Currency forecasting: an investigation of extrapolative judgement

Mary E. Wilkie-Thomson[a,*], Dilek Önkal-Atay[b], Andrew C. Pollock[c]

[a]Department of Consumer Studies, Glasgow Caledonian University, Park Campus, 1 Park Drive, Glasgow, G3 6LP, Scotland, UK
[b]Faculty of Business Administration, Bilkent University, 06533 Bilkent, Ankara, Turkey
[c]Department of Mathematics, Glasgow Caledonian University, City Campus, Cowcaddens Road, Glasgow, G4 0BA, Scotland, UK

Abstract

This paper aims to explore the potential effects of trend type, noise and forecast horizon on experts' and novices' probabilistic forecasts. The subjects made forecasts over six time horizons from simulated monthly currency series based on a random walk, with zero, constant and stochastic drift, at two noise levels. The difference between the Mean Absolute Probability Score of each participant and an AR(1) model was used to evaluate performance. The results showed that the experts performed better than the novices, although worse than the model except in the case of zero drift series. No clear expertise effects occurred over horizons, albeit subjects' performance relative to the model improved as the horizon increased. Possible explanations are offered and some suggestions for future research are outlined. © 1997 Elsevier Science B.V.

Keywords: Evaluation; Exchange rate; Expertise; Forecasting; Judgement

1. Introduction

Forecasts based on human judgement are widely used in practical situations (e.g. Dalrymple, 1987; Klein and Linneman, 1984). One such situation is currency forecasting, where predictions are often based on judgement alone or, at the very least, in combination with statistical models. This is especially the case with the 'chartist' forecasting approach, which essentially consists of two principal judgemental tasks (Murphy, 1986). The first of these tasks is to identify trends at the beginning of their development for the purpose of trading in the appropriate direction. The second task involves recognising when the price series is indicative of a trend reversal and distinguishing this situation from instances when apparent contradictory movements may only reflect noise. Despite the practical significance of judgement in this area, academic research has tended to be quantitatively based, focusing on the advantages of one statistical forecasting method relative to another. Consequently, very little is known about the quality of professional currency forecasting judgement and how it is affected by relevant characteristics such as the type of trend, the level of noise and the length of the forecast horizon. This paper reports an exploratory investigation of these issues within a probability forecasting framework.

*Corresponding author. Tel.: +44 141 3374035; fax: +44 141 3374420.

Our focus on probabilistic forecasts stems from their advantages over point forecasts in presenting quantitative descriptions of forecaster's uncertainty, hence, enabling their users to make more informed decisions (Murphy and Winkler, 1974, 1992). Comparative advantages of using judgemental probability forecasts have been emphasised in a variety of decision-making contexts (Wright et al., 1996), including financial domains. In a study carried out by Kabus (1976), seven top banking executives predicted the value of interest rates 3 months into the future and attached probability assessments to their predictions. These experts performed very well at predicting actual values, and the correct direction of movement was predicted in all cases. In contrast, much of the earlier work examining probabilistic forecasting of stock prices has reported poor results. For instance, Stael von Holstein (1972) compared stock price predictions of five subject groups – stock market experts, bankers, university business teachers, business students and statisticians. His subjects' predictions were astonishingly poor: only 3 out of 72 subjects performed better than a 'uniform forecaster' (i.e. a forecaster who assigns equal probabilities to all possible occurrences). Furthermore, the relationship between level of expertise and accuracy was almost the opposite of what one would expect. The statisticians performed best, followed by the stock market experts, students, teachers, and finally bankers. This 'inverted expertise' effect has also been illustrated in two recent stock market studies. Yates et al. (1991), in a study concerning both prices and earnings, found that the probabilistic forecasts of 'novices' (i.e. undergraduate business administration students) were more accurate than that of 'semi-experts' (i.e. graduate business students). Önkal and Muradoglu (1994) analysed stock price forecasts, and found that students who had previously made stock investment decisions (i.e. semi-experts) performed worse than students with no active trading experience. However, both studies used students as 'semi-experts' in concluding the effects of expertise. Also, Stael von Holstein (1972); Yates et al. (1991), and Önkal and Muradoglu (1994) have all employed multiple-interval task structures (where the forecaster is asked to report his/her predictions by assigning probabilities to a given number of intervals) as

opposed to dichotomous task structures (where the forecaster predicts which of the two possible outcomes will occur and then assigns a probability for the chosen outcome's occurrence). It is shown that the choice of task structure can have important implications for reporting and evaluating probability judgements (Ronis and Yates, 1987). Thus, the exclusive use of multiple-interval task format may be viewed as another important factor that should be considered in interpreting previous findings.

Focusing on the potential limitations of past research, Muradoglu and Önkal (1994) and Önkal and Muradoglu (1996) have investigated probabilistic forecasting performance of professional portfolio managers (i.e. experts) and other banking professionals participating in a portfolio management workshop (i.e. semi-experts). Results suggested that forecasting horizon and task format were significant determinants of forecasting performance. As governed by these two factors, the ecological validity of the forecasting task (i.e. its agreement with experts' natural environments) was found to be of critical importance in explaining experts' performance. This conclusion supports Bolger and Wright (1994) contention that ecological validity and learnability of tasks provide the critical variables for understanding the contradictory findings of expertise research. Accordingly, the alleged inverse-expertise effect of earlier studies was not found when performances of professional portfolio managers and other banking professionals were analysed (Önkal and Muradoglu, 1996). This research accentuated the need for further investigation to delineate the different dimensions of forecasting accuracy that can be expected at various levels of expertise. One objective of the present study was to examine this issue within a currency forecasting context, particularly in relation to important price series characteristics such as the types of trend and levels of noise. In order to proceed within this framework, we next review the literature specifically concerned with time series forecasting.

Many recent studies have focused on 'abstract' time series forecasting tasks, i.e. forecasting under conditions where no information on the nature of the series is provided to subjects (Goodwin and Wright, 1991; Webby and O'Connor, 1996). Although the abstract design is highly representative of the chartist

forecasting approach outlined initially.[1] this is not the case in other decision making domains where contextual information is utilised in addition to time series information in the forecasting process. However, even in the latter cases, the design is still valid. As O'Connor and Lawrence (1989) have pointed out, the quality of time series extrapolative judgement cannot be directly examined unless other data (i.e. environmental cues) are eliminated. If environmental cues are not controlled, the subject is able to retrieve relevant information from memory and this is likely to result in judgement based on both time series and non-time series information. As such, little can be said about the possible causes of either good or bad performance: it is impossible to determine whether poor judgement, for instance, is the result of salient non-time series information (Tversky and Kahneman, 1973) or factors specific to the series (e.g. Bolger and Harvey, 1993).

Abstract forecasting tasks have so far enabled various important issues to be addressed. Of particular relevance to the present investigation are studies which have examined subjects' ability to extrapolate from trended and random series. A pervasive finding that has emerged from previous research is the tendency to under-estimate the strength of the trend (Andreassen, 1988; Eggleton, 1982; Lawrence and Makridakis, 1989). This underestimation bias has been found to be particularly strong when subjects extrapolate from deterministic exponential functions (Wagenaar and Sagaria, 1975; Wagenaar and Timmers, 1978, 1979).

The ability to recognise randomness or to detect a trend from noisy data are further issues that are of paramount importance to a currency analyst. Strong negative statements have been made in the psychological literature about the human concept of randomness. However, this view is arguably unjustified. For example, in a critique of this literature, Ayton et

al. (1989) have shown that many of the randomness tasks presented to subjects are logically and methodologically problematic. Wagenaar (1972) claims that studies have shown people to be poor at recognising randomness, but fails to cite any examples. In fact, very few studies have focused on recognition, and those that did exhibited good performance (e.g. Baddeley, 1966; Cook, 1967). Further support that there is a performance difference between recognition and production tasks comes from a time series study carried out by Harvey (1988). In this study, individuals were able to acquire internal representations of the process used to generate data points, but did not use these representations in a forecasting task.

