



ELSEVIER

International Journal of Forecasting 19 (2003) 241–256

*international journal
of forecasting*

www.elsevier.com/locate/ijforecast

The influence of trend strength on directional probabilistic currency predictions

Mary E. Thomson^{a,*}, Dilek Önköl-Atay^b, Andrew C. Pollock^c, Alex Macaulay^c

^a*Division of Risk, Glasgow Caledonian Business School, Cowcaddens Road, Glasgow G4 0BA, UK*

^b*Faculty of Business Administration, Bilkent University, 06533 Bilkent, Ankara, Turkey*

^c*Department of Mathematics, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, UK*

Abstract

An experiment is reported which compares the judgmental forecasting performance of experts and novices using simulated currency series with differing trend strengths. Analyses of directional probability forecasts reveal: (1) significant effects of trend strength on all aspects of predictive performance being studied, with evidence for the hard–easy effect where overconfidence is exhibited on weak (i.e., more difficult) trends, while underconfidence is shown on strong (i.e., less difficult) trends; (2) lower performance of experts on relative accuracy and profitability measures, reflecting experts' resistance to strong trends; (3) better overall performance on negative trends; and (4) superior performance of composite forecasts. Possible explanations are offered for these results and future research directions are outlined.

© 2001 International Institute of Forecasters. Published by Elsevier Science B.V. All rights reserved.

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

1. Introduction

The trend constitutes a common element of many financial time series, with potentially serious implications for forecasting accuracy (Tvede, 1990). In attempting to delineate the effects of trend and other time-series components on forecasting performance, previous research has primarily employed constructed or simulated series (e.g., Ang & O'Connor, 1991; O'Connor & Lawrence, 1992; Lawrence &

O'Connor, 1992, 1993; O'Connor, Remus & Griggs, 1993, 1997; Bolger & Harvey, 1995; Harvey, 1995; Lim & O'Connor, 1995; Remus, O'Connor, & Griggs, 1995, 1996; Harvey & Bolger, 1996; Webby & O'Connor, 1996; Wilkie-Thomson, Önköl-Atay, & Pollock, 1997; Pollock, Macaulay, Önköl-Atay, & Wilkie-Thomson, 1999). The use of simulated data is particularly advocated in this framework, since it is argued it enables detailed analyses of predictive accuracy by providing systematic control of relevant time-series characteristics while deterring the effects of non-time-series cues (Goodwin & Wright, 1993; O'Connor & Lawrence, 1989).

In examining the influence of trend via constructed series, past work has suggested that both the presence and the direction of trend (i.e., trend versus no

*Corresponding author. Tel.: +44-141-331-8954; fax: +44-141-331-3229.

E-mail addresses: mwi@gcal.ac.uk (M.E. Thomson), onkal@bilkent.edu.tr (D. Önköl-Atay), a.c.pollock@gcal.ac.uk (A.C. Pollock), abma@gcal.ac.uk (A. Macaulay).

trend, along with upward versus downward trend) affect the accuracy of judgmental point forecasts as well as prediction intervals (Lawrence & Makridakis, 1989; O'Connor & Lawrence, 1992; O'Connor, Remus & Griggs, 1997). Furthermore, judges appear to underestimate the trend in a series, displaying a bias called trend-damping (Eggleton, 1982; Sanders, 1992; Bolger & Harvey, 1993, 1996; Harvey, 1995). It is interesting to note that, despite the considerable attention devoted to trend as an efficacious time-series component, the potential influence of *trend-strength* on forecasting performance has been virtually ignored. This paper argues that the strength of trend represents a potent factor that can seriously alter the degree of predictability of a series. In addition, while previous research has almost exclusively focused on judgmental point forecasts and prediction intervals, we aim to investigate the effects of trend-strength on the accuracy of directional probability forecasts (i.e., forecasts of whether a future value of a variable will rise or fall, accompanied by a subjective probability reflecting the forecaster's degree of belief that the predicted direction will indeed occur). Probability forecasts are used extensively in economic and financial forecasting (Önköl-Atay, Wilkie-Thomson & Pollock, in press). Such forecasts reveal detailed information about the forecaster's uncertainty, acting as a basic communication channel for the transmission of this uncertainty to the users of these predictions, who can, in turn, better interpret the forecast information (Murphy & Winkler, 1984). However, even though such directional forecasts are commonly employed in business contexts, there has only been one study exploring the effects of trend on predictive accuracy. In particular, Bolger and Harvey (1995) presented undergraduate students with trended and untrended series, asking them to judge the probability that the next point would be below specific reference values. They found that the subjects appeared to be underconfident in their assessments (i.e., subjective probability distributions were *flatter* than they should have been) with greater underconfidence for the trended series. Subjects were also more underconfident in judging the probability that the next point would be above the given values. Given their particular task structures, the authors concluded that the *framing* of a prediction problem could have serious implications for the elicited confidence judgements.

