and has published articles and papers in a variety of books and journals in the financial area. His particular research interest is the application of analytical techniques to financial forecasting involving judgemental and statistical approaches.

Authors' addresses:

Dilek Önkal, Bilkent University, Faculty of Business Administration, 06800 Ankara, Turkey.

Paul Goodwin, School of Management, University of Bath, Bath BA2 7AY, U.K.

Mary E. Thomson, Division of Psychology, Glasgow Caledonian University, Cowcaddens Road, Glasgow, G4 0BA, UK.

M. Sinan Gönül, Department of Business Administration, TOBB University of Economics and Technology, 06560 Ankara, Turkey.

Andrew C. Pollock, Division of Mathematics, School of Communications, Networking and Electronic Engineering, Glasgow Caledonian University, Cowcaddens Road, Glasgow, G4 0BA, UK.