Other studies have shown that people are able to detect a known trend from noisy data. For example, Mosteller et al. (1981) and Lawrence and Makridakis (1989) found that the level of noise did not affect the ability to identify a trend. However, this was not the case in a study by Andreassen and Kraus (1990) which found that subjects tended to identify a trend more often when the signal was strong relative to the noise level.

Studies of extrapolating, rather than detecting, trends from noisy data have also produced contradictory findings. Much of this research has compared human judgement to statistical models. Some studies have found human judgement to be less accurate than quantitative methods. For instance, Adam and Ebert (1976) conducted a comparison study to assess the impact of pattern complexity (comprising trend, trend with low and high seasonally) and the degree of noise and found these factors to have a significant detrimental effect on performance. However, it has been asserted that when the underlying signal of a series is unstable, human judgement can outperform, or at least rival, statistical models. For instance, Lawrence (1983) compared judgement with statistical forecasts obtained via exponential smoothing and Box-Jenkins techniques on a series of US airline passenger data and found little difference in accuracy. Similarly, Sanders and Ritzman (1992) found good judgemental performance relative to statistical models with higher variability series. However, it appears that people perform poorly relative to statistical models when extrapolating more complex

---

[1]Chartists do not use contextual information due to the belief that all indicators of change (i.e. economic, political, psychological or otherwise) are reflected in the pattern of the price series itself and, therefore, a study of price action is all that is needed to forecast future price movements (Murphy, 1986). The chartist is aware that there are causes for rises and falls in currency rates. However, he or she simply doesn't think that the forecasting task requires a knowledge of these causes.

stable signals from noisy data. For instance, for a high noise step function, Sanders (1992) found human judgement to perform much worse than a statistical method. In a similar vein, Remus et al. (1995) documented the forecasters' overreaction to immediate past information, implicating the problems that may be confronted in assessing randomness.

A number of studies have focused on the effect of length of the forecast horizon on judgemental accuracy. There is evidence relating to both novices and experts that an inverse relationship exists between accuracy and the length of the forecast horizon. Lawrence and O'Connor (1992), with non-experienced subjects, and Basi et al. (1976), with professional security analysts, found accuracy to be greater in the shorter horizons. A reason for this may be found in the Bolger and Harvey (1993) study. They suggested that subjects tended to make repetitions of previous forecasts as the horizon length increases (a form of anchoring and adjustment heuristic with adjustment set as zero). With the presence of a trend, this heuristic would result in a decrease in accuracy as the horizon is lengthened. However, in one of the few studies relating to currency forecasting, we (Wilkie and Pollock, 1994) found that professional forecasters performed worse in the short term. In this study, the professionals were compared to mathematicians (with no experience of currencies) and interesting horizon effects emerged although overall performance was similar. Overall, the study suggested that professionals and non-professionals are likely to be influenced differently by specific characteristics of the forecasting task.

In view of the literature cited above, this study is designed to explore time series extrapolative judgement in a currency forecasting context. The goal is to investigate the potential effects of trend, noise, and forecast horizon on judgemental probability forecasts based on abstract time series. The use of abstract series aids our attempts to discern the comparative forecasting performance of experts and non-experts operating under identical historical information. Accordingly, Section 2 presents the simulated data used in this study, and the methodology is given in Section 3. Section 4 provides the results, while

Section 5 presents conclusions and directions for further research.

2. Characteristics of exchange rate series and the simulation of the data

This section discusses the nature of exchange rate behaviour and the method by which the data used in the present study were obtained to exhibit the relevant characteristics. The principle feature of actual values of currency series is that they are not stationary: the variance and covariance depend on time even when logarithmic values are used. In particular, the variance tends to increase over time and first order serial correlation with a value close to unity is likely to be present. Series of this form can, however, be made stationary by some simple transformations. Taking first differences of the actual logarithmic values simultaneously takes out the effect of a linear trend in the series (i.e. giving constant drift in the difference data) and the autocorrelation (i.e. a first order serial correlation coefficient close to unity in the actual data has a value close to zero in the difference data). In other words, currency series tend to follow what Nelson and Plosser (1982) describe as a difference stationary process (i.e. non-stationary arising from the accumulation over time of stationary and inevitable first differences) rather than a trend stationary process (i.e. stationary fluctuations around a deterministic trend). In this difference stationary framework, the trend term in the actual series is associated with the drift term in the first differences. A constant drift gives rise to a linear trend and a variable drift gives rise to a non-linear trend. Zero drift implies that there is no trend.

The Efficient Markets Hypothesis (EMH) is often referred to as the random walk view and is supported by a number of studies (e.g. Crumby and Obstfeld, 1984; Boothe and Glassman, 1987). This view implies that currency movements follow an identical and independent distribution over time. This random walk process (for the actual logarithmic values) would tend to meander away from the starting value but exhibit no particular trend in doing so and is, therefore, dependent on its initial value and the

cumulative effect of random error movements from the initial period. Movements in this type of series are purely random with zero drift. As this type of series provides a basic starting point in examining currencies, it forms the basis of the first set of simulated series (i.e. Model 1) which is statistically defined below. The error term can be modelled as a normally distributed random variable.

The trend in the actual (logarithmic) series (drift in the logarithmic difference series) is the major characteristic in currency series that is of use to the forecaster when extrapolating from past and present values of the data. Both chartist and fundamental currency forecasting techniques are essentially designed to identify trends in financial series. The time series path of the spot exchange rate (as opposed to futures or forward exchange rates) often exhibits a major trend (e.g. an examination of the Swiss Fr./ UK £ clearly shows a relative depreciating £ over the last 20 years). Such a trend arises from fundamentals in the foreign exchange market, the most important of which is Purchasing Power Parity (PPP). PPP states that exchange rates adjust to offset differentials in relative price changes (i.e. inflation rates) between countries which can persist over the long term. Results from Officer (1982) and Pollock (1989a), (1989b), (1990b) support the long run validity of PPP. If it is assumed that relative price movements are roughly constant over time, the PPP view would support the presence of approximately linear trends in currency series: constant drift. As countries have differing rates of interest (high inflation countries tend to have higher rates of interest than low inflation countries), long term speculative gains on the movement of the currency would tend to be offset by interest rate differentials such that the trends can persist over time. An approximately linear trend in a logarithmic currency series is consistent with this view, hence it is appropriate to consider drift as non-zero and constant over time. This approach provides the second group of simulated series (i.e. Model 2). This model can have positive drift and negative drift and is consistent with the EMH if interest rate differentials fully explain the drift.

While major trends can persist over the long term, minor trends can occur due to the time it takes

information to be incorporated into exchange rates. Short term fundamentals can arise from asset market factors. These include: oil shocks arising from events such as the Iraqi invasion of Kuwait; political unrest in the former USSR; conflicts in the former Yugoslavia; and other political and economic changes or less spectacular events such as the resignation of a prime minister or an announcement of good trade figures. If information from such events is incorporated into the drift term over time, consecutive values will be influenced in the same direction causing the drift to show positive autocorrelation. That is, there would be an initial effect and subsequent effects that decrease over time, which is consistent with a short term variable drift pattern. This approach considers that over several periods the exchange rate moves in the same direction (subject to random variation and other things being equal) towards a mean (constant drift reflecting the major trend). If this mean is zero the model would suggest that the exchange rate is influenced by a series of events which form (by assumption) an irregular pattern. This pattern can be modelled by using a random error term that follows a normal distribution. Hence, the model contains two error terms, one that reflects pure random variation (as in the case of the random walk model) and another which reflects the effect of (random) events on drift, the effect of which decreases over time. This type of series provides the third group of simulated series – variable drift with a zero mean (i.e. Model 3).

