Judgmental forecasting: A review of progress over the last 25 years

Michael Lawrence [a],*, Paul Goodwin [b], Marcus O'Connor [c], Dilek Önkal [d]

[a] *University of New South Wales, Sydney, Australia*
[b] *University of Bath, Bath, UK*
[c] *University of Sydney, Sydney, Australia*
[d] *Bilkent University, Ankara, Turkey*

Abstract

The past 25 years has seen phenomenal growth of interest in judgemental approaches to forecasting and a significant change of attitude on the part of researchers to the role of judgement. While previously judgement was thought to be the enemy of accuracy, today judgement is recognised as an indispensable component of forecasting and much research attention has been directed at understanding and improving its use. Human judgement can be demonstrated to provide a significant benefit to forecasting accuracy but it can also be subject to many biases. Much of the research has been directed at understanding and managing these strengths and weaknesses. An indication of the explosion of research interest in this area can be gauged by the fact that over 200 studies are referenced in this review.
© 2006 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

Keywords: Judgement; Forecasting; Review; Improving judgement forecasts; Probability forecasts; Domain knowledge; Prediction intervals

1. Introduction

While judgement has always played an important role in forecasting, academic attitudes to the role and place of judgement have undergone a significant transformation in the last 25 years. It used to be commonplace for researchers to warn against judgement (e.g. Hogarth & Makridakis, 1981), but there is now an acceptance of its role and a desire to learn how to blend judgement with statistical methods to estimate the most accurate forecasts. The forecasting practitioner has never shared the scepticism of the researcher towards judgement. It is generally recognized that without management judgement in forecasting, serious problems can result. Worthen (2003) describes Nike's $400 million experiment with forecasting software which went disastrously wrong leading to massive inventory write-offs due to the system's inaccuracy and lack of management input. Worthen claims that "corporate America is littered with companies that invested heavily in demand software but have little or nothing to show from it". Good forecasting requires

* Corresponding author.
E-mail addresses: michael.lawrence@unsw.edu.au
(M. Lawrence), mnspg@management.bath.ac.uk (P. Goodwin),
M.Oconnor@econ.usyd.edu.au (M. O'Connor),
onkal@bilkent.edu.tr (D. Önkal).

that management judgement play its role and, equally important, that there be effective implementation of the forecasting systems (Fildes & Hastings, 1994). It is also important that the goal of the forecasting be clearly defined. The cost of lost sales and excess inventory are rarely equal and their costs fall on different organisational units. This often leads to different units having differing forecasting goals (Lawrence, O'Connor, & Edmundson, 2000).

The poor corporate experience of forecasting software claimed by Worthen may be partly responsible for the recent Sanders and Manrodt (2003) finding from a large survey of 240 US corporations, that only 11% reported using forecasting software. And of those who did use forecasting software, 60% indicated they routinely adjusted the forecasts based on their judgement. Thus, understanding the proper use of judgement is more than ever an important activity for researchers and practitioners.

It may be expected that judgement would play an important role in company sales forecasting where the impacts of promotions and competitor activity, generally known or anticipated by marketing staff, can be built into the forecasts. But judgement also plays an important role in macro-economic forecasting (Batchelor & Dua, 1990; Clements, 1995; Fildes & Stekler, 2002; McNees, 1990; Turner, 1990). Fildes and Stekler (2002) in their review of macroeconomic forecasting, summarise their findings by stating that "the evidence unequivocally favours (judgmental) interventions".

In Fig. 1, we show the steps in forecasting, say, the sales of a product. We propose viewing the total set of data useful for forecasting as made up of two classes; the history data and the domain or contextual data. The history data are the history of the sales of the product. The domain data are in effect all the other data which may be called on to help understand the past and to project the future. This includes past and future promotional plans, competitor data, manufacturing data and macroeconomic forecast data. The data usually input to a forecasting decision support system are the history data and occasionally promotion data. The adjustment review process is informed by both the history data and all the non-history data.

In this review of the past 25 years of research into judgmental forecasting, we have divided the field along the lines of Fig. 1. We first consider judgmental
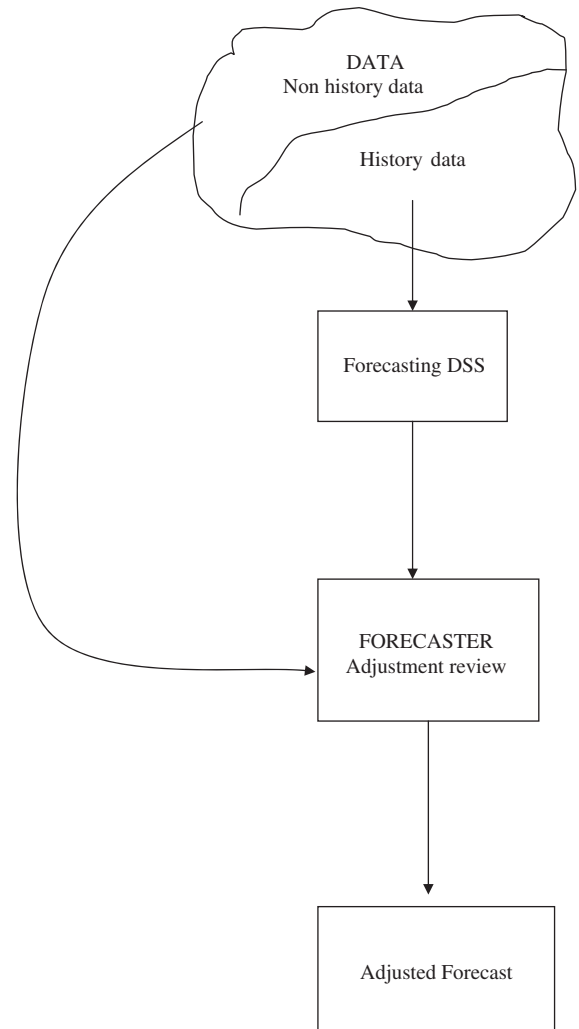

Fig. 1. Forecasting steps.

forecasting of a time series with no domain or contextual knowledge. Under this restriction, if we compare a judgmental forecaster with a quantitative model, as both are limited to the same data set, we gain a fair comparison of the strengths of each mode of forecasting. We then move to examine the influence of domain knowledge on the judgmental forecaster. Here we are specifically looking to see how the judgmental forecaster may use non-time series information to improve the forecast. Up to this stage in the review we have restricted ourselves to examining point forecasts. In the following section, we examine the research contribution aimed at investigating probabilistic or interval forecasts. Finally, we consider what research

has revealed over the last 25 years about how the role of judgement in forecasting can be improved.

2. Judgmental (point) forecasting without domain knowledge

This section reviews judgmental point forecasting[1] under the restriction that the judgmental forecaster has no domain knowledge.[2] This is, in practice, a most unlikely situation as a judgmental forecaster will almost always have some information about the value to be forecasted in addition to the time series values. However, it does form a useful basis for comparison with statistical methods as the two methods are both restricted to the same data set and thus are on a "level playing field."

Hogarth and Makridakis (1981), in a major review, analysed over 175 papers concerned with forecasting and planning and concluded without any hesitation that "quantitative models outperform judgmental forecasts" (p. 126). Furthermore, judgement was characterised as being associated with systematic biases and large errors, the tendency to see patterns where none exist, the illusion of control even when the underlying process is purely random and excessive and unfounded confidence in its correctness.

However, almost none of the cited studies involved judgement applied to time series forecasting where the cues are serially correlated. Most were psychological laboratory experiments using general knowledge questions (e.g. which is longer, the Suez Canal or the Panama Canal and how confident are you in this judgement), serially uncorrelated cues or simple gambles based on given probabilities. Hammond's Social Judgement Theory (Hammond, Stewart, Brehmer, & Steinmann, 1975) stressed the role of the task in influencing the effectiveness of judgement. The properties of the task can either help or hinder the application of judgement and can influence the judge's ability to acquire skill. Stewart, Roebber and Bosart (1999) and Lawrence and O'Connor (1996) have further demonstrated the importance of the task in studies of human judgement. In addition, the ecological validity of the task is an important

dimension influencing human expertise (Bolger & Wright, 1994). Thus, the conclusions of studies based on non-time series data may not apply to time series forecasting (Fischhoff, 1988).

In this section, we review the studies done in this area over the last 25 years. We first look at some forecasting competition comparisons of judgement and forecasting methods which have demonstrated the skill of non-experts to judgmentally extrapolate a time series, before examining research aimed at exploring why people are good at this task. We then proceed to investigate the influence of various data and task characteristics on the accuracy of the judgmental forecaster, including the organisation of the forecasting effort as a group activity.

2.1. How accurate is judgmental point forecasting?

Early comparisons of judgmental forecasting with statistical methods mostly used artificial data and reached varying conclusions about the relative accuracy of the two methods (Adam & Ebert, 1976; Eggleton, 1982; Lawrence, 1983). The first large-scale comparison of the accuracy of judgmental forecasting and quantitative model forecasting using real life data was performed by Lawrence, Edmundson and O'Connor (1985, 1986). Their study followed what has now been called the M1 forecasting competition (Makridakis et al., 1982), which compared the accuracy of most of the widely available forecasting models on a set of 1001 real-life time series comprising annual, quarterly and monthly series (some seasonal and some non-seasonal) drawn from a variety of domains including stock market, sales, demographic and financial. The competition concluded that simple methods outperformed more complex methods, with deseasonalised single exponential smoothing coming out in front. Lawrence et al. (1985) forecasted a subsample of 111 of the 1001 series using two alternative methods, a graphical method and a table method, with each method being applied by two sets of forecasters comprising the researchers themselves and around 200 undergraduate business students who each forecasted one series. The study concluded that it had "demonstrated judgmental forecasting to be at least as accurate as statistical techniques, while in a number of subgroupings of the time series a judgmental technique was the most accurate". In addition, the standard

---

[1] That is, where the forecast is the single most likely value.

[2] That is, only the set of time series data is available.

deviation of the judgmental forecast errors was uniformly less than for the statistical methods suggesting a greater consistency in their accuracy; specifically, the standard deviation of the table forecast errors averaged over all series was half that of the deseasonalised single exponential smoothing forecasts. The judgmental forecasts were less correlated with the model forecasts than the model forecasts were with each other, resulting in greater gains by combining a judgmental with a statistical forecast than two statistical forecasts (Lawrence et al., 1986). This appeared to suggest the value of an approach which combined the judgmental and statistical approaches.

A later forecasting competition (Makridakis et al., 1993) added further support to these conclusions. On the other hand, using a subsample of 10 series from the M1 competition, Carbone and Gorr (1985) concluded that judgement was less accurate than the statistical forecast. Clearly, the particular nature of the time series task is very important. This was further illustrated by Sanders (1992), who constructed 10 time series to simulate monthly data, each series of 60 periods. She found judgement to produce forecasts that were biased and less accurate than statistical methods. With artificial series it is to be expected that a statistical method, based on an assumption of a stable generating function, should perform better in a comparison with judgement. This is especially true as human judges anticipate change and instability, even when the underlying generating function is stable (Lawrence & Makridakis, 1989). Sometimes the judgmental forecasters' biases have been shown to be rational (e.g. De Bondt, 1993), while in other studies they have been shown to be irrational and resulting in suboptimal performance (e.g. Anderson & Goldsmith, 1994; Moore, Kutzberg, Fox, & Bazerman, 1999).

We opened this subsection with the question "how accurate is judgmental forecasting", and summarise the evidence as suggesting that they can be as good as the best statistical techniques and may have greater consistency in their accuracy, but this is not assured.

2.2. Why are people good at forecasting?

We have already alluded to a number of factors that distinguish time series forecasting from the laboratory tasks that have mostly been used in establishing the rich literature on human judgement. Two factors are particularly important in influencing accuracy. Firstly, the task has high ecological validity.[3] Bolger and Wright (1994) demonstrated that judgmental performance is "a function of the interaction between the dimensions of ecological validity and learnability". Secondly, the autocorrelated cues can be simply presented in a graph allowing "eye-ball" processing at which humans are relatively skilled (Lawrence and Makridakis, 1989; Mosteller, Siegel, Trapido and Youtz, 1981). Using policy capturing methods to uncover the methods and techniques used by the judgmental forecaster, Andreassen and Kraus (1990) and Lawrence and O'Connor (1992, 1995) demonstrated that judgmental forecasting can be modelled as single exponential smoothing, or alternatively as anchor and adjustment, where the anchor point is the average of recent time series values and the adjustment is the proportion of deviation of the most recent value from this average. When modelled as exponential smoothing, the judgmental forecaster appears to use a value of the smoothing constant dependent on the characteristics of the series. As the forecast horizon increases, less emphasis is placed on the last observation. Each of these characteristics is appropriate for achieving accuracy. Thus one can conclude that judgmental forecasting accuracy is the product of a good subjective model being applied. However, these results again depend on the characteristics of the series and the presentation of the task. But as Goodwin and Wright (1993) pointed out, although a good fit was obtained, a wide range of alternative models was not investigated. Highlighting the contingent nature of the time series and the task presentation, Harvey, Bolger and McClelland (1991), using a strongly cyclical series with low levels of noise and a tabular presentation of information (rather than graphical), did not observe behaviour matching exponential smoothing.