Instead of presenting specific reference values and asking for a probability for a specified direction, the present study employs directional probabilistic predictions where the forecaster first predicts a direction of change, and then assesses a probability that the predicted direction will occur. This task structure is more representative within the realm of currency forecasting — a domain where considerable variations in trend-strength are clearly observable.

Our research focus concerns the influence of trend strength on the performance of judgmental currency forecasts provided in the form of directional probabilistic predictions. Currency forecasting constitutes an important application domain where judgmental predictions are predominantly used in response to the ubiquitous uncertainties facing the forecasters, and where predictive accuracy carries significant financial consequences (Pollock & Wilkie, 1992, 1993, 1996; Wilkie & Pollock, 1994, 1996). In this context, the current paper attempts to address a research gap by investigating the potential effects of trend strength on various dimensions of forecasting performance. In so doing, judgmental forecasts of experts and novices are compared to depict any differences in predictive accuracy that could be attributed to expertise. Accordingly, the next section discusses the characteristics of exchange rate series and their relevance to the simulated series used in the present study. This is followed by a description of the methods used and a discussion of the results attained. The paper concludes with the implications of the findings and suggested directions for future work in this area.

2. Characteristics of exchange rate series and their simulation

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 principal feature of actual values of currency series is that they are not stationary, i.e., the mean, variance and covariance depend on time. 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 differ-

ences of the actual values simultaneously takes out the effects of linear trend in the series (i.e., giving constant drift in the differenced 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 differenced data). In other words, currency series tend to follow what Nelson and Plosser (1982) describe as a difference stationary process (i.e., non-stationarity arising from the accumulation over time of stationary and invertible 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 zero drift implies that there is no trend. The difference-stationary form of exchange rate series has important implications for the simulation of series: it is more appropriate to generate data using first differences than actual values (Wilkie-Thomson et al., 1997; Pollock et al., 1999).

The Efficient Markets Hypothesis (EMH), which portrays the view that currency movements follow an identical and independent distribution over time, is supported by a number of studies (e.g., Crumby & Obstfeld, 1984; Boothe & Glassman, 1987). Such a random walk process 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. This type of series provides a basic starting point in examining currencies and it forms the basis for the simulated series that are statistically defined below. The error term can be modelled as a normally distributed random variable.

The trend in the actual series (drift in the differenced 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. 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 that can persist over the long term, with results from Officer (1982) and Pollock (1989) supporting 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, i.e., constant drifts. 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 currency series is consistent with this view, hence it is appropriate to consider drift as non-zero and constant over time. This model can have a positive or negative drift and is consistent with the EMH if interest rate differentials fully explain the drift. Thus, a random walk with drift model would take into account long term (major) trends in the exchange rate.

In modelling the noise component, a natural choice is the normal distribution. It has been shown that, for weekly forecasts of exchange rate between the US\$ and UK£, the assumption of normally distributed first differences would be appropriate if allowance is made for time varying means and standard deviations (Pollock & Wilkie, 1996). The case for the assumption of Normality is even stronger in the case of the longer horizon, monthly data. 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 the underlying distribution is stable.