The assumption made above of a zero mean can be relaxed to allow positive or negative drift in the longer term resulting in a price trend model which allows major and minor trends in the currency series. It is this type of series that provides the fourth type of simulated series (i.e. Model 4) – variable drift with a positive or negative mean. This model exhibits both major and minor trends around random fluctuations and can be justified in the same way as the above models. In this case, however, both constant and stochastic drift occur in the same model.

These four models, therefore, take into account both long and short term (major and minor) trends in the exchange rate. Model 1 contains no long term or short term influences. Model 2 considers only long

term influences, Model 3 considers only short term influences, and Model 4 considers both long term and short term influences. These four models can be simulated by defining the drift term as a linear and/or stochastic variable that follows a first order autoregressive (AR) process. Pollock (1990a) used various models of this form in the context of Italian Lira/UK £ exchange rate forecasting. In the examination of exchange rate behaviour, an AR(1) model for the drift term is an appropriate specification (Taylor, 1980, 1986). Taylor (1989) illustrates a method for constructing daily financial data. By choosing appropriate parameters, Taylor's procedure can be applied to monthly exchange rate data. The design of the simulated series (described above) was based on this price trend model with parameters chosen to reflect a random walk with: (i) zero drift Model 1; (ii) constant drift – Model 2; (iii) stochastic drift – Model 3 and; (iv) constant and stochastic drift – Model 4.

In modelling the noise component a natural choice is the normal distribution. We (Pollock and Wilkie, 1996; Pollock et al., 1996) have found for weekly forecasts of the US $/UK £ and JapaneseYen/German DM that the assumption of normally distributed first differences was appropriate if allowance was made for time varying parameters. The case for the assumption of normality is even stronger in the case of the longer horizon, monthly data.[2]

In order to examine the impact of noise on the judgemental identification of the major and minor trends, high and low variance specifications for the four models defined above were included. No attempt was made to incorporate changing variances within particular series: the identification of changing variances within a series is a difficult task without statistical analysis. Each series, therefore, was given a constant variance.

The simulated currency series were obtained by using a modification of the Price Trend model of Taylor (1989). This model is set out in Eq. (1) and (2):

$$\Delta y_t = T_t + \varepsilon_t \qquad (1)$$

$$(T_t - \mu) = \rho(T_{t-1} - \mu) + \nu_t \qquad (2)$$

where: Δ is a first difference operator and y_t is the logarithm of the exchange rate such that $\Delta y_t = y_t - y_{t-1}$; T_t is the drift term; ρ is the autocorrelation coefficient; μ is the mean of Δy_t; ε_t and ν_t are independent and identically distributed normal random variables with expected values of zero and variances of σ_ε^2 and σ_ν^2 respectively; A is defined as the signal to noise ratio $\sigma_\nu^2/\sigma_\varepsilon^2$; subscripts t and $t-1$ denote time; variances are $V(T_t) = \sigma_\nu^2/(1-\rho^2)$ and $V(\Delta y_t) = \sigma_{\Delta y}^2 = \sigma_\varepsilon^2 + \sigma_\nu^2/(1-\rho^2)$; and the initial values for y and T are set at $y_0 = 0$ and $T_0 = \mu$.

To set the parameters (ρ, $\sigma_{\varepsilon+\nu}$, A, μ), the actual series of monthly cross rates between five major currencies (UK Pound, US Dollar, Japanese Yen, German DM and Swiss Franc) were obtained for the period December 1973 to December 1994. The figures for each series were indexed to a value of unity for December 1973. Logarithmic values to base ten were then obtained so that the value for December 1993 became zero. The data were then first difference giving a series for the period January 1974 to December 1994. The means, standard deviations and first order autocorrelation coefficients were obtained for each series (see Table 1 for estimates). These estimates provided the guidelines on which the parameters of the models were defined.

Using the results in Table 1 as a guide and taking into account the need for appropriate values that allow some degree of judgemental recognition in the series, the parameters chosen for the simulated series are defined as in Table 2.

To compare an individual's judgemental predictions with the optimal, it was necessary to obtain theoretical expected point values for the 1-6 month ahead forecasts (i.e. for months 61-66). These are set out in Appendix A.

3. Methodology

Participants of this study came from two groups. One group consisted of ten members of the EURO Working Group on Financial Modelling. This 'expert' group was comprised of academics and prac-

---

[2] The Central Limit Theorem suggests that, as exchange rate changes between two points in time are essentially the sum of changes over shorter horizons, the distribution will tend to normality, even if the underlying distribution is not normal, provided this underlying distribution is stable.

Table 1
Estimated parameters for the price trend model

Major cross exchange rates – first differences Rate	January 1974 to December 1994		
	Mean	S.D.	Autocorrelation coeff.
US Dollar/UK Pound	− 0.0007	0.0147	0.102
Japan Yen/UK Pound	− 0.0025	0.0143	0.109
German DM/UK Pound	− 0.0016	0.0118	0.107
Swiss Franc/UK Pound	− 0.0023	0.0130	0.121
Japanese Yen/US Dollar	− 0.0018	0.0145	0.025
German DM/UK Dollar	− 0.0010	0.0148	− 0.002
Swiss Franc/UK Dollar	− 0.0016	0.0164	0.042
Japanese Yen/German DM	− 0.0008	0.0131	0.055
Swiss Franc/German DM	− 0.0006	0.0070	0.169
Japanese Yen/Swiss Franc	− 0.0002	0.0138	0.036

Table 2
Parameter set for the simulated series

Model	Parameters			
	p	$\sigma_{e,i,v}$	A	μ
Zero drift – Model 1				
Low noise	0	0.01	0	0
High noise	0	0.02	0	0
Constant drift – Model 2				
Low noise, positive constant drift	0	0.01	0	0.002533
Low noise, negative constant drift	0	0.01	0	− 0.002533
High noise, positive constant drift	0	0.02	0	0.002533
High noise, negative constant drift	0	0.02	0	− 0.002533
Stochastic drift – Model 3				
Low noise	0.5	0.01	0.25	0
High noise	0.5	0.02	0.25	0
Constant and stochastic drift – Model 4				
Low noise, positive constant drift	0.5	0.01	0.25	0.002533
Low noise, negative constant drift	0.5	0.01	0.25	− 0.002533
High noise, positive constant drift	0.5	0.02	0.25	0.002533
High noise, negative constant drift	0.5	0.02	0.25	− 0.002533

Note: For Models 3 and 4 the values of p and A of 0.5 and 0.25 respectively are consistent with a first order autocorrelation coefficient of 0.125.

titioners from different European countries. All of these individuals had considerable expertise in financial forecasting including knowledge of the nature of currency series and sufficient understanding of judgemental probability forecasting. Finally, these individuals were proficient with chartist techniques.

The second group consisted of 30 third-year management students taking a forecasting course at Bilkent University, Turkey. This 'novice' group was exposed to judgemental probability forecasting via their forecasting course, and had limited domain knowledge via a previously-taken finance course.[3]

Simulated data for the time paths of 32 series were

[3]Students in the 'non-expert' group were exposed to random walk processes and EMH concepts at an elementary level. These subjects' comparatively limited domain knowledge and minimal experience induces their classification as 'novices'. On the other hand, professional qualifications of the members of the EURO Working Group on Financial Modelling substantiate their identification as the 'expert' group.

presented graphically to the subjects. The subjects were not told anything about the nature of the data or that they were constructed, only that they reflected logarithmic values of currency series. The series were presented for a 60 month period (months were numbered from 1 to 60) and indexed with the initial value (for month 0) set at zero.

The subjects were asked to study each series and make directional forecasts over six horizons (i.e. for months 61 to 66). They were also required to indicate how certain they were about each prediction by assigning a probability (between 50% and 100%). The subjects completed the task at their own pace and convenience.