2.3. The influence of data characteristics

To understand the ability of the judgmental forecaster to respond to various time series characteristics including trend, seasonality, randomness and discontinuities, researchers have conducted laboratory

---

[3] Ecological validity is the extent to which the data and the task setting correspond to the real world situation that is under study.

studies using, in general, artificial series so as to wash out all but the intended manipulation. While some practical principles can be drawn from the wealth of research done on this topic, the conclusion of Goodwin and Wright (1993) is still valid, that much of the evidence is contradictory due to the difficulty of characterising a time series and the influence of apparently quite minor changes in the series and in the presentation of the task.

One of the consistent findings, however, is that subjects damp both up and down trends, with downtrends damped more than up-trends (Eggleton, 1982; Lawrence & Makridakis, 1989; O'Connor, Remus, & Griggs, 1997). Lawrence and Makridakis (1989) presented subjects with a plot of data "evenly" distributed around a trended line and observed that subjects damped both the up and down trends, and suggested this indicated a commonsense view about the behaviour of economic time series. They also observed that subjects seemed less sure of down trends as they both widened their confidence bounds for these series and damped their most likely estimates more than for up trends. O'Connor et al. (1997) confirmed the difficulty presented by down trends and that the forecaster's behaviour suggests an anticipation of a reversal in slope. Bolger and Harvey (1993) found that people employed different heuristics for trended and untrended series and their approach to trended series depended on the extent of serial correlation.

However when trend is confounded with randomness, Lawrence and Makridakis (1989) and Mosteller et al. (1981) found no impact of randomness, while Andreassen and Kraus (1990) found that noise did impact the subject's ability to detect the trend. These different conclusions probably relate to the fact that Lawrence and Makridakis and Mosteller et al. both used a graphical display while Andreassen and Kraus provided only a tabular display. When the trend is not linear, judgmental extrapolations become significantly biased (Timmers and Wagenaar, 1977; Wagenaar and Sagaria, 1975), but these biases could be due to different beliefs about the nature of the series being forecasted. Hence, we may expect that a judgmental forecaster may encounter difficulty in forecasting more complex trends such as new product sales or other growth processes.

Goodwin and Wright (1993) argue that the complexity of a series includes three components: (1) the underlying signal, comprising its seasonality, cycles and trends and response to shocks; (2) the level of noise around the signal; and (3) the stability of the underlying signal. O'Connor, Remus, and Griggs (1993) investigated the impact of series instability by presenting subjects with a simple series exhibiting a major discontinuity at a certain point. The subjects estimated the next period forecast on a rolling basis after each new actual was given. The researchers expected that the human judge would be able to detect the occurrence of a discontinuity earlier than a statistical method, but found the opposite to be true. Although in many studies of judgement people have tended to respond to randomness as if it was signal (Andreassen, 1988; Harvey, 1995; Lopes & Oden, 1987), O'Connor et al. (1993) found that their subjects ignored the discontinuity signal for far too long. That is, they mistook the signal for randomness. On the other hand, Lawrence (1983), Edmundson, Lawrence and O'Connor (1988), Sanders (1992) and Sanders and Ritzman (1992), using less artificial series, all found that judgmental extrapolation outperformed or equalled statistical projection for more unstable or more volatile series. The time series forming the basis of the forecasting tasks in these papers are all quite different, again suggesting that judgements are sensitive to small differences. The presence of high and complex seasonality or a strongly cyclical component confuses judgement (Harvey et al., 1991; Lawrence & O'Connor, 1993).

In a series of two studies investigating how well the judgmental forecaster responds to information reliability, Remus, O'Connor, and Griggs (1995, 1998) showed that correct information leads to forecast accuracy while incorrect or unreliable information did not seem to have a big effect except at a turning point. However, people were not able to make good use of the information provided and estimated forecasts of lower accuracy than statistical methods.

2.3.1. Mode of task presentation

Although most of the laboratory studies on judgmental forecasting have used a graphical mode of presentation,[4] there is no clear evidence that

---

[4] Mode of presentation refers to how the task is presented to the subjects in the experiment.

uniformly supports its superiority over a table format (Desanctis & Jarvenpaa, 1989; Harvey & Bolger, 1996). Trends are better estimated from a graphical presentation but these seem to encourage inconsistency and overforecasting when compared to tabular format. In addition, Desanctis and Jarvenpaa (1989) only achieved incremental performance with subjects using graphs when they provided training.

2.4. Expertise

There has been much conflicting evidence on the existence of an "inverse expertise effect" such that novices perform equal to or better than experts (Önkal & Muradoglu, 1994; Thomson, Pollock, Henriksen, & Macaulay, 2004; Wilkie-Thomson, Önkal-Atay, & Pollock, 1997). It would appear that subtle changes in the presentation of the task can act as a mask preventing the expert exercising expertise. As most experts rely on domain knowledge this issue will be explored further in the later sections.

2.5. Impact of loss function shape

While the majority of studies examining judgmental forecasting have either explicitly or implicitly used a symmetric loss function for forecast errors, many forecasters operating in the sales and marketing field have expressed the view that their loss function is asymmetric, and demonstrate biases in their forecasts consistent with the reported asymmetry (Lawrence et al., 2000). Lawrence and O'Connor (2005) examined, in a laboratory experiment, the response of subjects to two alternative shapes of loss functions with each shape presented at three different symmetry/asymmetry settings. They found that their research subjects (business students at a large university) were able to respond appropriately to the different directions of the asymmetry and to the different kinds of shapes of the loss functions. However, Goodwin (2005), in an experiment for examining the value of providing support to the judgmental forecaster operating under conditions of asymmetric loss observed poor performance for the unaided judgmental forecaster. Once again, minor differences in the task may be responsible for the observed differences. Goodwin's experimental task used artificial time series and

differed in a number of other ways from Lawrence and O'Connor (2005).

2.6. Forecasting as a group activity

Almost all the research has examined individuals making forecasts, while most forecasting activity is undertaken by a group (Lawrence et al., 2000). A few studies have examined the influence of the group dimension by comparing techniques of group interaction. Ang and O'Connor (1991) and Sniezek (1989, 1990) concluded that the group does produce more accurate forecasts than simply averaging the individual pre-group judgements; there are some differences as to which structuring approach worked best, perhaps reflecting differences in the task. When all group members had access to the same information, Sniezek (1990) suggested the choice of group technique was less important. Possible light is shed on the optimism of many operational forecasts by Brenner, Griffin, and Koehler (2005). In a time prediction task, they found that predictions generated through group discussion were more optimistic than those generated individually. Group discussion acted to focus subjects' attention on those factors promoting "success", thus encouraging their optimism.

3. The influence of domain knowledge on judgmental forecasting

The previous section focused on the ability of people to engage in the task of time series extrapolation. In general, people do reasonably well, although they suffer from a number of cognitive traps and illusions. Nevertheless, it must be said that it is quite uncommon in practice for people to be faced with a time series they know nothing about. Typically, people will be aware of both the nature of the time series and its associated context. They might also be aware of information that is associated with the time series—e.g. some 'special' information that may help to explain the past behaviour of the time series or is likely to have some major impact in the future. This section addresses our understanding of the ways in which people utilise such non-time series information.

3.1. Towards an understanding of domain knowledge

We define domain knowledge as any information relevant to the forecasting task other than the time series—i.e., non-time series information. In its very simplest case, this could be an awareness of the nature of the time series. For example, knowledge that a downward trending time series is related to daily interest rates rather than sales of a product may (arguably) contain valuable information that may change the way in which a person forecasts the variable. Armstrong (1985) demonstrated this principle in a time series context, confirming the finding from studies using other tasks, that the context of a judgement is highly salient to performance (Adelman, 1981; Sniezek, 1986). At another level, domain knowledge could also involve the provision of causal information that is associated with the time series. For example, since new housing loan approvals may relate to new housing construction, knowledge of the former may assist with the latter. Lim and O'Connor (1996), in a laboratory study, examined the ability of people to use this causal information that is associated with a time series and compared it with their use of statistical forecasts (which were only time series based). They found that knowledge and use of this causal information contributed significantly to final forecast accuracy—even though it may have been far from optimal. Other studies (Andreassen, 1991; Harvey, Bolger, & McClelland, 1994) reported that people may not be as good as Lim and O'Connor (1996) suggest. Nevertheless, this situation probably does not describe a typical judgmental forecaster in a typical business related forecasting environment.

The most common situation in which a judgmental forecaster finds himself/herself is one where there is both contextual knowledge (knowledge of the nature of the time series) and some additional irregular knowledge that can be useful in either explaining the past behaviour of the series or in predicting the future (or both). Sometimes the impact of this information can be quite minor, but sometimes it can be quite important and can have a major influence on the behaviour of the variable. In any event, the distinguishing characteristic of such 'domain knowledge' is that it represents an 'unmodelled' component of the time series. This 'unmodelled' component is a necessary characteristic—if it is capable of being modelled by a statistical method, it

can be incorporated into the statistical forecast. For example, knowledge that a fixed and known marketing promotion will occur regularly in a particular month would not represent domain knowledge, since such regularity could be modelled as a seasonal component by a statistical method.

In the sections that follow, we review two contexts (earnings per share forecasting and sales forecasting) in which there has been substantial research into the impact of domain knowledge. These two fields provide some guidance on how we should effectively deal with such knowledge.

3.2. Earnings or earnings per share forecasting

A comparative analysis of various approaches to forecasting earnings per share (EPS) has a long history. In summarising this literature in 1983, Armstrong concluded that management forecasts of EPS (judgment) were more accurate than analysts' forecasts (judgement), which were more accurate than those of statistical models. The substantial domain (inside) knowledge that management possessed enabled their accuracy. Much of this research into EPS was undertaken by Brown. In commenting on it, Brown (1996) concluded that EPS forecasts produced by analysts and management achieved very high accuracy (within 3% of actual). Furthermore, this was almost always of substantially higher accuracy than the best available statistical models—in fact it did not matter whether simple or complex statistical models were used. He concluded that people should place far greater emphasis on these analysts' judgmental forecasts than they seem to—a conclusion that was also reached by Chatfield, Hein, and Moyer (1990).

There seem to be two sources/reasons for the analysts' superior forecasting accuracy. First, they possess better information and this may explain a considerable portion of the non-modelled component of the variance. This is the common interpretation of domain knowledge. But they also possess more timely information—information that has come to light after the last history point in the time series has been recorded and made available. Thus, their forecasts are based on more up-to-date data. At the time of release of the latest data, statistical forecasts do well: but they become progressively less accurate as time passes. Alexander (1995) showed that the superiority of the

analysts' forecasts came from both better sources of information and more up-to-date information. The value of the domain knowledge encapsulated in the analysts' and management forecasts is reflected in the impact that such forecasts have on the market prices. As Asquith, Mikhail, and Au (2005) showed, there is substantial information content in the management and analysts' forecasts when they are released. Furthermore, Clement and Tse (2005) have shown that the market reacts mostly to 'bold' forecasts and that these bold forecasts tend to be the most accurate. Ivkovic and Jegadeesh (2005) also found that the revisions that analysts' make to their previous forecasts have a positive effect on the market. The point here is that there seems to be a strong component of domain knowledge in analysts' forecasts and that this is considered useful by the market. However, in some situations domain knowledge may not be the prime cause of this superior forecast accuracy—managers may be able to exert control over the earnings they are forecasting and hence produce highly accurate EPS forecasts (Beneish, 2001; Brown, 1996; Holthausen & Leftwich, 1983; Watts & Zimmerman, 1990).