To examine the identification of drift strength, it is necessary to construct simulated data in a way that balances characteristics of exchange rate series with the experimental requirements of presenting graphs that would be acceptable and realistic to the subjects. Hence, the difference-stationary framework discussed above was adopted in constructing the data. In order to examine the impact of noise on the judgemental identification of the major trends an error term was included which followed a Standard

Normal distribution (i.e., zero mean and unit variance). 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 unit variance. It is also worth noting that the ratios of drift relative to the standard deviation used in this study are consistent with the empirical evidence from the currency markets.

The simulated currency series were obtained by generating first differences as set out in Eq. (1):

$$\Delta Y_t = \mu + e_t \text{ for } t = 1, 2, \dots, 60 \quad (1)$$

where Δ is a first difference operator and Y_t is the logarithm of the exchange rate so that $\Delta Y_t = Y_t - Y_{t-1}$; μ is the drift term or the mean of ΔY_t ; $\{e_t\}$ are independently and identically distributed Normal random variables with expected value of zero and variance of unity. In this study μ is set at ± 0.2533 , ± 0.5244 , ± 0.8416 and ± 1.2816 .

The resulting undifferenced series, Y_t , was obtained by setting a starting value $Y_0 = 1000$ at $t = 0$, using Eq. (2):

$$Y_t = Y_0 + t\mu + \sum_{i=1}^t e_i \text{ for } t = 1, 2, \dots, 60. \quad (2)$$

The starting value of 1000 was chosen largely for presentation purposes. This value allows an exchange rate range that is consistent with the use of linear trends without the need for a logarithmic transformation of the series. Lower starting values, given the drift and standard deviation, would have resulted in an inappropriate linear trend since the series values would have come close to the horizontal axis for the strong trends. The choice of the starting value has, of course, no impact on the actual monthly movements of the series or the probabilities.

To compare an individual's judgemental predictions with the optimal, it was necessary to obtain theoretical expected changes, $E(\Delta Y_{61})$, for the 1 month ahead forecasts (i.e., for month 61). That is: $E(\Delta Y_{61}) = \mu$. As the variance of e_t is unity, the expected cumulative probability for a rise in the series (α_{61}) is given as $\alpha_{61} = \Phi\{\mu\}$ where Φ is the cumulative distribution function of the Standard Normal. Therefore, drift (μ) of -1.2816 would be

associated with a probability of a rise (α_{61}) of 0.1 (or fall of 0.9), -0.8416 with probability 0.2 (0.8), -0.5244 with probability 0.3 (0.7), -0.2533 with probability 0.4 (0.6), 0.2533 with probability 0.6 (0.4), 0.5244 with probability 0.7 (0.3), 0.8416 with probability 0.8 (0.2) and 1.2816 with probability 0.9 (0.1).

3. Method

3.1. Subjects

Two groups participated in the experiment. The first group (i.e., the 'expert' group) comprised 18 members of the EURO Working Group on Financial Modelling. These participants were academics and practitioners from different European countries who were familiar with chartist techniques. They all had considerable expertise in financial forecasting, including knowledge of the nature of currency series and understanding of judgmental probability forecasting. The second group (i.e., the 'novice' group) consisted of 18 fourth-year students who were undertaking the Financial Studies option of the B.Sc. Mathematics for Business Analysis course at Glasgow Caledonian University. These participants had sufficient knowledge to understand the task, (i.e., an understanding of the nature of exchange rate series, probability and chartist techniques) but had little practical experience of judgemental financial forecasting.

3.2. Materials

Simulated data for the time paths of 24 series were presented graphically to the subjects. The subjects were not told how the data were constructed, only that they were obtained through a statistical procedure to simulate currency price series. These series were presented graphically for a 60-month period (months were numbered from 1 to 60) and indexed with the initial value in month 0 set at 1000.