A comparison of subjects' predictions with expected probabilities were made using a range of probability accuracy measures which essentially involved the calculation of the Mean Absolute Probability Score (MAPS) and the associated measures of the Mean Response $\{M(r)\}$ and Bias (B). These essentially follow the lines of the covariance decomposition approach, set out in Yates (1982), (1988), but with modifications to take into account the magnitude of movements in the series (see Wilkie and Pollock, 1996). These are outlined below.

Once the subjects' forecasts were obtained a weighted outcome index (c_i) for each forecast i was calculated for each forecaster as defined in Eq. (3):

$$c_i = 0.5 + w_i \qquad (3)$$

To apply the proposed framework, it was necessary to calculate the weight (w_i) in the weighted outcome index (c_i) for each forecast i. As defined in Wilkie and Pollock (1996), the quantity, 0.5, plus the absolute value of this weight (i.e. $0.5 + |w_i|$) can be viewed as a probability that reflects the relative magnitude of a movement in the currency series at period i. The sign of w_i reflects whether the forecaster is correct $(+)$ or incorrect $(-)$. Since the series used in the present study were simulated, this weight was known with certainty as the signal and error terms could be identified. In this case, $\{0.5 + |w_i|\}$ was the theoretical probability of the predicted change in the series at forecast i (i.e. in the appropriate direction).

The subjects' performance was compared with the hypothetical random walk forecaster. The random walk forecaster assigns all probabilities as 0.5 with an arbitrary direction. An individual who views the currency market as efficient with exchange rate movements following a random pattern would make predictions in a similar way. The expected value of the weighted outcome index {i.e. $M(c) = \Sigma\ c_i/n$} for the random walk forecaster is 0.5.

The MAPS, which is closely related to the Mean Absolute Error (MAE), was computed using the modified outcome index. This is defined in Eq. (4):

$$MAPS = \Sigma\ |r_i - c_i|/n \qquad (4)$$

where r_i is the probability response for forecast i. The MAPS has an expected value for the random walk forecaster of $\Sigma\ |p|/n$.

The MAPS represents a form of linear loss function (the penalty attached to the error is proportional to the size of the error) in contrast to the widely used Mean Probability Score (MPS) which takes the form of a quadratic loss function (the penalty attached is proportional to the square of the error). It was considered more appropriate to use MAPS in this study as it is likely that the subjects would have tended, intuitively, to view the consequences of the error in a linear way. It has been pointed out by Keren (1991) that the loss function used in assessing probabilistic forecasting performance should be approximately consistent with the framework in which subjects make their predictions.

To supplement the interpretation of MAPS, two other accuracy measures were calculated. These measures were the Mean Response $\{M(r)\}$ and Bias $\{B = M(r) - M(c)\}$. Bias measures the degree of under/overconfidence in predictions. It is positive in cases of overconfidence and negative in cases of under confidence. The expected value of B is zero for the random walk forecaster.

The MAPS and associated measures, however, vary across the types of series with different characteristics and random variation with the result that interpreting a subject's performance between different situations becomes difficult. It was, therefore, appropriate to use a relative standard of comparison. In this study, the MAPS Difference (MAPSD) was used, which is defined as the difference between each subject's MAPS (and $M(r)$ and B) and the MAPS (and $M(r)$ and B, respectively) obtained from

applying a first order Autoregressive Model Order One {AR(1)} to the first differences of the series. Each subject's performance was, therefore, measured relative to the model, which facilitated comparisons of experts and novices based on various series characteristics. While the MAPS can only take positive values with the best possible measure attainable being zero, the MAPSD can take positive or negative values. A positive value would indicate that the subject's performance was worse than that of the AR(1) model and a negative value would indicate that the subject's performance was better. To provide additional information the $M(r)$ Difference {$M(r)$D} and Bias Difference (BD) were also considered. The AR(1) model was chosen because it has been used in a currency forecasting context (Pollock and Wilkie, 1992) and because it can be used to identify both the linear trend (constant drift) and the low level of autocorrelation (a feature of stochastic drift). Due to the statistical problems associated with the identification and separation of the two error terms (ε_t and η_t), variable parameter techniques were not considered suitable for providing a more appropriate model.

4. Results

A Split Plot (Mixed) ANOVA was applied to the dependent variable, MAPSD, with four independent factors: (1) Expertise (expert/novice); (2) Horizon (1–6 months); (3) Series Type (1, 2, 3 and 4 derived from Models 1, 2, 3 and 4 respectively, i.e. zero, constant, stochastic, and stochastic with constant drift); and (4) Noise (low/high). Expertise was a between-subjects factor and Horizon, Series Type and Noise were within-subjects factors. As the subjects were chosen from the members of the Euro Working Group on Financial Modelling (in the experts case) and management students at Bilkent University (in the novice case) they were treated as fixed factors. In addition, as there were 10 experts and 30 students, the ANOVA took the form of an unbalanced design. The four factor interaction terms were excluded from the analysis to provide the error term. To complement the results and provide additional information, the procedure was also repeated with Mean Response Difference {$M(r)$D} and Bias Difference (BD) as dependent variables. Accuracy

components such as the Weighted Outcome Index, Slope and Scatter could not be included as the zero drift model gives the same constant values in all cases. The mean values for the MAPSD, {$M(r)$D} and (BD) for single factor effects and two way interaction with series type are set out in Table 3 Table 4 Table 5. Table 3 also gives respective values of the MAPSD for the random walk forecaster relative to the AR(1) model.

Important single factor effects were highlighted by the analysis for the MAPSD. There was a significant expertise effect {$F(1,585) = 238.55, P < 0.001$} which reflected that experts clearly performed better than novices, although performance in both cases was poorer than the AR(1) model (Table 3). This was probably due to the experts giving a much lower mean response than the AR(1) model while the novices gave similar levels of response to the model but exhibited a poorer directional probability performance (Table 4). Hence, the experts' bias scores were similar to the model's, whereas the novices proved to be quite overconfident (Table 5). Expertise, therefore, did appear to improve performance. There was also a significant horizon effect {$F(5,585) = 16.15, P < 0.001$} which illustrated that, with the exception of the one month horizon, relative performance over the model improved as the horizon length increased (Table 3). The model, however, still performed better than the subjects in all horizons and better than the random walk forecaster. One explanation for this is that the subjects' mean response decreased relative to the model as the horizon increased, so that for horizons of 2 or more months, it was less than the AR(1) model (Table 4). The result was that the clear overconfidence relative to the model displayed for the 1-month horizon was significantly reduced to reveal slight under confidence for the 6-month horizon (Table 5). The subjects, therefore, appeared less confident of a drift persisting into the future than the model. The type of series also had a major effect {$F(5,585) = 1416.90, P < 0.001$}. The subjects performed similar to the random walk forecaster where the series contained a constant drift element but their performance was worse than the AR(1) model (Table 3). The subjects performed similarly to the AR(1) model in the zero drift case, but this performance decreased with the presence of stochastic and constant drift. Perform-