3.3. Sales forecasting

There are two main groups of studies that have examined the role of domain knowledge in sales/ product forecasting. In the first group are those forecasts that have been produced without any major input from a statistical forecast model, where the forecast has simply been the product of an individual or consensus judgement. One of the first studies to examine this was Edmundson, Lawrence, and O'Connor (1988). They examined the judgmental forecasting process at a large consumer products corporation. The unique aspect of this study was that they were able to assess the relative contribution of two levels of domain knowledge. The product forecasts were made by a consensus of people including the product manager, the marketing manager, and a finance representative; but they were mainly 'driven' by the product manager. In this study, these 'domain knowledge' laden forecasts were compared with two benchmarks—the first was the judgmental forecasts produced by other product managers in the same organisation where the time series details (including scale values) were hidden from them, and the second

benchmark was a statistical forecast. Analysis was also made in terms of the 'key' and 'non-key' products. Surprisingly, the product managers forecasting the current company products (but not their own products) were only as accurate as the statistical forecast. So, industry domain knowledge had no influence on accuracy. But for the key products, the consensus meeting forecasts were substantially and significantly more accurate than all the benchmarks. However, this did not occur for the non-key products. This suggests that intimate product knowledge was a major contributor to accuracy. The content of that 'intimate' product knowledge was typically about individual promotions, distribution outlets, competitor actions, sourcing or raw materials, etc.

A broader, though not as intensive, study of the contribution of domain knowledge in product forecasting was reported by Lawrence et al. (2000). The forecast accuracies of thirteen organisations were studied and compared with simple benchmarks (e.g. the naïve forecast). On first analysis, the accuracy of the domain knowledge laden company forecasts was mixed: some organisations were clearly better and some worse than the naïve forecasts. However, they observed that these forecasts were the outcome of organisational budgeting and incentive issues, which tended to bias the forecasts; e.g. if the forecast was used as a target to be achieved, the forecasts tended to be below the actual. When this organisational bias was (statistically) removed from the forecast, they tended to be significantly and substantially more accurate than the naïve benchmark—thus confirming the results from Edmundson et al. (1988) that domain knowledge was a major contributor to forecast accuracy.

The second main group of studies in sales forecasting are those where the domain knowledge is incorporated in conjunction with a statistical forecast, the so-called judgmental adjustment studies. While it is true that some work on judgmental adjustment to statistical forecasts has been done in fields other than sales forecasting (McNees, 1975, 1990), the majority of the studies have been in the context of sales forecasting. As Mathews and Diamantopolous (1986, 1989, 1990) demonstrated, the revisions of statistically generated forecasts using relevant domain knowledge enabled greater final forecast accuracy. These revisions are particularly

important given the tendency for judgmental adjustments without domain knowledge to impair final accuracy (Willemain, 1991). Sanders and Ritzman (1992, 1995) reinforced these conclusions and also showed that the greatest advantage of such adjustments to statistical forecasts were in conditions of high variability—where the un-modelled component was relatively large. A comprehensive study of product forecasting in a UK-based household consumer products company was recently undertaken and reported by Nikolopolous, Fildes, Goodwin, and Lawrence (2005). Statistical forecasts were produced for all products and these were adjusted in about half of the cases. The adjustments for domain knowledge were overall beneficial, but were most advantageous when large adjustments were made. When small adjustments were made, they seemed to be less than useful—perhaps reflecting the tendency to 'tinker at the edges'.

What is the main benefit of domain knowledge in aiding forecast accuracy? Like the EPS forecast accuracy, there appear to be three reasons for it. First, there may be a 'timing' advantage—the domain-rich forecasts are normally produced with very up-to-date knowledge, while the statistical methods may lag in data availability. While this may appear to be an important issue for EPS forecasts, we are unsure as to its influence in sales/product forecasts. Second, there is clearly an advantage that can be attributed to the domain knowledge which represents the un-modelled component. Intimate knowledge of marketing campaigns, competitor actions, raw material sourcing, etc. may be quite important. Certainly, if the nature of the discussion at the consensus meetings is any guide to importance (and that may be debatable), this may be a powerful reason for its greater accuracy. Finally, in a similar way to the EPS forecasting situation reviewed earlier, there is an element of self-serving bias in the domain forecasts. But in the case of sales forecasting, it may be an even more powerful determinant of comparative forecast accuracy advantage. In the course of one such meeting reported in Lawrence et al. (2000), substantial marketing expenses were committed to arrest declining sales with the comment that sales needed to reach a certain level. In another case, sales levels were manipulated by promotional spending to ensure that sales targets were met. One major multi-national commented that they always

ensure that sales are within 0.5% of the forecast—they made sure it happened that way! Thus, in some situations management can influence the accuracy of the forecasting process by the manipulation of the 'drivers' of the forecast variable. To our knowledge, this aspect of the sales forecasting process has been largely ignored by researchers.

4. Judgmental probability forecasts and prediction intervals

Point forecasting is not the only format for providing judgmental forecasts. Especially in domains like economics and finance, users of forecasts may specifically demand information about providers' uncertainties surrounding the given predictions (Tay & Wallis, 2002). Probability forecasts and prediction intervals offer two effective formats for explicating such uncertainties, thus prohibiting false assumptions of precision.

4.1. Probability forecasts

Probability forecasts provide an elicitation format whereby subjective probabilities supply the communicative means for facilitating the users' understanding of the 'vagueness' surrounding the presented forecasts, by enabling the forecast provider to give a more complete judgmental portrayal. Probability forecasting is used in various domains like weather forecasting, portfolio analysis, risk analysis, economic forecasting, pharmaceutical forecasting, and technological forecasting (Martino, 2003; Murphy & Winkler, 1984, 1992; Poland & Wada, 2001), with the International Journal of Forecasting dedicating special issues to Probability Forecasting (1995, Vol. 11, No. 1), and to Probability Judgmental Forecasting (1996, Vol. 12, No. 1).

Research in the use of judgmental probabilistic forecasts started in the 1980s with studies showing that the results from probability judgment research using almanac questions have limited applicability to forecasting tasks (Carlson, 1993; Wagenaar & Keren, 1985; Wright, 1982; Wright & Wisudha, 1982). Accordingly, the last 25 years have witnessed a surge of research into evaluating probability forecasts and expert performance, with fewer studies on construct-

ing probability forecasts and understanding user-provider perspectives.

4.2. Evaluating probability forecasts

Various measures addressing distinct aspects of probabilistic forecasting performance were proposed prior to 1980 (Murphy, 1972a, 1972b, 1973; Sanders, 1963). Further research on evaluating probability forecasts exploded during the 1980–2005 period with the development of additional performance measures (Björkman, 1994; Murphy, 1988; Pollock, Macaulay, Thomson, & Önkal, 2005; Wilkie & Pollock, 1996; Yates, 1982, 1988), portrayed via graphical tools (Blattenberger & Lad, 1985; Hsu & Murphy, 1986; Yates & Curley, 1985) and extensively used in assessing predictive performance for stock prices (Muradoglu & Önkal, 1994; Önkal & Muradoglu, 1994, 1995, 1996; Yates, McDaniel, & Brown, 1991), earnings (Whitecotton, 1996; Yates et al., 1991), and exchange rates (Önkal, Yates, Simga-Mugan, & Oztin, 2003; Pollock & Wilkie, 1992, 1993; Wilkie & Pollock, 1994; Wilkie-Thomson, Önkal-Atay, & Pollock, 1997).

In evaluating predictive performance, a significant emphasis has been on to calibration (a measure of the correspondence of forecast probabilities with the realized proportion of correct predictions or with the relative frequencies of occurrence of the predicted event, depending on the task structure used) and over/underconfidence (an index of the provider's probability assessments exceeding or falling short of the attained proportion correct for the corresponding events). While the focus on these performance aspects may be attributed to the appealing comparisons they provide of the forecaster's probabilities with the empirical reality (Lichtenstein, Fischhoff, & Phillips, 1982), their limitations in addressing user needs have also been noted, as discussed below.

Limitations notwithstanding, calibration has been extensively studied using mainly general knowledge questions (see Lichtenstein et al., 1982 and McClelland & Bolger, 1994 for detailed reviews), with the 'demonstrated' overconfidence debated via explanations on misleading item selection (Juslin, 1994), as well as via probabilistic mental models emphasizing ecological validities of predictive cues and the frequentistic assessor (Gigerenzer, Hoffrage, & Klein-bölting, 1991). Acknowledging the limited applicability of these results to forecasting situations, some studies used real prediction tasks to reveal good calibration for weather forecasters' probability of precipitation forecasts (Murphy & Winkler, 1984; Stewart, Roebber, & Bosart, 1999), hockey players' forecasts of future game results (Vertinsky, Kanetkar, Vertinsky, & Wilson, 1986), and experienced bridge players' probability forecasts of making the contracts that they had bid (Keren, 1987). On the other hand, overconfidence and poor calibration are reported for predictions of weather forecasters not trained in probability forecasting (Daan & Murphy, 1982), professional economic forecasters' future recession predictions (Braun & Yaniv, 1992), sports experts' forecasts for the World Cup soccer games (Andersson, Edman, & Ekman, 2005), and the Russian managers' economic forecasts (Aukutsionek & Belianin, 2001). Forecasts of earnings (Davis, Lohse, & Kottemann, 1994), election results (Babad, Hills, & O'Driscoll, 1992), starting salaries and job offers (Hoch, 1985), sports events (Ayton & Önkal, 1996; Carlson, 1993; Peterson & Pitz, 1988), and general events that could happen within a month (Fischhoff & MacGregor, 1982) all demonstrated overconfidence. Signalling the influential role of task characteristics, currency forecasts given by finance professionals showed underconfidence for exchange rate series with strong trends, while displaying overconfidence for series with weak trends (Thomson, Önkal-Atay, Pollock, & Macaulay, 2003). Task format is viewed as a significant factor overall, with a higher overconfidence shown in tasks where the forecaster assigns a probability to a pre-selected outcome, in comparison to tasks where the forecaster first selects one of two possible outcomes and then assigns a probability to his selected outcome (Ronis & Yates, 1987).

Individuals tend to give higher probabilities (expressing more certainty) for forecasts of personal events as compared to impersonal events (Wright & Ayton, 1989). Furthermore, for non-personal events, predictions for desirable events are better calibrated and less overconfident in the immediate time period than the subsequent time periods (Wright & Ayton, 1992). Desirable events are also judged as being more likely to happen to one while the undesirable events

are judged to be more likely to happen to others (Zakay, 1983).

In addition to the anticipated effects of desirability, imminence, and time period (Wright & Ayton, 1986), it is proposed that the perceived controllability of events may affect probability assessments (Langer, 1982; Weinstein, 1980) and their calibration (Wright & Ayton, 1989). Overconfidence may in fact be related to a need to feel in control, failure of imagination, information distortions, and problems in assessing or weighing probabilities (Schoemaker, 2004). Furthermore, perceived controllability of events may be positively correlated with the amount of optimistic bias (Weinstein, 1980). Within areas like financial and economic forecasting, "the combination of overconfidence and optimism is a potent brew, which causes people to overestimate their knowledge, underestimate risks, and exaggerate their ability to control events" (Kahneman & Riepe, 1998, p. 54). Overall, decision makers appear to take an *inside view*, resulting in overly optimistic (and "bold") forecasts; rather than taking an *outside view* that involves a broader and more comparative framing (Kahneman & Lovallo, 1993).

Klayman, Soll, Conzález-Vallejo, and Barlas (1999) propose that variations in findings of over-confidence (between individuals in a given domain as well as differences among domains) call attention to the interaction effects of information processing and information content. Interestingly, confidence is found to increase regardless of the relevance of information for the forecasting task (Davis et al., 1994). Along similar lines, Wright and Ayton (1988) report evidence for *personologism* (where traits or cognitive styles are argued to be the main determinants of performance, as opposed to *situationism*, where environmental factors act as the primary source) in probabilistic forecasting performance, suggesting 'acuity' (number of different probability categories that can be differentiated) as an important performance dimension (Wright & Ayton, 1987). In relation to this, forecast providers' overconfidence can be thought of as expressing 'overprediction' (consistent preference to assign probabilities that are too high), or 'overextremity' (tendency to assess probabilities that are consistently very close to 0 or 1.0), with different implications (Griffin & Brenner, 2004).