The data were based on six randomly generated series from a Standard Normal distribution. Cumulative values were then obtained and the series were given a starting value of 1000. To each series, drift of varying magnitude was added. The drift could be

categorised as: (i) mild: drift of ± 0.2533 giving a probability for an increase or decrease, as appropriate, of 0.6; (ii) medium: drift of ± 0.5244 giving a probability of 0.7; (iii) strong: drift of ± 0.8416 giving a probability of 0.8; and (iv) very strong: drift of $\pm 0.1.2816$ giving a probability of 0.9. For each of these probabilities, three positive and three negative trends were obtained. This resulted in 24 random walks with varying degrees of drift series (12 positive and 12 negative). These series were presented to the subjects in a random fashion.

It should be noted that the data constructed in this way contain only one signal (drift) for subjects to identify. When actual data are used, a variety of signals may or may not be present, and this makes it almost impossible to distinguish between valid and invalid interpretations of the data.

3.3. Procedure

The subjects were instructed to study each series and make directional forecasts for the subsequent 1-month horizon. They were also asked to indicate how ‘certain’ they were about their predicted directions by assigning probabilities using a half-range percentage scale from 50 to 100% (see Appendix A for specific instructions). A percentage probability response of 50% implies a no change prediction. These percentages were then expressed as proportions for the resulting analysis (i.e., 0.5 to 1). The participants were permitted to complete the task at their own convenience and at their own pace.

3.4. The probability performance measures

On completion of the task, four probability performance measures were computed for each group: the weighted outcome index $\{M(c)\}$, the relative root mean probability score $\{U_{RMPS}\}$, mean response $\{M(r)\}$, and bias $\{B\}$. These measures are defined below and are essentially based on Yates’ (1982) covariance decomposition approach. They follow the framework described in Wilkie and Pollock (1996), which suggests that by weighting the outcome index in Yates’ various prescribed statistics to take into account the relative magnitude of change in a series, measures of performance more suitable to currency forecasting contexts can be obtained. Specifically,

the effect of this particular modification on the scoring rule is such that errors are penalised more heavily when incorrect decisions are made with large-scaled changes than with small-scaled changes, and forecasters are not discouraged from using low probabilities to the same extent as in the original form. When the scaled change in a series is close to zero, the best possible assigned probability is a value close to 0.5, whether in the correct direction or not. This is particularly important in currency forecasting as probability responses are often in the lower half of the 0.5 to 1 range due to the difficulty of forecasting quasi-random walk series. In addition, the modification uses more information from currency series than the original: it takes advantage of the precise values of the data which, in turn, results in more stable accuracy measures that can be used to examine performance over a relatively short set of predictions.

The Mean Weighted Outcome Index $\{M(c)\}$ measures the correctness of the directional forecast taking into account the magnitude of movement in the series. As pointed out above, in terms of accuracy, the magnitude of movement is vitally important whether or not the selected direction is correct. To obtain this measure it is first necessary to calculate a weighted outcome index (c_i) for forecast period i . This is defined in Eq. (3):

$$c_i = 0.5 + p_i \quad (3)$$

where p_i is a weight that reflects the scaled movement in the exchange rate between the beginning and end of forecast period i . The sign of p_i indicates whether the correct direction of the exchange rate movement was forecast. If the correct direction was predicted, p_i is positive and c_i is greater than 0.5. If the incorrect direction was predicted, p_i is negative and c_i is less than 0.5. A zero value of p_i corresponds to a situation where the exchange rate has not changed between the beginning and end of the forecast period. Thus p_i takes a value between -0.5 and $+0.5$. The scaled magnitude of the exchange rate change is reflected in the absolute value ($|p_i|$) which takes a value between zero and 0.5. Hence, p_i is a compound variable that reflects both the absolute relative magnitude of the exchange rate change and whether or not the predicted direction of change is correct. This variable is crucial to the application of

the approach and is considered further below. The resulting c_i takes a value between zero and unity and can be viewed as a continuous variable. The mean weighted outcome index $\{M(c)\}$ is the mean of the c_i 's, viz. $\sum c_i/n$, where n is the number of forecasts.