Table 3
MAPS differences – subjects and random walk forecaster

	Series type/drift				All	Noise	
	1 Zero	2 Constant	3 Stochastic	4 Stochastic and constant		Low	High
Expertise							
All	0.002	0.062	0.029	0.136	0.057	0.056	0.059
Novice	0.014	0.066	0.039	0.139	0.064	0.063	0.065
Expert	−0.032	0.051	0.001	0.127	0.037	0.034	0.039
	(−0.083)	(0.062)	(−0.027)	(0.135)	(0.022)	(0.012)	(0.031)
Horizon							
1 month	0.048	0.064	0.043	0.103	0.065	0.078	0.052
	(−0.046)	(0.025)	(−0.019)	(0.061)	(0.005)	(0.019)	(−0.009)
2 month	0.028	0.060	0.037	0.142	0.067	0.071	0.062
	(−0.063)	(0.045)	(−0.015)	(0.113)	(0.020)	(0.021)	(0.018)
3 month	0.008	0.063	0.036	0.136	0.061	0.059	0.063
	(−0.080)	(0.059)	(−0.019)	(0.135)	(0.024)	(0.016)	(0.032)
4 month	−0.013	0.057	0.024	0.146	0.054	0.049	0.058
	(−0.092)	(0.071)	(−0.028)	(0.154)	(0.026)	(0.010)	(0.042)
5 month	−0.025	0.062	0.021	0.141	0.050	0.041	0.058
	(−0.103)	(0.082)	(−0.037)	(0.167)	(0.027)	(0.005)	(0.049)
6 month	−0.032	0.066	0.016	0.148	0.049	0.039	0.060
	(−0.113)	(0.092)	(−0.045)	(0.179)	(0.028)	(0.001)	(0.056)
Noise							
Low	−0.038	0.039	0.078	0.144	0.056		
	(−0.137)	(0.020)	(0.022)	(0.144)	(0.012)		
High	0.042	0.084	−0.019	0.128	0.059		
	(−0.029)	(0.104)	(−0.077)	(0.126)	(0.031)		
Expertise/noise							
Novice/low	−0.024	0.046	0.085	0.147	0.063		
Expert/low	−0.080	0.021	0.058	0.138	0.034		
	(−0.137)	(0.020)	(0.022)	(0.144)	(0.012)		
Novice/high	0.051	0.085	−0.007	0.132	0.065		
Expert/high	0.016	0.082	−0.056	0.116	0.039		
	(−0.029)	(0.105)	(−0.077)	(0.126)	(0.031)		

Notes:
Means – single factor effects and two way interactions with series type and noise relative to AR(1) model {random walk relative to AR(1) model results in brackets}.
The lower the value the better the performance relative to the AR(1) model.
Positive values indicate performance worse than the AR(1) model and negative values indicate performance better than the AR(1) model.

ance was worse than the model in the stochastic drift case, much worse in the constant drift case, and worst of all in the constant with stochastic drift case. It would appear that the model was much better in picking up the constant drift, and to a lesser extent, the stochastic drift, than the subjects. The model's ability to identify the zero drift situation was similar to the subjects as a whole. The mean response

indicates that the subjects particularly underestimated the constant drift in the series, giving lower responses in these cases (Table 4). Subjects, however, were still overconfident relative to the model reflecting that they not only underestimated the drift but were poor at identifying it (Table 5). Hence, the model appeared, as would be expected, to perform much better than the subjects particularly where

Table 4
Mean response difference – subjects

| | Series type/drift | | | | All | Noise | |
	1 Zero	2 Constant	3 Stochastic	4 Stochastic and constant		Low	High
Expertise							
All	0.002	−0.073	0.014	−0.058	0.029	−0.031	−0.027
Novice	0.013	−0.064	0.028	−0.048	−0.018	−0.019	−0.017
Expert	−0.032	−0.101	−0.028	−0.087	−0.062	−0.068	−0.056
Horizon							
1 month	0.048	0.016	0.034	0.041	0.035	0.015	0.055
2 month	0.026	−0.034	0.022	−0.037	−0.006	−0.007	−0.005
3 month	0.008	−0.071	0.017	−0.068	−0.029	−0.027	−0.030
4 month	−0.013	−0.097	0.009	−0.083	−0.046	−0.043	−0.047
5 month	−0.025	−0.119	0.004	−0.096	−0.056	−0.058	−0.060
6 month	−0.032	−0.135	0.000	−0.106	−0.068	−0.066	−0.071
Noise							
Low	−0.039	−0.067	0.033	−0.051	−0.031		
High	0.042	−0.080	−0.004	−0.065	−0.027		
Expertise/noise							
Novice/low	−0.025	−0.058	0.047	−0.038	−0.019		
Expert/low	−0.079	−0.093	−0.008	−0.092	−0.068		
Novice/high	0.051	−0.070	0.010	−0.059	−0.017		
Expert/high	0.016	−0.109	−0.048	−0.083	−0.056		

Notes:
Means – single factor effects and two way interactions with series type and noise relative to AR(1) model.
Positive values indicate a higher mean response than the AR(1) model and negative values indicate a lower mean response than the AR(1) model.

constant drift was present in the series. Of the four factors, noise appeared the least important, giving non-significant results {$F(1,585) = 2.72$, ns}.

There were also important two-way interactions for the MAPSD. The interaction between expertise and series type was significant {$F(3,585) = 22.80$, $P < 0.001$} with the main difference occurring between the performance of experts and novices on the zero and stochastic drift cases (Table 3). For series types displaying a constant drift element, experts performed better than the random walk forecaster while the novices performed worse. It was, of course, impossible for the subjects to perform better than the random walk forecaster on the zero drift type and, as the expected directional movement and probabilities for the stochastic drift type approached those of the random walk series when the forecast horizon was increased (i.e. the expected effect of the

stochastic drift shock diminished over time), it was not surprising that subjects would perform worse than the random walk forecaster on this series type also. Experts performed better than the AR(1) model while the novices performed considerably worse. On all series types experts gave much lower probability responses than the novices and the AR(1) model (Table 4). For the zero drift series type, in particular, the experts showed under confidence relative to the model while novices showed overconfidence (Table 5). These results suggest that experts, who are familiar with the efficient market hypothesis and understand that currencies can often move in an apparently random way, are more ready to accept situations where they could not predict the direction of change in the series than novices. The interaction between horizon and series type was also significant {$F(15,585) = 27.22$, $P < 0.001$} with the best per-

Table 5
Bias difference – subjects

	Series type/drift				All	Noise	
	1 Zero	2 Constant	3 Stochastic	4 Stochastic and constant		Low	High
Expertise							
All	0.002	0.021	0.025	0.083	0.033	0.039	0.026
Novice	0.014	0.033	0.038	0.097	0.045	0.052	0.038
Expert	−0.032	−0.014	−0.013	0.040	−0.005	0.000	−0.010
Horizon							
1 month	0.048	0.080	0.049	0.116	0.074	0.068	0.079
2 month	0.026	0.044	0.035	0.126	0.058	0.062	0.053
3 month	0.008	0.032	0.028	0.098	0.041	0.045	0.037
4 month	−0.013	0.007	0.017	0.075	0.021	0.033	0.010
5 month	−0.025	−0.008	0.013	0.040	0.005	0.014	−0.005
6 month	−0.032	−0.029	0.009	0.042	−0.003	0.011	−0.016
Noise							
Low	−0.039	0.007	0.074	0.113	0.039		
High	0.042	−0.035	−0.024	0.052	0.026		
Expertise/noise							
Novice/low	−0.026	0.020	0.086	0.128	0.052		
Expert/low	−0.079	−0.029	0.040	0.069	0.000		
Novice/high	0.051	0.046	−0.010	0.066	0.038		
Expert/high	0.016	0.000	−0.066	0.012	−00.010		

Notes:
Means – single factor effects and two way interactions with series type and noise relative to AR(1) model.
Positive values reflect overconfidence relative to the AR(1) model and negative values reflect underconfidence relative to the AR(1) model.