Aside from the work on overconfidence, there appear to be almost no studies investigating heuristics and biases in judgmental probabilistic forecasting contexts. Among several notable exceptions are the findings on (i) insensitivity to base rates in probabilistic bankruptcy predictions (Johnson, 1983), and (ii) higher probabilities for forecasts of Apple's earnings given by investors generating supporting reasons (Moser, 1989). There are clearly gaps in research on how the various probability judgment biases apply to forecasting contexts, as well as on the implications of specific biases like the 'recency bias' (forecast influenced by recent events; Hogarth & Makridakis, 1981) and the 'advocacy bias' (overpromising on forecasts) (Tyebjee, 1987) for probabilistic forecasting performance.

Another research issue concerning probability forecasting performance is *consistency*, i.e., the extent of agreement between a forecaster's probabilities assessed at different times but under identical information conditions. Given the infeasibility of restricting the information flow in real forecasting settings, consistency presents an exigent question addressed to date via decomposition techniques (Salo & Bunn, 1995) and integrative frameworks (Pollock, Macaulay, Önkal-Atay, & Thomson, 2002).

4.3. Experts' probability forecasting performance

Investigating the experts' probabilistic predictions has constituted an attractive research stream in the last 25 years, even with the barriers to working in ecologically valid settings. Experts gave better probability forecasts than non-experts in predicting earnings (Whitecotton, 1996), exchange rates (Önkal et al., 2003; Wilkie-Thomson et al., 1997), sports game outcomes (Andersson et al., 2005), and stock prices, with the moderating effects of task format and forecasting horizon (Muradoglu & Önkal, 1994; Önkal & Muradoglu, 1996). Also, the probability forecasts inferred from bookmakers' odds outperformed statistical model predictions in a high-stakes environment (Forrest, Goddard, & Simmons, 2005). On the other hand, studies employing only graduate and undergraduate students revealed that the relative novices (i.e., undergraduate students) performed better in stock price forecasting (Önkal & Muradoglu, 1994; Yates et al., 1991), although the non-expert

participant groups and the task structures (multiple-interval format) used hinder any direct comparisons. Interestingly, even self-rated expertise appears to be a good predictor of probabilistic forecasting performance, with individuals rating themselves as more expert attaining higher proportions of correct predictions, better calibration and less overconfidence (Wright, Rowe, Bolger, & Gammack, 1994).

Information effects remain an ongoing concern for studies with experts. For instance, in predicting soccer-game outcomes, giving additional information to non-experts is found to increase their confidence, while not improving their predictive performance (Andersson et al., 2005; Ayton & Önkal, 1996). Whether incomplete representations of the forecasting problem (Wright & Ayton, 1987) and/or information access affects confidence, presents a challenging research question.

4.4. Constructing probability forecasts

How probability forecasts are constructed is another research issue with direct repercussions for expert performance. It is suggested that a mismatch between the problem structure and the forecaster's internal model may lead to poor probabilities and that restructuring to access the assessor's experience is imperative (Phillips, 1987). Along similar lines, probability forecast construction is viewed as consisting of an initial belief assessment phase followed by a second phase entailing an assessment of a probability qualifying the belief (Benson, Curley, & Smith, 1995; Curley & Benson, 1994; Curley, Brown, Smith, & Benson, 1995; Ferrell & McGoey, 1980; Smith, Benson, & Curley, 1991). Accordingly, belief assessment (dominated by reasoning) and response assessment (dominated by judgment) are thought to represent two distinct phases directly affecting the 'goodness' of probability forecasts; with the former stage (starting with data screening and construction of arguments, and ending with combining the formed arguments) remaining a neglected area carrying significant implications for forecasting. As Evans (1987) points out, economic forecasting is a good example of a field vulnerable to personal and political factors which lead to belief maintenance where "forecasts are frequently wrong ... (and) rival theories are persistently maintained in the face of all evidence" (p. 43).

4.5. User-provider perspectives

Given the users' preference for forecasts with reliable depictions of predictive uncertainties (Murphy, 1998), and the providers' desire to produce 'good' forecasts, provider-to-user communication of judgmental probability forecasts presents a challenging and under-researched topic (Abramson & Clemen, 1995; Önkal-Atay, Thomson, & Pollock, 2002). Studies on the user perspective indicate that the users seem to emphasize performance dimensions other than the typical calibration-overconfidence focus that dominates most of the work in this field. In an interesting experiment, Yates, Price, Lee, and Ramirez (1996) reported that the consumers of forecasts focus particularly on extreme probability (close to 0% and/or 100%) usage. That is, users either inferred the forecaster's competence via the extremity of the assessed probabilities (thinking such high probabilities would not be assigned if the implied certainty was not justified), or concluded that these forecasters knew little about the uncertainties to the extent of not recognizing their recklessness in assigning the high probabilities. Along similar lines, Keren (1997) found that the participants preferred a forecaster who only used 90% probabilities in all cases to a forecaster who exclusively used 75% probabilities, stating that the former forecaster made clear and conclusive predictions whereas the latter forecaster's predictions were not sufficiently decisive. Finding complementary results, Price and Stone (2004) argued that the users of probability forecasts employed a 'confidence heuristic', i.e., they used confidence as a cue to the forecaster's knowledge and competence in making more categorically correct forecasts. Given the users' clear preference for forecasters assigning extreme probabilities, providers' overconfidence could actually be a result of their efforts to impress the users and to 'prove their expertise'. Although future research is needed to improve our understanding of their concerns, it may well be that from the providers' perspective, "...we are encouraged to be confident in what we do. We are constantly reminded that the person with confidence is the person who succeeds. Furthermore, an admission of a lack of confidence is an admission of failure" (O'Connor, 1989, p. 161).

In a different vein, it is suggested that a probability forecast could be flexibly interpreted depending on

the context information that may shape our intuitive reactions (Flugstad & Windschitl, 2003; Teigen & Brun, 1999, 2000; Windschitl, Martin, & Flugstad, 2002). That is, a probability forecast of 15% may be interpreted as an alarm situation or as an infinitesimal chance depending on contextual factors triggering positive or negative reactions.

Irrespective of the contextual contingencies, it is argued that the "probabilistic forecasts mollify the potential for misperception of responsibilities and misattribution of decisions" (Krzysztofowicz, 2001, p. 5) by decoupling the forecasting and the decision-making tasks. Since probability forecasts are found to have greater decision-making value than deterministic predictions for users of forecasts in certain domains (Mlyne, 2002), the communication and interpretation of such forecasts promises to be a potent agenda for multi-disciplinary research.

4.6. Prediction intervals

Prediction intervals, also called 'interval forecasts', consist of prediction bounds that specify upper and lower forecast limits within which the future value of the predicted variable is expected to lie with a specified probability. Although their importance is repeatedly accentuated in domains like weather forecasting (Hamill & Wilks, 1995), economic predictions (Corker, Holly, & Ellis, 1986; Christoffersen, 1998; Clements & Taylor, 2003) and financial forecasting (Önkal-Atay, 1998), there exist fewer studies on judgmental prediction intervals as compared to probabilistic predictions. Evaluating prediction intervals, expert performance, and user-provider perspectives again constitute the core themes for research.

4.6.1. Evaluating prediction intervals

The evaluation of prediction intervals is an important concern given their frequent use with financial time series (Taylor, 1999) as well as in sales forecasting contexts (Dalrymple, 1987). Most of the research on interval forecasts has focused on the effects of time series characteristics and presentation format. Studies show that the prediction intervals appear to be influenced by the trend, seasonality, and variability in the series in addition to the choice of the presentation scale (Lawrence & Makridakis, 1989;

Lawrence & O'Connor, 1993; O'Connor & Lawrence, 1992). In particular, prediction intervals become wider for trended time series (Eggleton, 1982; Lawrence & Makridakis, 1989), with seasonality significantly influencing the interval width (O'Connor & Lawrence, 1992). Furthermore, choice of the presentation scale alters the interval width (Lawrence & O'Connor, 1993), with randomness showing little effect (Lawrence & Makridakis, 1989).

Overall, the predominant finding of the last 25 years is that judgmental prediction intervals exhibit overconfidence (i.e., for intervals given a confidence coefficient of XX%, less than XX% of the intervals actually include the true value) (Lawrence & Makridakis, 1989; Lichtenstein et al., 1982; O'Connor & Lawrence, 1989; Russo & Schoemaker, 1992). When data series has high variability, more overconfidence is observed and reflected via narrower intervals (Lawrence & O'Connor, 1993); however, the interval width does not appear to increase when more skewness is perceived (De Bondt, 1993). Overconfidence appears to be higher when participants have to report a point forecast in addition to the prediction interval (Russo & Schoemaker, 1992), and groups also give interval forecasts that display overconfidence (Ang & O'Connor, 1991). Choice of the confidence percentage seems to matter: although overconfidence is found with 95% prediction intervals, it disappears when the same forecasters are asked to give 50% prediction intervals for the same stock price series (Önkal & Bolger, 2004). Interestingly, when the question format does not entail direct interval forecasting, but rather asks for the probability that the actual outcome will fall within a specified range, probabilities are underestimated and underconfidence is revealed (Bolger & Harvey, 1995; Harvey, 1988).

4.6.2. Experts' interval forecasting performance

Experts' prediction intervals are overconfident according to studies with managers predicting industry-related and firm-related outcomes ((Russo & Schoemaker, 1992), and with finance professionals making currency forecasts (Önkal et al., 2003). Overconfidence is also reported with software professionals' *effort prediction intervals* (i.e., judgmental prediction intervals that reveal the uncertainties in software development effort) (Connolly & Dean,

1997; Jørgensen, Teigen, & Moløkken, 2004). Jørgensen and Sjøberg (2003) propose the following reasons for overly narrow effort prediction intervals: (i) interpretation difficulties (what a given confidence percentage actually means to a professional is unclear), (ii) hidden agendas (narrow interval may be perceived as a sign of software professionals' skill), (iii) narrow intervals facilitate project planning and execution, (iv) point predictions serve as anchors (cannot adjust sufficiently to attain a wider interval), and (v) the lack of meaningful and immediate feedback. While these reasons might be applicable to many contexts, it may also be that, if more evidence bolsters confidence and if experts have greater access to evidence that they may perceive as supporting their predictions, then experts may simply carry a greater overconfidence risk (Arkes, 2001).

4.6.3. User-provider perspectives

Studies reveal a clear user preference for prediction intervals over point forecasts (Baginski, Conrad, & Hassell, 1993; Pownall, Wasley, & Waymire, 1993). Prediction intervals are claimed to provide information which enables the users to better assess future uncertainties, to plan for alternative strategies addressing the range of possible future outcomes, and to compare the predictions obtained from alternative forecasting sources (Chatfield, 2001). Moreover, giving intervals instead of single values is found to enhance decision performance (Johnson, 1982).

How the users actually perceive and employ the narrow prediction intervals assessed by the forecasters remains to be studied in detail. However, a potential explanation for the forecast providers' apparent insistence on giving narrow intervals is provided by the so-called *accuracy-informativeness trade-off* (Yaniv & Foster, 1995, 1997). This view suggests that there exists a trade-off in setting interval bounds since widening the prediction interval increases accuracy (as indexed by high hit rates; i.e., high proportion of intervals including the realized value), but reduces informativeness (precision as indexed by narrowness of the interval width). That is, a forecaster insists on not widening the forecast limits to avoid being non-informative while simultaneously appearing more credible to the users. Also, the rewards for being informative are immediately provided by users (who seem to prefer narrow intervals), whereas

accuracy can only be assessed when the future outcome transpires.

Another signal concerning the relative importance of informativeness over accuracy is conveyed via the asymmetric prediction intervals of forecasters (O'Connor, Remus, & Griggs, 2001). The information that the forecast providers are trying to communicate via prediction intervals may not have anything to do with accuracy concerns, but may rather be focused on providing bounds that will be regarded as 'useful' or 'meaningful' by the forecast users (Bolger & Önkal-Atay, 2004). Hence, the motivation for providing interval forecasts and the motivation for giving point forecasts may be totally different, even leading to potential 'hedging' strategies (with forecasters placing asymmetric bounds to convey the differential risks in either direction of the point forecast) (O'Connor et al., 2001). Overall, the intended use of prediction intervals dictates the assessment concerns. For instance, detailed assessments of the predictive distribution or of the probabilities in certain tail areas gain prominence depending on whether the users are concerned with the probability of incurring a loss (e.g. VAR analysis in finance) or whether they have more asymmetric concerns about comparative losses.