To apply the proposed framework, it is necessary to calculate the weight (p_i) for each forecast i . The quantity $0.5 + |p_i|$ can be viewed as a probability that reflects the relative magnitude of movement in the currency series at period i . Since the series used were simulated, the weight (p_i) was known as the signal and error terms could be identified. Thus the quantity $0.5 + |p_i|$ is the theoretical probability of the predicted change in the series for forecast i (i.e., in the appropriate direction). As values of p_i are positive when the forecaster is correct and negative when the forecaster is incorrect the weighted outcome index (c_i) would take a value of $0.5 + |p_i|$ or $0.5 - |p_i|$, respectively. For instance, in the case of the very strong magnitude drift, p_i would take a value of 0.4 when correct or -0.4 when incorrect resulting in a weighted outcome index of 0.9 or 0.1, respectively.

When analysing judgement, it is appropriate to compare the subjects' performance with that of a hypothetical random walk forecaster. The random walk forecaster assigns all probabilities as 0.5 with an arbitrary direction. The expected value of mean weighted outcome index $\{M(c)\}$ for the random walk forecaster is 0.5.

The Mean Response $\{M(r)\}$ is the mean of the r_i 's, viz. $\sum r_i/n$, where r_i (which is between 0.5 and 1) is the probability response for forecast period i ignoring whether or not the prediction is in the correct direction. The Mean Response $\{M(r)\}$ has, of course, a value of 0.5 for the random walk forecaster.

Bias, B , is the difference between the mean response and the mean weighted outcome index $\{B = M(r) - M(c)\}$ and measures the degree of overconfidence (if positive) or underconfidence (if negative) in the assignment of probabilities and directional responses. This measure has a value of zero for the random walk forecaster.

The relative root mean probability score (U_{RMPS}) is an overall probability accuracy measure and is defined in Eq. (4):

$$U_{\text{RMPS}} = \sqrt{\{[\sum (r_i - c_i)^2/n] / [\sum p_i^2/n]\}}. \quad (4)$$

This statistic involves a quadratic loss function (attaching a penalty to error that is proportional to the square of the error) and measures performance relative to the random walk forecaster. For the random walk forecaster the root mean probability score (U_{RMPS}) would have a value of unity. A subject whose U_{RMPS} score is less than unity is performing better than the random walk forecaster. The upper value for this statistic is theoretically infinite as $p_i = 0$, for all i , when the series is in the form of a constant. For practical purposes, given the series used in the present study, values in excess of 2 would indicate very poor performance.

3.5. The profitability measure

The measures discussed above do not provide an indication of the profit or loss that would have accrued to a market participant acting on a set of predictions. A measure of profitability is, therefore, appropriate. We consider the profit or loss on taking a particular action (i.e., long or short position), assuming an absence of transactions costs. We describe a method for measuring profitability which considers the predicted probability along with the direction of movement. As an illustration, suppose that a forecaster has asset holdings in two currencies — e.g. US\$ assets and UK£ assets — which are held in equal proportions at the beginning of the forecast period. For example, if we assume that, initially, he holds total assets of \$3m {which is equivalent to £2m at the current US\$ per unit UK£ (\$:£) rate of 1.5}, then his holdings, divided equally, would be \$1.5m and £1m. These proportional holdings can be adjusted in relation to a directional probability prediction. For example, if the forecaster predicts that the probability of an appreciation of the UK£ against the US\$ is 0.75, then his holdings can be adjusted in the \$:£ ratio 25%:75%, which amounts to holdings at the current exchange rate of \$0.75m and £1.5m. Now if we assume that the \$:£ exchange rate rises by 10% (i.e., to 1.65), then the total value of his holdings, in terms of \$'s, would be $\{0.75m + 1.65(1.5m)\} = \$3.225m$. His profit would therefore be \$0.225m or 7.5%. Nevertheless, it is necessary to take into account that in this case the individual would have made a profit in terms of \$ even if he had kept his holdings in equal proportions. In this case, the value