formance occurring with zero drift and improving as the horizon increased. The subjects, in fact, performed better than the model in the longer horizons (Table 3). Performance on the other three types were, however, worse than the model and much more constant over the horizon. The mean probability responses indicated the subjects gave responses similar to those of the model. In particular, even though the probability responses were slightly higher than the model in the 1-month horizon, they declined as horizons increased (Table 4). While the zero and stochastic drift responses did not show a marked difference from the model, this was not the case when constant drift was present. In these cases, the responses were considerably less than the model. Overconfidence relative to the model was, however, greatest in this combined case but generally declined for all series types over the horizons (Table 5). It appears that the subjects' poor performance relative to the model reflected their inability correctly to

identify the constant drift situations. The subjects' performance tended to be worse than the model when constant drift occurred in a series, but when it was not present the subjects' performance was similar to the model. The interaction between drift type and noise was also significant {$F(3,585) = 644.96$, $P<0.001$}, indicating that the main differences occurred in series which did not contain constant drift (Table 3). In the zero drift case with low noise and in the stochastic drift case with high noise, the performance of the subjects was better than the AR(1) model. In terms of probability responses, the only marked difference occurred in the zero-drift case with much higher probability responses being given in the high-noise (greater than the model) compared with the low-drift case (less than the model) (Table 4). In this zero-drift high-noise case, predictions were more overconfident than the model but the low-noise predictions were more under confident (Table 5). This was reversed, how-

ever, in the stochastic drift case. These results suggest that the level of noise can have both positive and negative effects on judgemental extrapolation. There was also a significant interaction between horizon and noise $\{F(5,585)=21.46, P<0.001\}$ indicating that, as the forecast horizon increased, there was a consistent improvement in performance relative to the model in the low-noise case with a fairly constant performance in the high-noise case (Table 3). Performance was, however, worse than the model in all cases. The results for the probability responses did not indicate that this could be explained by differences in the probability responses (Table 4), but overconfidence tended to be higher in the low-noise case (Table 5). In fact, the high-noise situation with horizons of 5 months or more showed under confidence relative to the model. The poorer performance in the high-noise case at the longer horizons could be explained by less accurate directional probability responses. It appears that it was much easier for the subjects to identify the signal in the low-noise situation in comparison with the high-noise situation.

As for the three-way interactions for the MAPSD, expertise, drift type and noise were significant $\{F(3,585)=8.6, P<0.001\}$. Table 3 shows that the experts performed better than the novices on all four series types at both noise levels; however, marked differences occurred on the zero drift series with low noise (i.e. experts had a mean value of -0.080 as compared to that of the novices of -0.024) and the stochastic drift series with high noise (i.e. experts had a mean value of -0.056 as compared to that of the novices of -0.007). These results suggest that the experts were more skilled at identifying stochastic drift in series as well as distinguishing it from random fluctuations. Further evidence that the experts behaved differently where randomness was concerned is reflected in their mean probability responses over the four series types as compared with those of novices. The novices had higher mean responses than the experts in all cases but exhibited relative constancy across series types (i.e. 0.60, 0.60, 0.60 and 0.61, respectively). The experts, on the other hand, exhibited lower mean responses on the zero and stochastic drift series (i.e. 0.55, 0.57, 0.54 and 0.57 respectively). These results suggest that while the novices viewed the four series types as

being of roughly equal difficulty to forecast, the experts appreciated that series with random characteristics were particularly difficult to forecast. There was also a significant interaction between horizon, drift type and noise $\{F(5,585)=7.20, P<0.001\}$. This result indicated that the zero-drift case enabled better predictions relative to the model in all but the first horizon in the low-noise case (i.e. mean values for the 1 to 6 month horizon of: 0.013, -0.004, -0.028, -0.055, -0.070 and -0.085) and that with stochastic drift gave better predictions than the model over all horizons in the high-noise case (i.e. mean values for the 1-6 month horizon of: -0.009, -0.016, -0.015, -0.025, -0.023 and -0.029). This suggests that different levels of noise can have an influence on the identification of zero drift and stochastic drift series with the subjects' performance tending to improve relative to the AR(1) model as the forecast horizon is increased.

5. Conclusion

The present investigation reveals crucial insights for the financial forecasting domain. Our results suggest that experts' probabilistic currency forecasts are clearly more accurate than non-experts' forecasts. These findings confirm Whitecotton (1996) results regarding the superior accuracy of financial analysts' probabilistic earnings forecasts under conditions of constrained information.[4] Our findings are also congruent with the results of previous studies in financial markets showing better performance of experts under representative task conditions (Kabus, 1976; Muradoglu and Önkal, 1994; Önkal and Muradoglu, 1996).

Current results have important implications for financial decision making in that they extend the voluminous research demonstrating the accuracy of

---

[4] In efforts to examine comparative performance under shared information, Whitecotton (1996) presented analysts and students with limited financial ratios and previous earnings data, while hiding the company names and time frames. Analysts' probability forecasts were found to outperform undergraduates' forecasts, leading to the conclusion that experts could demonstrate their performance edge if given a constrained information set. Similarly, our subjects' data were constrained in that they were simulated and, therefore, cross-rate currency names were not supplied.

financial analysts' judgemental point forecasts, especially of earnings (Brown and Rozeff, 1978; Fried and Givoly, 1982; Armstrong, 1983; Collins et al., 1984; Brown et al., 1987; O'Brien, 1988; Schipper, 1991). Comparisons with time series models have suggested that the analysts' forecasting accuracy could largely be due to their use of non-time series information (see Brown (1993) for an extensive review). This suggestion has also been supported by the Affleck-Graves et al. (1990) study, which compared the earning forecasts of students (having only time series information) with those of analysts (having non-time series information as well), yielding superior accuracy for the latter group.

Following O'Connor and Lawrence (1989), we argue that a detailed investigation of time series extrapolative judgement necessarily entails eliminating non-time series information and exploring expert performance under those conditions. The current study presents such an attempt in a situation where the provision of time series information alone does not reduce ecological validity. We employ probability forecasts as a means for consolidating the inherent uncertainties in financial markets not reflected by point forecasts. Within a currency forecasting framework, we find that experts can effectively outperform non-experts under conditions of equal access to time series information. One potential explanation for this finding may involve the nature of expertise in currency forecasting. In particular, the experts in this study possessed specific knowledge of the nature of currency series in addition to their general knowledge of financial forecasting. Unlike the experts who had substantive knowledge about the existence and nature of random walk processes and market efficiency, students may not have been aware of the important theoretical implications of these concepts to currency forecasting, leading to poorer performance. Further research may test this assertion by concealing the currency identification of series and using participants with differing levels of expertise in financial forecasting.

Another explanation may relate to proposed arguments on potential hazards of experts' richer cognitive representations. As summarised by Whitecotton (1996), this view suggests that the presentation of selective information may serve to prevent the experts from using irrelevant and unproductive cues,

hence enabling better accuracy. Belatedly, Yates et al. (1991) maintain that increased experience within a domain leads to more beliefs being formed about what types of information are predictive of relevant target events. False beliefs are corrected relatively easily in domains where feedback is reliable (e.g. Kaiser and Proffitt, 1984); but in some complex systems the correction of erroneous beliefs is practically impossible. Consequently, greater experience in such systems can lead to a greater reliance on weak cues (e.g. Gaeth and Shanteau, 1984; Poses et al., 1985). Secondly, Yates et al. (1991) contend that, even if additional cues are valid, better performance is not guaranteed. For instance, lens model research has demonstrated that even the addition of valid cues can be detrimental to performance; additional cues cannot only be misused, but they can reduce the individual's reliability by making the task more difficult (Dudycha and Naylor, 1966). These arguments have direct implications for designing support systems to aid forecasters in effective and efficient processing of information. Future research examining forecasters' search for and use of different levels of contextual and time-series information may enhance our understanding of these important issues.

Another critical result emerging from the present study reflects the experts' ability to deal with random series. Not only is this expert ability superior to that of novices, but also it outperforms the AR(1) model. These results support the findings of Lawrence (1983); Edmundson et al. (1988) and Sanders and Ritzman (1992). The superior performance of human judgement in this case perhaps reflects two undesirable characteristics of models in general. Firstly, models tend to underestimate uncertainty because they cannot take all of its sources into account. Secondly, models attempt to identify signals in the data even when they are non-existent. Our experts, on the other hand, familiar with the characteristics of currency data, were able to accept that such series can exhibit random movements. In the present study, the experts were faced with a task which was, arguably, consistent with Ayton et al. (1989) criteria of being logically and methodologically appropriate, and this further supports the view that humans can recognise randomness (Baddeley, 1966; Cook, 1967; Harvey, 1988). Further research delineating the effects of feedback on such tasks would be extremely

valuable for users of judgemental forecasts (Benson and Önkal, 1992; Bolger and Wright, 1993, 1994; Önkal and Muradoglu, 1995; Yates et al., 1996).