Similar to the findings with judgmental probability forecasts, the use of higher confidence percentages in prediction intervals is regarded as a display of expertise. When allowed to choose a confidence percentage (as opposed to being forced to use an imposed confidence percentage), providers seem to prefer using higher percentages as demonstrations of the extent of their knowledge (Bolger & Önkal-Atay, 2004). Similarly, in using provided sets of prediction intervals, participants would use the 50% intervals only when they are given performance-based incentives (Foong, Lawrence, & O'Connor, 2003). Such direct or implied preferences for higher percentages by both the providers and the users again reinforce the significant role of a confidence percentage as an acknowledged pointer to the providers' expertise.

5. Improving judgmental forecasts

The earlier sections of this paper have demonstrated that we have learned much over the last 25 years about the psychology of judgmental forecasting and

its associated biases, together with its performance relative to statistical methods under different conditions. But what have we learned about how to improve the accuracy of judgmental forecasting? This section looks at the main improvement strategies that researchers have investigated, including provision of feedback, decomposition, combining and correction.

5.1. Feedback

One of the key findings in the last 25 years is that feedback can be valuable because it enables the judgmental forecaster to learn. Feedback presented to the judgmental forecaster can take a number of forms. Outcome feedback is perhaps the most common type encountered in practice and simply involves informing the forecaster of the latest observation in a series. Performance feedback provides information on the accuracy, calibration or bias associated with the forecaster's past forecasts, while in cognitive process feedback the judgmental forecaster is provided with information about the strategy that he or she is adopting to produce the forecasts. For example, such feedback might involve a graphical display of the weights that the forecaster appears to be attaching to the different cues. Finally, task properties feedback provides the forecaster with statistical information about the task (e.g. correlations of possible predictor variables with the forecast variable or details of the underlying time series structure). Note that Björkman, 1972, has argued that "task properties feedback" should not be regarded as feedback. Instead, it is feedforward since it is usually provided before the initial judgment.

Feedback has been shown to improve the accuracy of point forecasts (Goodwin & Fildes, 1999; Remus, O'Connor, & Griggs, 1996; Sanders, 1997; Welch, Bretschneider, & Rohrbaugh, 1998) and the calibration of both probability forecasts (Benson & Önkal, 1992; Murphy & Winkler, 1984; Önkal & Muradoglu, 1995) and judgmental prediction intervals (Bolger & Önkal-Atay, 2004; Goodwin, Önkal-Atay, Thomson, Pollock, & Macaulay, 2004; O'Connor & Lawrence, 1989). However, these studies have tended to show that outcome feedback is the least effective form. This is probably because the most recent outcome contains noise and hence it is difficult for the forecaster to distinguish the error arising from a systematic

deficiency in their judgement from the error caused by random factors (Klayman, 1988). Task properties feedback has generally been found to be the most effective (Balzer, Doherty, & O'Connor, 1989), possibly because, in providing statistical information about the task, it helps forecasters to reject erroneous hypotheses that they are entertaining (Kluger & DeNisi, 1996). Nevertheless, in some contexts different types of feedback may be more appropriate for improving different elements of the forecasting task. For example, Stone & Opel (2000) found that performance feedback was only effective in improving the calibration of probability forecasts, while task properties feedback only improved the forecasters' discrimination (i.e., their ability to distinguish between cases where the target event will occur from those where it will not occur), and actually worsened calibration. Moreover, the relative effectiveness of the different types of feedback is likely to depend closely on the characteristics of the forecasting task and hence vary between tasks (Fischer & Harvey, 1999).

Of course, the value of any type of feedback is also likely to depend on its understandability, timeliness, accuracy and presentation. For example, Lim, O'Connor and Remus (2005) found that, when improvements were needed in a decision-making task, presenting cognitive process feedback via a text message was more effective than presenting it in the form of a multimedia display showing an expert delivering the message. This was apparently because the cognitive resources needed to process the text message were more closely matched to the level of resources required to improve the accuracy of the decisions (Keller & Block, 1997).

5.2. Decomposition

Decomposition methods are designed to improve accuracy by splitting the judgmental task into a series of smaller and cognitively less demanding tasks, and then combining the resulting judgements. Armstrong (2001) distinguishes between decomposition, where the breakdown of the task is multiplicative (e.g. sales forecast = market size forecast × market share forecast), and segmentation, where it is additive (e.g. sales forecast = Northern region forecast + Western region forecast + Central region forecast), but we will use the term for both approaches here. Surprisingly, there has

been relatively little research over the last 25 years into the value of decomposition and the conditions under which it is likely to improve accuracy. In only a few cases has the accuracy of forecasts resulting from decomposition been tested against those of control groups making forecasts holistically. One exception is Edmundson (1990) who found that for a time series extrapolation task, obtaining separate estimates of the trend, seasonal and random components and then combining these to obtain forecasts led to greater accuracy than could be obtained from holistic forecasts. Similarly, Webby, O'Connor and Edmundson (2005) showed that, when a time series was disturbed in some periods by several simultaneous special events, accuracy was greater when forecasters were required to make separate estimates for the effect of each event, rather than estimating the combined effects holistically. Armstrong and Collopy (1993) also constructed more accurate forecasts by structuring the selection and weighting of statistical forecasts around the judge's knowledge of separate factors that influence the trends in time series (causal forces).

Many other proposals for decomposition methods have been based on an "act of faith" that breaking down judgmental tasks is bound to improve accuracy or upon the fact that decomposition yields an audit trail and hence a defensible rationale for the forecasts (Abramson & Finizza, 1991; Bunn & Wright, 1991; Flores, Olson, & Wolfe, 1992; Saaty & Vargas, 1991; Salo & Bunn, 1995; Wolfe & Flores, 1990). Yet, as Goodwin and Wright (1993) point out, decomposition is not guaranteed to improve accuracy and may actually reduce it when the decomposed judgements are psychologically more complex or less familiar than holistic judgements, or where the increased number of judgements required by the decomposition induces fatigue.

5.3. Combining forecasts

A third strategy for improving judgmental forecasts involves combining these forecasts either with statistical forecasts or with other judgmental forecasts. Combination procedures can range from mechanical methods (e.g. taking a simple or weighted average of the constituent forecasts) to the use of judgement to determine how the forecasts should be combined. Combination has been one of the major areas of

forecasting research over the last 25 years (Clemen, 1989), and the method is thought to work because the forecasts being combined draw upon different information sources and hence increase the information upon which the forecast is based. This combination of many independent estimates accounts for the observed accuracy of prediction markets, such as the Iowa electronic markets (Surowiecki, 2004). Indeed, it can be shown that mechanical combinations of forecasts are most effective when the forecasts are negatively correlated (Bunn, 1987; Goodwin, 2000a). The complementary strengths of judgmental and statistical forecasts suggest that combinations of these two methods are likely to be worth considering in many contexts (Blattberg & Hoch, 1990; Lawrence et al., 1986).

How should combination be carried out? Earlier reviews of the literature suggested that mechanical combinations of point forecasts are likely to lead to greater accuracy than those based on judgement (Goodwin & Wright, 1994; Webby & O'Connor, 1996). However, Fischer and Harvey (1999) found that when the judge received performance feedback relating to each of the individual forecasts available for combination, the accuracy of judgmental combination surpassed that of the simple average. If mechanical combination is to be used, de Menezes, Bunn, and Taylor (2000) present guidelines on which approach is likely to be most appropriate, depending on the forecaster's objectives.

Other researchers have investigated the nature of forecasts which should be included in a combination. When judgmental combination is employed, Harvey and Harries (2004) argue that the person combining the forecasts should not include their own forecast in the combination because they are likely to over-weight this. Indeed, they should consider refraining from making their own forecasts at all. When judgmental forecasts are to be mechanically combined with univariate statistical forecasts, Sanders and Ritzman (1995) argue that the judgmental forecast should be based on contextual information (or domain knowledge), especially where the time series has a high degree of variability. When probability forecasts are required, Clemen, Murphy and Winkler (1995) demonstrate a method for screening out forecasting methods that are inferior to other methods or do not add any information to the combination.

5.4. Taking advice

Taking advice from others on what the forecast should be is similar to combining in that the judgmental forecaster faces the task of combining the advice with a prior estimate of the appropriate forecast. Indeed, the 'advice' might be available in the form of a statistical forecast (e.g. Lim & O'Connor, 1995), rather than the recommendation of a human judge. As with combination, advice is likely to be particularly beneficial where it comes from independent sources (Yaniv, 2004). Advice which is highly correlated with the initial forecast or where the multiple sources are themselves correlated is unlikely to improve accuracy.

There is evidence that people tend to attach less weight to the advised forecast than to their own prior estimate, possibly because they have greater access to and belief in the rationale underlying their own view than to the reasons underpinning the advice (Yaniv & Kleinberger, 2000). In addition, the weight they attach to the advice is dependent on the reputation of the advisor, but reputations are more easily lost than gained in that negative information about the advisor is perceived to be more diagnostic than positive information (Yaniv & Kleinberger, 2000). When advice is available from multiple sources, people will also give more weight to advice that they perceive to come from more experienced people (Harvey & Fischer, 1997). Harvey, Harries and Fischer (2000) found that forecasters are also better at assessing the quality of the different sources than they are at using that advice. They therefore suggested that forecasters should be asked only to assess the weights that are appropriate for each source of advice and that these weights should then be used mechanically to combine the estimates from the different sources.

5.5. Bootstrapping and correction

A major finding of the 'general' literature on human judgment has been that the use of a statistical model of how a judge arrives at predictions tends to lead to more accurate predictions than the judge—a process known as (judgmental) bootstrapping (e.g. Dawes, 1975). This occurs because the model 'averages out' the judge's inconsistency. However, time series forecasting tasks differ from those used in

earlier studies of bootstrapping in that there is usually a very large, or even infinite, number of possible cues available to the forecaster, many of which may not be available to the model (Yaniv & Hogarth, 1993). Moreover, some of the cues will be serially correlated or configural. These factors will tend to favour the original judgments relative to the model, and therefore some studies (e.g. see Lawrence & O'Connor, 1996) have cast doubt on the value of bootstrapping in this context.

However, statistical forecasting methods can play another role in improving judgmental forecast accuracy—they can be used to forecast the error in judgmental forecasts. This predicted error can then be used to correct the judgmental forecast. Correction methods are appropriate when the biases associated with judgmental forecasts are systematic and sustained (e.g. Lawrence et al., 2000). A number of correction methods have been proposed. Theil's method (Theil, 1971), which was suggested over 30 years ago, has been found to be effective in improving accuracy in more recent research (e.g. Ahlburg, 1984; Elgers et al., 1995; Goodwin, 1996, 2000a). This method involves regressing past outcomes onto judgmental forecasts and using the resulting model to correct future forecasts. Goodwin (1997) extended Theil's method by using discounted weighted regression to allow the correction procedure to adapt to changes in the nature of the judgmental forecaster's biases over time. Fildes (1991) found that when judgmental forecasters had access to cue information, a correction based on a regression of past forecast errors on to the values of the cues led to improved accuracy. This improvement occurred because the correction removed a tendency of the forecasters to overweight recently released information.

5.6. Judgmental adjustments of statistical forecasts

An alternative way of integrating statistical methods and judgement is to allow the forecaster to apply judgmental adjustments to statistical forecasts. Twenty-five years ago researchers were suggesting that judgmental adjustment should be discouraged because it harmed accuracy (e.g. Carbone, Andersen, Corriveau, & Corson, 1983). For example, Armstrong (1985, p. 273) argued "Business people and econo-

metricians revise sales forecasts that have been produced by objective methods. Not only is this bad practice, but it also leads to poor forecasts."

The main problem with judgmental adjustment is that forecasters read systematic patterns in the noise associated with a series (Eggleton, 1982; O'Connor et al., 1993). As a result they adjust statistical forecasts that they think have missed these patterns, with the result that forecast accuracy is reduced. However, more recent research has found that adjustments can improve the accuracy of statistical forecasts under the right conditions. The first condition is where the statistical method is deficient in its estimation of the underlying time series pattern (Willemain, 1989). Willemain (1991) has suggested that this can be detected by comparing the accuracy of the statistical method with that of a naïve forecast. He found that forecast accuracy could be improved by restricting adjustments to cases where the naïve forecast performed best. As indicated earlier, the second condition where judgmental adjustment is likely to be beneficial is where the forecaster has important domain knowledge that is not available to the statistical method, such as knowledge about a forthcoming sales promotion campaign (e.g. see Donihue, 1993; Goodwin & Fildes, 1999; Mathews & Diamantopoulos, 1990; McNees, 1990; Turner, 1990; Vere & Griffiths, 1995; Wolfe & Flores, 1990).