would have been $\$1.5\text{m} + 1.65(1\text{m}) = \3.15m , giving a profit of $\$0.15\text{m}$ or 5%. Hence the additional profit, in terms of \$, from acting on the probability recommendation would be $\$0.225\text{m} - \$0.15\text{m} = \$0.075\text{m}$ (2.5%). More formally, the additional percentage profit or loss (π_i^*) from taking a particular action in month t (i.e., long or short position) in respect of currency i , over a 1 month horizon, ignoring transactions costs, is defined in Eq. (5):

$$\pi_i^* = 100[(|Y_{t+1,i} - Y_{t,i}|)/Y_{t,i}]v_i. \quad (5)$$

Here, $Y_{t,i}$ is the actual exchange rate when the prediction is made (i.e., the value at the end of the current month), $Y_{t+1,i}$ is the actual exchange rate at the end of the predictive horizon, $|Y_{t+1,i} - Y_{t,i}|$ is the absolute value of the change in the exchange rate, and v_i , set out in Eq. (6), is a variable dependent on: (i) whether the prediction is in the correct direction: this is described by an indicator variable d_i taking values 1 or 0 depending on whether or not the direction is correct; and (ii) the probability response, r_i :

$$v_i = (2d_i - 1)(r_i - 0.5). \quad (6)$$

Using the above example to illustrate, Eq. (6) gives $v_i = 0.25$, hence Eq. (5) gives the percentage profitability, in terms of \$, as $\pi_i^* = 100(0.1)0.25 = 2.5\%$. The profitability per month measure for a specific horizon would, therefore, reflect the percentage monthly profit from these holdings expressed in terms of \$'s (say $\pi_{\$,i}^*$). But profitability can also be expressed in terms of £'s (say $\pi_{£ ,i}^*$). The average (π_i) of the two profitability measures with respect to the £ and \$ {i.e. $\pi_i = (\pi_{\$,i}^* + \pi_{£ ,i}^*)/2$ } is taken to remove any bias arising from using only one currency direction. The mean profitability measure, $M(\pi)$, is defined as the mean of the π_i 's, viz. $\sum \pi_i/n$.

In this study, as simulated series were used, the signals were actually known. It was, therefore, more appropriate to use the expected value at the end of the predictive horizon rather than the actual value. This is because the actual value would be influenced by the random variation from the error term. The expected value is, therefore, used as $Y_{t+1,i}$ in Eq. (5) to calculate profitability.

3.6. Composite forecasts

Measures for a *composite forecaster* were derived for each group (experts and novices) by aggregating the results for all the subjects in the group as follows. For a given series and forecast horizon, the probability predicted by the subject was transformed, if necessary, into a probability of a rise in the exchange rate. (For example, if a subject predicted a rise in the rate with probability 0.7, this was left as 0.7; whereas, if the subject predicted a fall in the rate with probability 0.7, this was transformed to $1 - 0.7 = 0.3$.) These probabilities were obtained for all 18 individuals in each group and the mean was calculated. If the mean probability was 0.5 or above, the composite forecaster was deemed to give a directional prediction of a rise with the assigned probability equal to this mean probability. If the mean probability was below 0.5, the composite forecaster was deemed to give a directional forecast as a fall with the assigned probability equal to unity less this mean probability. The calculations were repeated for all series and forecast horizons giving a complete set of forecasts for the expert composite forecaster and the novice composite forecaster.

4. Results and discussion

The mean, across the 24 series, of each of the four performance measures and the single profitability measure discussed above were each analysed as a single factor (expertise) fixed effects design using 4 (strength of trend) by 2 (direction of trend) repeated measures (see Appendix B for details of the model). The ANOVA results are displayed in Table 1. The group mean scores for each of the five performance measures for the 16 cells defined by the levels of expertise, strength of trend and direction are displayed in Table 2.