Another interesting result is that the experts and the AR(1) model performed similarly on the stochastic drift series, with the experts significantly outperforming the model in the high-noise case. The comparative performance of experts and the AR(1) model supports Yaniv and Hogarth (1993) assertion that, in dynamic (high-noise) environments, humans may better utilise some infrequently-occurring cues that are difficult to include in statistical models. Accordingly, our results could also be viewed as suggesting that the experts were also able to concentrate on recent movements of the series as well as the overall trend. Support for this explanation comes from point forecasting studies concerned with the anchoring and adjustment heuristic (Bolger and Harvey, 1993; Goodwin and Wright, 1994; Lawrence and O'Connor, 1995). The relevance of this heuristic in a currency forecasting context could provide a promising direction for future research.

The interaction of series type and horizon is also intriguing. When the series contains no overall trend, subjects' performance, relative to the model is found to improve as the horizon is extended. However, when an overall trend is present, the subjects' performance, relative to the model is similar for all horizons. Not only do these results help explain the contradictory horizon effects discussed in the introduction but they suggest that the whole issue of the effect of forecast horizon on performance is much more complicated than was previously thought, and that it depends largely on the nature of the data and the experience of the forecaster. These issues also warrant further investigation.

Acknowledgements

Appendix A

Obtaining the theoretical expected direction and probability values of the series

The expected logarithmic exchange rate changes for one step ahead or more {i.e. $E(\Delta y_{t+\tau|t})$} are given by Eq. (A1).

$$E(\Delta y_{t+\tau|t}) = \mu + \rho^{\tau}(T_t - \mu) \tag{A1}$$

The optimal directional change for the 1–6 months ahead forecasts the sign of $E(\Delta y_{t+\tau|t})$ from Eq. (A1). The variances $V(\Delta y_{t+\tau|t})$ are given in Eq. (A2):

$$V(\Delta y_{t+\tau|t}) = \sigma_{\varepsilon}^2 + \sigma_{\eta}^2[(1 - \rho^{2\tau})/(1 - \rho^2)] \tag{A2}$$

The combined variances over τ periods are given in Eq. (A3):

$$\sum_{t-1}^{\tau} V(\Delta y_{t+\tau|t}) = \tau\sigma_{\varepsilon}^2 + \{\sigma_{\eta}^2/(1 - \rho^2)\}$$
$$\times \{\tau - \rho^2[(1 - \rho^{2\tau})/(1 - \rho^2)]\} \tag{A3}$$

The normally distributed values (z_τ) for the τ step ahead forecasts are given in Eq. (A4):

$$z_\tau = \{\tau\mu + \{\rho[(1 - \rho^{\tau})/(1 - \rho)]\}(T_t - \mu)\}/$$
$$\{\tau\sigma_{\varepsilon}^2 + \{\sigma_{\eta}^2/(1 - \rho^2)\}$$
$$\times \{\tau - \rho^2[(1 - \rho^{2\tau})/(1 - \rho^2)]\}\}^{1/2} \tag{A4}$$

As z_τ follows a standard normal distribution probability estimates for the directional change are directly obtained. That is, taking the absolute value of z_τ, $|z_\tau|$, the probability associated with the expected directional change was obtained from the cumulative distribution function of the standard normal distribution for the given values of z {i.e. $\Phi(|z_\tau|)$}. This probability has a minimum value of

0.5 {i.e. $\Phi(|z_r|) = 0.5$, when $z_r = 0$} and a maximum value of unity {i.e. $\Phi(|z_r|) = > 1$, when $|z_r| = > \infty$}.

References

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

Affleck-Graves, J., Davis, L.R., Mendenhall, R.R., 1990. Forecasts of earnings per share: possible sources of analysts' superiority and bias. Contemporary Accounting Research 6, 501–517.

Andreassen, P.B., 1988. Examining the price-volume relationship: the difference between price changes and changing prices. Organizational Behaviour and Human Decision Processes 41, 3711–3789.

Andreassen, P.B., Kraus, S.J., 1990. Judgemental extrapolation and the salience of change. Journal of Forecasting 9, 347–372.

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

Ayton, P., Hunt, A.J., Wright, G., 1989. Psychological conceptions of randomness. Journal of Behavioural Decision Making 2, 221–238.

Baddeley, A.D., 1966. The capacity for generating information by randomisation. Quarterly Journal of Experimental Psychology 18, 119–129.

Basi, B.A., Carey, K.J., Twark, R.D., 1976. A comparison of the accuracy of corporate and security forecasts of earnings. Accounting Review 51, 244–254.

Benson, P.G., Onkal, D., 1992. The effects of feedback on the performance of probability forecasters. International Journal of Forecasting 8, 559–573.

Bolger, F., Harvey, N., 1993. Context-sensitive heuristics in statistical reasoning. Quarterly Journal of Experimental Psychology 46A, 779–811.

Bolger, F., Wright, G., 1993. Coherence and calibration in expert probability judgement. Omega 21, 629–644.

Bolger, F., Wright, G., 1994. Assessing the quality of expert judgement: issues and analysis. Decision Support Systems 11, 1–24.

Boothe, P., Glassman, D., 1987. Comparing exchange rate forecasting models, accuracy versus profitability. Journal of Forecasting 3, 65–79.

Brown, L.D., 1993. Earnings forecast research: its implications for capital markets research. International Journal of Forecasting 9, 295–320.

Brown, L.D., Hagerman, R.L., Griffin, P.A., Zmijewski, M.E., 1987. Security analysts' superiority relative to univariate time-series models in forecasting quarterly earnings. Journal of Accounting Research 9, 61–87.

Brown, L.D., Rozeff, M.S., 1978. The superiority of analysts' forecasts as measures of expectations: evidence from earnings. The Journal of Finance 33, 1–16.

Collins, W.A., Hopwood, W., McKeown, J.C., 1984. The predictability of interim earnings over alternative quarters. Journal of Accounting Research 22, 467–479.

Cook, A., 1967. Recognition of bias in strings of binary digits. Perception and Motor Skills 24, 1003–1006.

Crumby, R.E., Obstfeld, M., 1984. International interest rate linkages under flexible exchange rates: a review of recent evidence. In: Bilson, J.F.O., Marston, R.C. (Eds.), Exchange Rate Theory and Practice. Chicago Press, Chicago.

Dalrymple, D.J., 1987. Sales forecasting practices. International Journal of Forecasting 3, 379–391.

Dudycha, L.W., Naylor, J.C., 1966. Characteristics of the human inference process in complex choice behaviour situations. Organizational Behaviour and Human Performance 1, 110–128.

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.

Fried, D., Givoly, D., 1982. Financial analysts' forecasts of earnings: a better surrogate for market expectations. Journal of Accounting and Economics 4, 85–107.

Gaeth, G.J., Shanteau, J., 1984. Reducing the influence of irrelevant information on experienced decision makers. Organizational Behaviour and Human Performance 33, 263–282.

Goodwin, P., Wright, G., 1991. Decision Analysis for Management Judgement. Wiley, Chichester.

Goodwin, P., Wright, G., 1994. Heuristics and biases and improvement strategies in judgemental time series forecasting. Omega 22, 553–568.

Harvey, N., 1988. Judgemental forecasting of univariate time series. Journal of Behavioural Decision Making 1, 95–110.

Kaiser, M.K., Proffitt, D.R., 1984. The development of sensitivity to causality. Child Development 55, 1614–1624.