The adjustment process involves two stages: (i) a decision on whether a statistical forecast needs adjusting and (ii) an estimate of the size of the adjustment that is required. Most research over the last 25 years has focussed on the first stage and investigated ways in which forecasters can be discouraged from making gratuitous adjustments to statistical forecasts. Strategies such as advertising the relative inaccuracy of judgmental adjustments by providing performance feedback do not appear to work (Lim & O'Connor, 1995). Allowing the forecaster to participate in the selection of the statistical forecast has had mixed results (Lawrence, Goodwin, & Fildes, 2002). Forecasters made fewer adjustments when they were involved in the selection of the statistical forecast, but the forecasts they selected were often poor. Requiring forecasters to record the reasons for their adjustments (Goodwin, 2000b) has had some success, though even then, 35% of forecasts were still adjusted unnecessarily to the detriment of

forecast accuracy. Even when a forecaster is aware that a forecasting model has omitted a variable, Bunn and Salo (1996) show that making a judgmental adjustment to take into account the effects of this variable may be undesirable where the omitted variable is collinear with a variable included in the model. Adjustment in these circumstances may lead to a double counting bias. To mitigate this danger, Bunn and Salo propose a formal screening procedure which indicates whether adjustments are necessary. Bunn (1996) also emphasizes the potential benefits of transparent statistical methods that are conducive to more insightful judgmental interactions.

When it comes to improving the estimation of the size of adjustment that is required, researchers have tended to suggest methods based on structured decomposition (Bunn & Wright, 1991; Bunn, 1996), though few specific structures have been proposed and investigated. Wolfe and Flores (1990) and Flores et al. (1992) have applied decomposed adjustments based on the Analytic Hierarchy Process but this approach has been questioned by Belton and Goodwin (1996) and Salo and Bunn (1995).

5.7. Ensuring that forecasters have technical knowledge

Sanders and Ritzman (1992) investigated whether providing the judgmental forecaster with technical knowledge improved accuracy. Their definition of technical knowledge embraced knowledge not only of statistical forecasting methods, but also knowledge of the biases inherent in human judgement. They found that possession of technical knowledge did not improve judgmental forecasters' accuracy. Edmundson (1990) reached a similar conclusion, while Lawrence et al. (1985) found that technical knowledge improved accuracy when data was presented in a tabular, rather than graphical, format. However, all of these studies considered situations where only time series data was available.

6. Suggestions for future research

Much has been learned over the last 25 years about the important role that judgment can play in

forecasting if it is applied appropriately. Neverthe-less, many questions remain to be answered. So where should research efforts be directed over the next 25 years? We need to know more about the performance of experts in forecasting since different studies have yielded contradictory findings abut the value of expertise (Bolger & Wright, 1994; Griffin & Brenner, 2004). We also need to know more about how people acquire and use information when they make forecasts and the effects of differences in the availability of information (Camerer, Loewenstien, & Weber, 1989; Stewart, 2001; Yaniv & Hogarth, 1993; Yaniv et al., 1991). There is still scope for more research on the influence of heuristics and biases on forecast accuracy. The role of biases like illusory correlation and hindsight bias (the 'knew-it-all-along' effect) has been largely neglected (Wilkie, Tuohy, & Pollock, 1993). We also anticipate the contribution of research on the value of forecasts and the commu-nication of these forecasts.

Much also remains to be researched to develop improved methods for supporting judgmental fore-casters, particularly support in identifying when judgmental intervention is needed and when it is not needed. Promising areas of research for acompli-shing this objective include: the role of restrictive-ness and guidance in forecasting support systems (Silver, 1991), effort manipulation (Todd & Benba-sat, 1999), support for dialectical inquiry (Kasper, 1996), the different support needs of novice and experienced forecasters, support for use of analo-gies, support for the groups of forecasters and the use of new technologies in providing intelligent agents and on-line advisors. The next 25 years promises to be as exciting for researchers as the last.

References

Abramson, B., & Clemen, R. (1995). Probability forecasting. *International Journal of Forecasting*, *11*, 1–4.

Abramson, B., & Finizza, A. (1991). Using belief networks to forecast oil prices. *International Journal of Forecasting*, *7*, 299–315.

Adam, E. E., & Ebert, R. J. (1976). A comparison of human and statistical forecasting. *AIIE Transactions*, *8*(1), 120–127.

Adelman, L. (1981). The influence of formal, substantive, and contextual task properties on the relative effectiveness of different forms of feedback in multiple cue probability learning tasks. *Organizational Behavior and Human Decision Processes*, *27*, 423–442.

Ahlburg, D. A. (1984). Forecasting evaluation and improve-ment using Theil's decomposition. *Journal of Forecasting*, *3*, 345–351.

Alexander, J. (1995). Refining the degree of earnings surprise: A comparison of statistical and analysts' forecasts. *Financial Review*, *30*(3), 469–506.

Anderson, M. A., & Goldsmith, A. H. (1994). Rationality in the mind's eye: An alternative test of rational expectations using subjective forecast and evaluation data. *Journal of Economic Psychology*, *15*(3), 379–404.

Andersson, P., Edman, J., & Ekman, M. (2005). Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, *21*, 565–576.

Andreassen, P. B. (1988). Explaining the price–volume relationship: The difference between price changes and changing prices. *Organizational Behavior and Human Decision Processes*, *41*, 371–389.

Andreassen, P. B. (1991). *Causal prediction versus extrapolation: Effects of information source on judgmental forecast accuracy*, Working paper. MIT.

Andreassen, P. B., & Kraus, S. J. (1990). Judgmental extrapo-lation and the salience of change. *Journal of Forecasting*, *9*(4), 347–372.

Ang, S., & O'Connor, M. (1991). The effect of group interaction strategies on performance in time series extrapolation. *Interna-tional Journal of Forecasting*, *7*, 141–149.

Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 495–515). Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S. (1983). Relative accuracy of judgmental and extrapolative methods in forecasting annual earnings. *Journal of Forecasting*, *2*, 437–447.

Armstrong, J. S. (1985). *Long range forecasting: From crystal ball to computer* (second edition). NY: Wiley.

Armstrong, J. S. (2001). The forecasting dictionary. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 761–824). Boston: Kluwer Academic Publishers.

Armstrong, J. S., & Collopy, F. (1993). Causal forces—structuring knowledge for time-series extrapolation. *Journal of Forecasting*, *12*, 103–115.

Asquith, P., Mikhail, M., & Au, A. (2005). Information content of equity analyst reports. *Journal of Financial Economics*, *75*(2), 245–282.

Aukutsionek, S. P., & Belianin, A. V. (2001). Quality of forecasts and business performance: A survey study of Russian managers. *Journal of Economic Psychology*, *22*, 661–692.

Ayton, P., & Önkal, D. (1996). *Effects of expertise on forecasts and confidence in forecasts*. Paper presented at the 16th International Symposium on Forecasting, Istanbul, Turkey.

Babad, E., Hills, M., & O'Driscoll, M. (1992). Factors influencing wishful thinking and predictions of election outcomes. *Basic and Applied Social Psychology*, *13*, 461–476.

Baginski, S. P., Conrad, E. J., & Hassell, J. M. (1993). The effects of management forecast precision on equity pricing and on the assessment of earnings uncertainty. *The Accounting Review*, *68*, 913–927.

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, *106*, 410–433.

Batchelor, R., & Dua, P. (1990). Forecaster ideology, forecasting technique, and the accuracy of economic forecasts. *International Journal of Forecasting*, *6*(1), 3–11.

Belton, V., & Goodwin, P. (1996). Remarks on the application of the analytic hierarchy process to judgmental forecasting. *International Journal of Forecasting*, *12*, 155–161.

Beneish, M. (2001). Earnings management: A perspective. *Managerial Finance*, *27*(12), 3–17.

Benson, P. G., Curley, S. P., & Smith, G. F. (1995). Belief assessment: An underdeveloped phase of probability elicitation. *Management Science*, *41*, 1639–1653.

Benson, P. G., & Önkal, D. (1992). The effects of feedback and training on the performance of probability forecasters. *International Journal of Forecasting*, *8*, 559–573.

Björkman, M. (1972). Feedforward and feedback as determiners of knowledge and policy-notes on a neglected issue. *Scandinavian Journal of Psychology*, *13*, 152–158.

Björkman, M. (1994). Internal cue theory: Calibration and resolution of confidence in general knowledge. *Organizational Behavior and Human Decision Processes*, *58*, 386–405.

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

Blattenberger, G., & Lad, F. (1985). Separating the Brier score into calibration and refinement components: A graphical exposition. *The American Statistician*, *39*, 26–32.

Bolger, F., & Harvey, N. (1993). Context sensitive heuristics in statistical reasoning. *Quarterly Journal of Experimental Psychology*, *46A*(4), 779–811.

Bolger, F., & Harvey, N. (1995). Judging the probability that the next point in an observed time-series will be below, or above, a given value. *Journal of Forecasting*, *14*, 597–607.

Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting*, *20*, 29–39.

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision Support Systems*, *11*, 1–24.

Braun, P. A., & Yaniv, I. (1992). A case study of expert judgment: Economists' probabilities versus base-rate model forecasts. *Journal of Behavioral Decision Making*, *5*, 217–231.

Brenner, L., Griffin, D., & Koehler, D. J. (2005). Collaborative planning and predication: Does group discussion affect optimistic case-based judgement. *Organizational Behavior and Human Decision Processes*, *97*(1), 64–81.

Brown, L. (1996). Analyst forecasting errors and their implications for security analysts: An alternative perspective. *Financial Analysts Journal*, *52*(1), 40–47.

Bunn, D. (1987). Expert use of forecasts: Bootstrapping and linear models. In G. Wright, & P. Ayton (Eds.), *Judgmental forecasting* (pp. 229–241). Chichester: Wiley.

Bunn, D., & Salo, A. A. (1996). Adjustment of forecasts with model consistent expectations. *International Journal of Forecasting*, *12*, 163–170.

Bunn, D., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods. Issues and analysis. *Management Science*, *37*, 501–518.

Bunn, D. W. (1996). Non-traditional methods of forecasting. *European Journal of Operational Research*, *92*, 528–536.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*, 1232–1255.

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*, *20*, 559–566.

Carbone, R., & Gorr, W. (1985). Accuracy of judgmental forecasting of time series. *Decision Sciences*, *16*, 153–160.

Carlson, B. (1993). The accuracy of future forecasts and past judgments. *Organizational Behavior and Human Decision Processes*, *54*, 410–420.

Chatfield, C. (2001). Prediction intervals for time-series forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 475–494). Norwell, MA: Kluwer Academic Publishers.

Chatfield, R., Hein, S., & Moyer, S. (1990). Long term earnings forecasts in the electric utility industry: Accuracy and valuation implications. *Financial Review*, *25*(3), 421–439.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*, 841–862.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583.

Clemen, R. T., Murphy, A. H., & Winkler, R. L. (1995). Screening probability forecasts; contrasts between choosing and combining. *International Journal of Forecasting*, *11*, 133–146.

Clements, M., & Tse, S. (2005). Financial analyst characteristics and herding behavior in forecasting. *Journal of Finance*, *60*(1), 307–341.

Clements, M. P. (1995). Rationality and the role of judgement in macroeconomic forecasting. *The Economic Journal*, *105*(429), 410.

Clements, M. P., & Taylor, N. (2003). Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics*, *18*, 445–456.

Connolly, T., & Dean, D. (1997). Decomposed versus holistic estimates of effort required for software writing tasks. *Management Science*, *43*, 1029–1045.

Corker, R. J., Holly, S., & Ellis, R. G. (1986). Uncertainty and forecast precision. *International Journal of Forecasting*, *2*, 53–70.

Curley, S. P., & Benson, P. G. (1994). Applying a cognitive perspective to probability construction. In P. Ayton, & G. Wright (Eds.), *Subjective probability* (pp. 185–209). Chichester: John Wiley & Sons.

Curley, S. P., Browne, G. J., Smith, G. F., & Benson, P. G. (1995). Arguments in the practical reasoning underlying constructive probability responses. *Journal of Behavioral Decision Making*, *8*, 1–20.

Daan, H., & Murphy, A. H. (1982). Subjective probability forecasting in the Netherlands: Some operational and experimental results. *Meteorologische Rundschau*, *35*, 99–112.

Dalrymple, D. J. (1987). Sales forecasting practices: Results from a United States survey. *International Journal of Forecasting*, *3*, 379–391.