Kabus, I., 1976. You can bank on uncertainty. Harvard Business Review 54, 95–105.

Keren, G., 1991. Calibration and probability judgements: conceptual and methodological issues. Acta Psychologica 77, 217–223.

Klein, H.E., Linneman, R.E., 1984. Environmental assessment: an international study of corporate practice. Journal of Business Strategy 5, 66–84.

Lawrence, M.J., 1983. The exploration of some practical issues in the use of quantitative forecasting. Journal of Forecasting 2, 169–179.

Lawrence, M.J., Makridakis, S., 1989. Factors affecting judgemental forecasts and confidence intervals. Organizational Behaviour and Human Decision Processes 42, 172–187.

Lawrence, M.J., O'Connor, M.J., 1992. Exploring judgemental forecasting. International Journal of Forecasting 8, 15–26.

Lawrence, M.J., O'Connor, M.J., 1995. The anchor and adjustment heuristic in time-series forecasting. Journal of Forecasting 14, 443–451.

Mosteller, F., Siegel, A.F., Trapido, E., Youtz, C., 1981. Eye fitting straight lines. The American Statistician 35, 150–152.

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

Murphy, A.H., Winkler, R.L., 1974. Subjective probability experiments in meteorology: some preliminary results. Bulletin of the American Meteorological Society 55, 1206–1216.

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

Murphy, J.J., 1986. Technical Analysis of Futures Markets. Prentice Hall, New York.

Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. Journal of Monetary Economics 10, 139–162.

O'Brien, P., 1988. Analysts' forecasts as earnings expectations. Journal of Accounting and Economics 10, 53–83.

O'Connor, M.J., Lawrence, M.J., 1989. An examination of the accuracy of judgemental confidence intervals in time series forecasting. Journal of Forecasting 8, 141–155.

Officer, L.H., 1982. Purchasing Power Parity and Exchange Rates: Theory, Evidence and Relevance. Contemporary Studies in Economic and Financial Analysis, Volume 35. J.A.I., London.

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

Onkal, D., Muradoglu, G., 1995. Effects of feedback on probabilistic forecasting of stock prices. International Journal of Forecasting 11, 307–319.

Onkal, D., Muradoglu, G., 1996. Effects of task format on probabilistic forecasting of stock prices. International Journal of Forecasting 12, 9–24.

Pollock, A.C., 1989. A model of UK real effective exchange rate behaviour. Applied Economics 2, 1563–1567.

Pollock, A.C., 1989. The time series characteristics of quarterly real and nominal lira/pound sterling exchange rate movements, 1973–1988. Rivista di Matematica per le Scienze Economiche e Sociali 12, 167–193.

Pollock, A.C., 1990. Forecasting quarterly movements of the lira/pound sterling: random walks, drift, seasonally and variable parameters. Rivista di Matematica per le Scienze Economiche e Sociali 13, 23–41.

Pollock, A.C., 1990. Alternative exchange rate forecasting models: an application to quarterly movements of the lira/pound sterling. Giornale degli Economisti e Annali di Economia 49, 527–555.

Pollock, A.C., Wilkie, M.E., 1992. Currency forecasting: human judgement or models?. VBA-Journaal, September, 21–29.

Pollock, A.C., Wilkie, M.E., 1996. The quality of bank forecasts: The dollar–pound exchange rate, 1990–1993. European Journal of Operational Research 91, 306–314.

Pollock, A.C., Wilkie, M.E., Pollock, J.E., 1996. A framework for examining the quality of bank currency forecasts: the Yen–DM, 1990–1993. Paper presented at the International Symposium on Forecasting, Istanbul, Turkey.

Poses, R.M., Cebul, R.D., Collins, M., Fager, S.S., 1985. The accuracy of experimental physicians' probability estimates for patients with sore throats. Journal of the American Medical Association 254, 925–929.

Remus, W., O'Connor, M.J., Griggs, K., 1995. Does reliable information improve the accuracy of judgemental forecasts?. International Journal of Forecasting 11, 285–293.

Ronis, D.L., Yates, J.F., 1987. Components of probability judgement accuracy: individual consistency and effects of subject matter and assessment method. Organizational Behaviour and Human Decision Processes 40, 193–218.

Sanders, N.R., 1992. Accuracy of judgemental forecasts. Omega 20, 353–364.

Sanders, N.R., Ritzman, L.P., 1992. The need for contextual and technical knowledge in judgemental forecasting. Journal of Behavioural Decision Making 5, 39–52.

Schipper, K., 1991. Commentary on analysts' forecasts. Accounting Horizons 5, 105–121.

Stael von Holstein, C.A.S., 1972. Probability forecasting: An experiment related to the stock market. Organizational Behaviour and Human Performance 8, 139–158.

Taylor, S.J., 1980. Conjectured models for trends in financial prices, tests and forecasts. Journal of the Royal Statistical Society, Series A 143, 338–362.

Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, Chichester.

Taylor, S.J., 1989. Simulating financial prices. Journal of the Operational Research Society 40, 567–569.

Tversky, A., Kahneman, D., 1973. Availability: A heuristic for judging frequency and probability. Cognitive Psychology 5, 207–232.

Wagenaar, W.A., 1972. Generation of random sequences by human subjects: a critical survey of the literature. Psychological Bulletin 77, 65–72.

Wagenaar, W.A., Sagaria, S.D., 1975. Misconceptions of exponential growth. Perception and Psychophysics 18, 416–422.

Wagenaar, W.A., Timmers, H., 1978. Extrapolation of exponential time series is not enhanced by having more data points. Perception and Psychophysics 24, 182–184.

Wagenaar, W.A., Timmers, H., 1979. The pond and duckweed problem: three experiments on the misperception of economic growth. Acta Psychologica 43, 239–251.

Webby, R., O'Connor, M.J., 1996. Judgemental and statistical time series forecasting: a review of the literature. International Journal of Forecasting 12, 91–118.

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

Wilkie, M.E., Pollock, A.C., 1994. Directional currency forecasting: an investigation into probability judgement accuracy. In: Peccati, L., Viren, M. (Eds.), Financial Modelling. Physica-Verlag, Heidelberg.

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

Wright, G., Lawrence, M.J., Collopy, F., 1996. The role and validity of judgement in forecasting. International Journal of Forecasting 12, 1–8.

Yaniv, I., Hogarth, R.M., 1993. Judgement versus statistical prediction: information asymmetry and combination rules. Psychological Science 4, 58–62.

Yates, J.F., 1982. External correspondence: decompositions of the mean probability score. Organizational Behaviour and Human Performance 30, 132–156.

Yates, J.F., 1988. Analyzing the accuracy of probability judgements for multiple events: an extension of covariance decomposition. Organizational Behaviour and Human Decision Processes 41, 281–299.

Yates, J.F., McDaniel, L.S., Brown, E.S., 1991. Probabilistic forecasts of stock prices and earnings: the hazards of nascent expertise. Organizational Behaviour and Human Performance 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.

Biographies: Mary E. WILKIE-THOMSON is a Lecturer in the Department of Consumer Studies, Glasgow Caledonian University, and has published articles and papers in a variety of books and journals. Her research interests focus on the role of judgement in financial forecasting and decision making.

Dilek ÖNKAL-ATAY is an Associate Professor of Decision Sciences at Bilkent University, Turkey. She received a Ph.D. in Decision Sciences from the University of Minnesota, and is doing research on decision analysis and probability forecasting. She has published in the European Journal of Operational Research, International Forum for Information and Documentation, International Journal of Forecasting, Journal of Behavioural Decision Making, and the Journal of Forecasting.

Andrew C. POLLOCK is a Lecturer in the Department of Mathematics, Glasgow Caledonian University. He completed a Ph.D. on exchange rates in 1988, and has subsequently published a variety of articles and papers in this area. His particular research interest is the application of analytical techniques to the forecasting of currency and, more generally, financial time series.