Davis, F. D., Lohse, G. L., & Kottemann, J. E. (1994). Harmful effects of seemingly helpful information on forecasts of stock earnings. *Journal of Economic Psychology*, *15*, 253–267.

Dawes, R. (1975). Graduate admission variables and future success. *Science*, *187*, 721–743.

De Bondt, W. F. M. (1993). Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting*, *9*, 355–371.

de Menezes, L. M., Bunn, D. W., & Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, *120*, 190–204.

Desanctis, G., & Jarvenpaa, S. L. (1989). Graphical presentation of accounting data for financial forecasting: An experimental investigation. *Accounting, Organizations and Society*, *14*(5,6), 5–9.

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

Edmundson, R. (1990). Decomposition: A strategy for judgmental forecasting. *Journal of Forecasting*, *9*, 305–314.

Edmundson, R., Lawrence, M., & O'Connor, M. (1988). The use of non time series information in sales forecasting: A case study. *Journal of Forecasting*, *7*, 201–211.

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

Elgers, P. T., May, H. L., & Murray, D. (1995). Note on adjustments to analysts' earning forecasts based upon systematic cross-sectional components of prior-period errors. *Management Science*, *41*, 1392–1396.

Evans, J. St. B. T. (1987). Beliefs and expectations as causes of judgmental bias. In G. Wright, & P. Ayton (Eds.), *Judgmental forecasting* (pp. 31–47). Chichester: Wiley.

Ferrell, W. R., & McGoey, P. J. (1980). A model of calibration for subjective probabilities. *Organizational Behavior and Human Performance*, *26*, 32–53.

Fildes, R. (1991). Efficient use of information in the formation of subjective industry forecasts. *Journal of Forecasting*, *10*, 597–617.

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

Fildes, R., & Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics*, *24*(4), 435–468.

Fischer, I., & Harvey, N. (1999). Combining forecasts: What information do judges need to outperform the simple average? *International Journal of Forecasting*, *15*, 227–246.

Fischhoff, B. (1988). Judgmental aspects of forecasting: Needs and possible trends. *International Journal of Forecasting*, *4*, 331–339.

Fischhoff, B., & MacGregor, D. (1982). Subjective confidence in forecasts. *Journal of Forecasting*, *1*, 155–172.

Flores, B. E., Olson, D. L., & Wolfe, C. (1992). Judgmental adjustment of forecasts—a comparison of methods. *International Journal of Forecasting*, *7*, 421–433.

Flugstad, A. R., & Windschitl, P. D. (2003). The influence of reasons on interpretations of probability forecasts. *Journal of Behavioral Decision Making*, *16*, 107–126.

Foong, S. -Y., Lawrence, M., & O'Connor, M. (2003). The interaction of incentives and information disclosure: The case of confidence interval information. *The British Accounting Review*, *35*, 233–255.

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, *21*, 551–564.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.

Goodwin, P. (1996). Statistical correction of judgmental point forecasts and decisions. *Omega*, *24*, 551–559.

Goodwin, P. (1997). Adjusting judgmental extrapolations using Theil's method and discounted weighted regression. *Journal of Forecasting*, *16*, 37–46.

Goodwin, P. (2000a). Correct or combine? Mechanically integrating judgmental forecasts with statistical methods. *International Journal of Forecasting*, *16*, 261–275.

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

Goodwin, P. (2005). Providing support for decisions based on time series information under conditions of asymmetric loss. *European Journal of Operational Research*, *163*(2), 388–402.

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., Önkal-Atay, D., Thomson, M. E., Pollock, A. C., & Macaulay, A. (2004). Feedback-labelling synergies in judgmental stock price forecasting. *Decision Support Systems*, *37*, 175–186.

Goodwin, P., & Wright, G. (1993). Improving judgmental time series forecasting: A review of the guidance provided by research. *International Journal of Forecasting*, *9*, 147–161.

Goodwin, P., & Wright, G. (1994). Heuristics, biases and improvement strategies in judgmental time series forecasting. *Omega*, *22*, 553–568.

Griffin, D., & Brenner, L. (2004). Perspectives on probability judgment calibration. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 177–199). Oxford: Blackwell.

Hamill, T. M., & Wilks, D. S. (1995). A probabilistic forecast contest and the difficulty in assessing short-range forecast uncertainty. *Weather and Forecasting*, *10*, 620–631.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. (1975). Social judgement theory. In M. F. Kaplan, & S. Schwartz (Eds.), *Human judgement and decision processes* (pp. 277–312). New York: Academic Press.

Harvey, N. (1988). Judgmental forecasting of univariate time series. *Journal of Behavioral Decision Making*, *1*, 95–110.

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

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

Harvey, N., Bolger, F., & McClelland, A. (1991). *Judgmental forecasting within and across correlated time series*. Department of Psychology, University College London Working paper.

Harvey, N., Bolger, F., & McClelland, A. (1994). On the nature of expectations. *British Journal of Psychology*, *85*, 203–229.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.

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

Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational Behavior and Human Decision Processes*, *81*, 252–273.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *11*, 719–731.

Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science*, *27*, 115–138.

Holthausen, R., & Leftwich, R. (1983). The economics consequences of accounting choice. *Journal of Accounting and Economics*, *5*(2), 77–117.

Hsu, W. R., & Murphy, A. H. (1986). The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *International Journal of Forecasting*, *2*, 285–293.

Ivkovic, Z., & Jegadeesh, N. (2005). The timing and value of forecast and recommendation revisions. *Journal of Financial Economics*, *73*(3), 433–463.

Johnson, W. B. (1982). The impact of confidence interval information on probability judgements. *Accounting, Organizations and Society*, *7*, 349–367.

Johnson, W. B. (1983). Representativeness in judgmental predictions of corporate bankruptcy. *The Accounting Review*, *LVIII*, 78–97.

Jørgensen, M., & Sjøberg, D. I. K. (2003). An effort prediction interval approach based on the empirical distribution of previous estimation accuracy. *Information and Software Technology*, *45*, 123–136.

Jørgensen, M., Teigen, K. H., & Moløkken, K. (2004). Better sure than safe? Over-confidence in judgement based software development effort prediction intervals. *The Journal of Systems and Software*, *70*, 79–93.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, *57*, 226–246.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*, 17–31.

Kahneman, D., & Riepe, M. W. (1998). Aspects of investor psychology. *Journal of Portfolio Management*, *24*, 52–65.

Kasper, G. M. (1996). A theory of decision support system design for user calibration. *Information Systems Research*, *7*, 215–232.

Keller, P. A., & Block, L. G. (1997). Vividness effects: A resource-matching perspective. *Journal of Consumer Research*, *24*, 295–304.

Keren, G. (1997). On the calibration of probability judgments: Some critical comments and alternative perspectives. *Journal of Behavioral Decision Making*, *10*, 269–278.

Keren, G. B. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes*, *39*, 98–114.

Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer, & C. J. B. Joyce (Eds.), *Human judgement: The SJT view* (pp. 115–156). Elsevier Science Publisher.

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*, 216–247.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, *249*, 2–9.

Langer, E. J. (1982). *The psychology of control*. Beverly Hills: Sage.

Lawrence, M. (1983). An exploration of some practical issues in the use of quantitative forecasting models. *Journal of Forecasting*, *1*, 169–179.

Lawrence, M., Edmundson, R., & O'Connor, M. (1985). An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting*, *1*, 25–35.

Lawrence, M., Edmundson, R., & O'Connor, M. (1986). The accuracy of combining judgmental and statistical forecasts. *Management Science*, *32*, 1521–1532.

Lawrence, M., Goodwin, P., & Fildes, R. (2002). Influence of user participation on DSS use and decision accuracy. *Omega, International Journal of Management Science*, *30*, 381–392.

Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, *43*, 172–187.

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

Lawrence, M., & O'Connor, M. (1993). Scale, randomness and the calibration of judgmental confidence intervals. *Organizational Behavior and Human Decision Processes*, *56*, 441–458.

Lawrence, M., & O'Connor, M. (1995). The anchoring and adjustment heuristic in time series forecasting. *Journal of Forecasting*, *14*, 443–451.

Lawrence, M., & O'Connor, M. (1996). Judgement or models: The importance of task differences. *Omega*, *24*, 245–254.

Lawrence, M., & O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions. *International Journal of Forecasting*, *21*, 3–14.

Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research*, *122*, 151–160.

Lichtenstein, S., Fischhoff, B., & Phillips, P. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge: Cambridge University Press.

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

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

Lim, K. H., O'Connor, M. J., & Remus, W. E. (2005). The impact of presentation media on decision making: Does multimedia improve the effectiveness of feedback? *Information & Management*, *42*, 305–316.

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and non-random events. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *13*, 392–400.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 111–153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., et al. (1993). The M2 competition: A real time judgmentally based forecasting study. *International Journal of Forecasting*, *9*, 5–22.

Martino, J. P. (2003). A review of selected recent advances in technological forecasting. *Technological Forecasting & Social Change*, *70*, 719–733.

Mathews, B., & Diamantopolous, A. (1986). Managerial intervention in forecasting: An empirical investigation of forecast manipulation. *International Journal of Research in Marketing*, *3*, 3–10.

Mathews, B., & Diamantopolous, A. (1989). Judgmental revision of statistical forecasts: A longitudinal extension. *Journal of Forecasting*, *8*, 129–140.

Mathews, B., & Diamantopolous, A. (1990). Judgmental revision of sales forecasts: Effectiveness of forecast selection. *Journal of Forecasting*, *9*, 407–415.

McClelland, A. G. R., & Bolger, F. (1994). The calibration of subjective probabilities: Theories and models 1980–94. In G. Wright, & P. Ayton (Eds.), *Subjective probability* (pp. 453–482). Chichester: Wiley.

McNees, S. (1975, Nov–Dec). An evaluation of economic forecasts. *New England Economic Review*, *16*, 3–39.

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

Mlyne, K. R. (2002). Decision-making from probability forecasts based on forecast value. *Meteorological Applications*, *9*, 307–315.

Moore, D. A., Kurtzberg, T. R., Fox, C. R., & Bazerman, M. H. (1999). Positive illusions and forecasting errors in mutual fund investment decisions. *Organizational Behavior and Human Decision Processes*, *79*(2), 95–115.

Moser, D. V. (1989). The effects of output interference, availability, and accounting information on investors' predictive judgments. *The Accounting Review*, *64*, 433–448.

Mosteller, F., Siegel, A., Trapido, E., & Youtz, C. (1981). Eye fitting straight lines. *American Statistician*, *35*(3), 150.

Muradoglu, G., & Önkal, D. (1994). An exploratory analysis of the portfolio managers' probabilistic forecasts of stock prices. *Journal of Forecasting*, *13*, 565–578.

Murphy, A. H. (1972a). Scalar and vector partitions of the probability score: Part I. Two-state situation. *Journal of Applied Meteorology*, *11*, 273–282.

Murphy, A. H. (1972b). Scalar and vector partitions of the probability score: Part II. N-state situation. *Journal of Applied Meteorology*, *11*, 1183–1192.

Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology*, *12*, 595–600.

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, *116*, 2417–2424.

Murphy, A. H. (1998). The early history of probability forecasts: Some extensions and clarifications. *Weather and Forecasting*, *13*, 5–15.

Murphy, A. H., & Winkler, R. L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association*, *79*, 489–500.

Murphy, A. H., & Winkler, R. L. (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting*, *7*, 435–455.

Nikolopolous, K., Fildes, R., Goodwin, P., & Lawrence, M. (2005). *On the accuracy of judgmental interventions on forecasting support systems*. Lancaster University Management School Working Paper 2005/022.

O'Connor, M. (1989). Models of human behaviour and confidence in judgment: A review. *International Journal of Forecasting*, *5*, 159–169.

O'Connor, M., & Lawrence, M. (1989). An examination of the accuracy of judgmental confidence intervals in time series forecasting. *Journal of Forecasting*, *8*, 141–155.

O'Connor, M., & Lawrence, M. (1992). Time series characteristics and the widths of judgmental confidence intervals. *International Journal of Forecasting*, *7*, 413–420.

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

O'Connor, M., Remus, W., & Griggs, K. (1997). Going up-going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting*, *16*(3), 165–176.

O'Connor, M., Remus, W., & Griggs, K. (2001). The asymmetry of judgmental confidence intervals in time series forecasting. *International Journal of Forecasting*, *17*, 623–633.

Önkal, D., & Bolger, F. (2004). Provider-user differences in perceived usefulness of forecasting formats. *Omega*, *32*, 31–39.

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

Önkal, D., & Muradoglu, G. (1995). Effects of feedback on probabilistic forecasts of stock prices. *International Journal of Forecasting*, *11*, 307–319.

Önkal, D., & Muradoglu, G. (1996). Effects of task format on probabilistic forecasting of stock prices. *International Journal of Forecasting*, *12*, 9–24.

Önkal, D., Yates, J., Simga-Mugan, F., & Oztin, C. (2003). Professional vs. amateur judgment accuracy: The case of foreign exchange rates. *Organizational Behavior and Human Decision Processes*, *91*, 169–185.

Önkal-Atay, D. (1998). Financial forecasting with judgment. In G. Wright, & P. Goodwin (Eds.), *Forecasting with judgment* (pp. 139–167). Chichester: John Wiley & Sons.

Önkal-Atay, D., Thomson, M. E., & Pollock, A. C. (2002). Judgmental forecasting. In M. P. Clements, & D. Hendry (Eds.), *A companion to economic forecasting* (pp. 133–151). Oxford: Blackwell Publishers.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*, 85–92.

Phillips, L. D. (1987). On the adequacy of judgmental probability forecasts. In G. Wright, & P. Ayton (Eds.), *Judgmental forecasting* (pp. 11–30). Chichester: Wiley.

Poland, B., & Wada, R. (2001). Combining drug-disease and economic modelling to inform drug development decisions. *Drug Discovery Today*, *6*, 1165–1170.

Pollock, A. C., Macaulay, A., Önkal-Atay, D., & Thomson, M. E. (2002). Consistent judgmental directional probability exchange rate predictions. In K. D. Lawrence, M. D. Geurts, & J. G. Guerard Jr. (Eds.), *Advances in business and management forecasting*, *vol. 3* (pp. 161–175). Oxford: Elsevier Science.

Pollock, A. C., Macaulay, A., Thomson, M. E., & Önkal, D. (2005). Performance evaluation of judgmental directional exchange rate predictions. *International Journal of Forecasting*, *21*, 473–489.

Pollock, A. C., & Wilkie, M. E. (1992). Currency forecasting: Human judgment or models. *VBA Journaal*, *3*, 21–29.

Pollock, A. C., & Wilkie, M. E. (1993). Directional judgmental financial forecasting: Trends and random walks. In R. Flavell (Ed.), *Modeling reality and personal modelling* (pp. 253–271). Heidelberg: Physica-Verlag.

Pownall, G., Wasley, C., & Waymire, G. (1993). Alternative forms of management earnings forecasts: Incidence and stock price effects. *The Accounting Review*, *68*, 896–912.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*, 39–57.

Remus, W., O'Connor, M., & Griggs, K. (1995). Does reliable information improve the accuracy of judgmental forecasts? *International Journal of Forecasting*, *11*(2), 285–293.

Remus, W., O'Connor, M., & Griggs, K. (1996). Does feedback improve the accuracy of recurrent judgmental forecasts? *Organizational Behavior and Human Decision Processes*, *66*, 22–30.

Remus, W., O'Connor, M., & Griggs, K. (1998). The impact of incentives on the accuracy of subjects in judgmental forecasting experiments. *International Journal of Forecasting*, *14*, 515–522.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes*, *40*, 193–218.

Russo, J. E., & Schoemaker, P. J. H. (1992). Managing overconfidence. *Sloan Management Review*, *33*, 7–17.

Saaty, T. L., & Vargas, L. G. (1991). *Prediction projection and forecasting*. Norwell, MA: Kluwer Academic Publishers.

Salo, A. A., & Bunn, D. W. (1995). Decomposition in the assessment of judgmental probability forecasts. *Technological Forecasting & Social Change*, *49*, 13–25.

Sanders, F. (1963). On subjective probability forecasting. *Journal of Applied Meteorology*, *2*, 191–201.

Sanders, N., & Ritzman, L. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, *5*, 39–52.

Sanders, N. R. (1992). Accuracy of judgmental forecasts: A comparison. *Omega*, *20*(3), 353.

Sanders, N. R. (1997). The impact of task properties feedback on time series judgmental forecasting tasks. *Omega*, *25*, 135–144.

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

Sanders, N. R., & Ritzman, L. P. (1995). Bringing judgement into combination forecasts. *Journal of Operations Management*, *13*, 311–321.

Schoemaker, P. J. H. (2004). Forecasting and scenario planning: The challenges of uncertainty and complexity. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 247–296). Oxford: Blackwell.

Silver, M. S. (1991). Decisional guidance for computer-based support. *MIS Quarterly*, *15*, 105–133.

Smith, G. G., Benson, P. G., & Curley, S. P. (1991). Belief, knowledge, and uncertainty: A cognitive perspective on subjective probability. *Organizational Behavior and Human Decision Processes*, *48*, 291–321.

Sniezek, J. (1990). A comparison of techniques for judgmental forecasting by groups with common information. *Group & Organization Studies*, *15*(1), 5–20.

Sniezek, J. A. (1986). The role of variable labels in cue probability learning tasks. *Organizational Behavior and Human Decision Processes*, *38*, 141–161.

Sniezek, J. A. (1989). An examination of group process in judgmental forecasting. *International Journal of Forecasting*, *5*(2), 171–178.

Stewart, T. R. (2001). Improving reliability of judgmental forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 81–106). Norwell, MA: Kluwer Academic Publishers.

Stewart, T. R., Roebber, P. J., & Bosart, L. F. (1999). The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision Processes*, *69*, 205–219.

Stone, E. R., & Opel, R. B. (2000). Training to improve calibration and discrimination: The effects of performance and environmental feedback. *Organizational Behavior and Human Decision Processes*, *83*, 282–309.

Surowiecki, J. (2004). *The wisdom of crowds*. London: Abacus.

Tay, A. S., & Wallis, K. F. (2002). Density forecasting: A survey. In M. P. Clements, & D. F. Hendry (Eds.), *A companion to economic forecasting* (pp. 45–68). Oxford: Blackwell Publishers.

Taylor, J. W. (1999). Evaluating volatility and interval forecasts. *Journal of Forecasting*, *18*, 111–128.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, *80*, 155–190.

Teigen, K. H., & Brun, W. (2000). Ambiguous probabilities: When does $p = 0.3$ reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision Making*, *13*, 345–362.

Theil, H. (1971). *Applied economic forecasting*. Amsterdam: North-Holland Publishing Company.

Thomson, M. E., Önkal-Atay, D., Pollock, A. C., & Macaulay, A. (2003). The influence of trend strength on directional probabilistic currency predictions. *International Journal of Forecasting*, *19*, 241–256.

Thomson, M. E., Pollock, A. C., Henriksen, K. B., & Macaulay, A. (2004). The influence of the forecast horizon on judgmental probability forecasts of exchange rate movements. *The European Journal of Finance*, *10*(4), 290–307.

Timmers, H., & Wagenaar, W. A. (1977). Inverse statistics and misperception of exponential-growth. *Perception & Psychophysics*, *21*, 558–562.

Todd, P., & Benbasat, I. (1999). Evaluating the impact of DSS, cognitive effort, and incentives on strategy selection. *Information Systems Research*, *10*, 356–374.

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

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.

Vertinsky, P., Kanetkar, V., Vertinsky, I., & Wilson, G. (1986). Predictions of "wins and losses" in a series of field hockey games: A study of probability assessment quality and cognitive information-processing models of players. *Organizational Behavior and Human Decision Processes*, *38*, 392–404.

Wagenaar, W. A., & Keren, G. B. (1985). Calibration of probability assessments by professional blackjack dealers, statistical experts and lay people. *Organizational Behavior and Human Decision Processes*, *36*, 406–416.

Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential-growth. *Perception & Psychophysics*, *18*, 16–422.

Watts, R., & Zimmerman, J. (1990). Positive accounting theory: A ten year perspective. *The Accounting Review*, *65*(1), 131–156.

Webby, R., & O'Connor, M. (1996). Judgmental and statistical time series forecasting: A review of the literature. *International Journal of Forecasting*, *12*, 91–118.

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

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*, 806–820.

Welch, E., Bretschneider, S., & Rohrbaugh, J. (1998). Accuracy of judgmental extrapolation of time series data. Characteristics, causes, and remediation strategies for forecasting. *International Journal of Forecasting*, *14*, 95–110.

Whitecotton, S. M. (1996). The effects of experience and a decision aid on the slope, scatter, and bias of earnings forecasts. *Organizational Behavior and Human Decision Processes*, *66*, 111–121.

Wilkie, M. E., & Pollock, A. C. (1994). Directional currency forecasting: An investigation into probability judgement accuracy. In L. Peccati, & M. Viren (Eds.), *Financial modeling* (pp. 354–364). Heidelberg: Physica-Verlag.

Wilkie, M. E., & Pollock, A. C. (1996). An application of probability judgement accuracy measures to currency forecasting. *International Journal of Forecasting*, *12*, 25–40.

Wilkie, M. E., Tuohy, A. P., & Pollock, A. C. (1993). Examining heuristics and biases in judgmental currency forecasting. *VBA Journal*, *2*, 12–17.

Wilkie-Thomson, M. E., Önkal-Atay, D., & Pollock, A. C. (1997). Currency forecasting: An investigation of extrapolative judgment. *International Journal of Forecasting*, *13*, 509–526.

Willemain, T. R. (1989). Graphical adjustment of statistical forecasts. *International Journal of Forecasting*, *5*, 179–185.

Willemain, T. R. (1991). The effect of graphical adjustment on forecast accuracy. *International Journal of Forecasting*, *7*, 151–154.

Windschitl, P. D., Martin, R., & Flugstad, A. R. (2002). Context and the interpretation of likelihood information: The role of intergroup comparisons on perceived vulnerability. *Journal of Personality and Social Psychology*, *82*, 742–755.

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

Worthen, B. (2003). Future results not guaranteed; contrary to what vendors tell you, computer systems alone are incapable of producing accurate forecasts. *CIO*, *16*(19), 1.

Wright, G. (1982). Changes in the realism and distribution of probability assessments as a function of question type. *Acta Psychologica*, *52*, 165–174.

Wright, G., & Ayton, P. (1986). Subjective confidence in forecasts: A response to Fischhoff and MacGregor. *Journal of Forecasting*, *5*, 117–123.

Wright, G., & Ayton, P. (1987). The psychology of forecasting. In G. Wright, & P. Ayton (Eds.), *Judgmental forecasting* (pp. 83–105). Chichester: Wiley.

Wright, G., & Ayton, P. (1988). Immediate and short-term judgmental forecasting: Personologism, situationism or interactionism? *Personality and Individual Differences*, *9*, 109–120.

Wright, G., & Ayton, P. (1989). Judgmental probability forecasts for personal and impersonal events. *International Journal of Forecasting*, *5*, 117–125.

Wright, G., & Ayton, P. (1992). Judgmental probability forecasting in the immediate and medium term. *Organizational Behavior and Human Decision Processes*, *51*, 344–363.

Wright, G., Rowe, G., Bolger, F., & Gammack, J. (1994). Coherence, calibration, and expertise in judgmental probability forecasting. *Organizational Behavior and Human Decision Processes*, *57*, 1–25.

Wright, G., & Wisudha, A. (1982). Distribution of probability assessments for almanac and future event questions. *Scandinavian Journal of Psychology*, *23*, 219–224.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and 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., & Hogarth, R. M. (1993). Judgmental versus statistical prediction: Information asymmetry and combination rules. *Psychological Science*, *4*, 58–62.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260–281.

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, *110*, 611–617.

Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, *30*, 132–156.

Yates, J. F. (1988). Analyzing the accuracy of probability judgments for multiple events: An extension of the covariance decomposition. *Organizational Behavior and Human Decision Processes*, *41*, 281–299.

Yates, J. F., & Curley, S. P. (1985). Conditional distribution analyses of probabilistic forecasts. *Journal of Forecasting*, *4*, 61–73.

Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational Behavior and Human Decision Processes*, *49*, 60–79.

Yates, J. F., Price, P. C., Lee, J. -W., & Ramirez, J. (1996). Good probabilistic forecasters: The 'consumer's' perspective. *International Journal of Forecasting*, *12*, 41–56.

Zakay, D. (1983). The relationship between the probability assessor and the outcomes of an event as a determiner of subjective probability. *Acta Psychologica*, *53*, 271–280.