Table 1 clearly depicts the importance of trend strength for forecasting performance. Mean scores for the overall measures demonstrate that predictive performance improves with the strength of trend, and this is supported by highly significant trend effects on all measures: on $M(c)$ $\{F(3,102) = 434.81, P < 0.001\}$, on U_{RMPS} $\{F(3,102) = 119.99, P < 0.001\}$, on $M(r)$ $\{F(3,102) = 234.80, P < 0.001\}$, on B

Table 1
ANOVA F -ratio results for trend strength, direction, and expertise

	Performance measures				
	$M(c)$	U_{RMPS}	$M(r)$	B	$M(\pi)$
Strength	434.81***	119.99***	234.80***	53.69***	884.16***
Direction	0.41	5.72*	0.01	0.25	37.79***
Expertise	0.80	7.17**	14.99***	3.73	21.63***
Stren.*Dir.	0.85	0.20	4.32**	0.89	16.58***
Dir.*Exp.	0.07	0.50	0.84	0.74	0.05
Stren.*Exp.	0.80	0.08	5.72**	3.05*	3.52*
Exp.*Stren.*Dir	0.81	1.22	2.63	2.44	0.48

Note: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

$\{F(3,102) = 53.69, P < 0.001\}$, and on $M(\pi)$ $\{F(3,102) = 884.16, P < 0.001\}$.

The direction of trend appears to influence the relative accuracy (as indexed by U_{RMPS} $\{F(1,102) = 5.72, P < 0.05\}$), as well as $M(\pi)$ $\{F(1,102) = 37.79, P < 0.001\}$, with no significant effects on the mean weighted outcome index, mean probability response, and bias (under/overconfidence).

Regarding expertise, Table 1 shows a significant inverse expertise effect on overall U_{RMPS} $\{F(1,34) = 7.17, P < 0.01\}$ and on $M(\pi)$ $\{F(1,34) = 21.63, P < 0.001\}$, indicating better relative accuracy and higher profitability for novices' directional probability forecasts as compared to those of experts. A significant main effect is also found for $M(r)$ $\{F(1,34) = 14.99, P < 0.001\}$, affirming the higher probabilities assigned by the novice group.

The main effects discussed above have to be interpreted with caution, however, since Table 1 displays significant interactions for trend strength by direction as well as trend strength by expertise. In particular, trend strength by expertise interactions are found for three of the performance measures: for $M(r)$ $\{F(3,102) = 5.72, P < 0.01\}$, for B $\{F(3,102) = 3.05, P < 0.05\}$, and for $M(\pi)$ $\{F(3,102) = 3.52, P < 0.05\}$. These interactions are illustrated in Fig. 1a, b and c, respectively, and in Table 2. Fig. 1a clearly indicates that, although experts exhibited slightly more confidence (as revealed by higher probability assessments) than the novices on the weakest trend, they were less able to accept the stronger trends, assigning lower probabilities than novices for drift strengths of 0.7, 0.8 and 0.9. As a result of these mean probability profiles of experts and novices, we can observe from Fig. 1b that experts appear to be

slightly more overconfident than the novices on the mild trend condition (drift strength of 0.6), while displaying more underconfidence than novices for strong and very strong trend conditions (drift strengths of 0.8 and 0.9). Fig. 1b may also be interpreted as revealing evidence of the hard–easy effect, with overconfidence on the more difficult trends (i.e., trends that are harder to discern, with drift strengths of 0.6 and 0.7) and underconfidence on the less difficult trends (i.e., strong trends that are easier to detect, with drift strengths of 0.8 and 0.9).

Experts' probability profiles are also largely responsible for the profitability differences shown in Fig. 1c. It can be seen that, while experts and novices attain similar scores in the case of the weakest trend, novices improve their profitability more than the experts as the trend gets stronger.

Taken together, the results may be viewed as suggesting that the inferior overall performance of the experts could be due to their underestimation of strong trends (high magnitude drift). Perhaps, in retrospect, this is not too surprising since these individuals are likely to have been 'nourished' with the Efficient Market Hypothesis (EMH) in the course of their academic upbringing making them sceptical that the stronger trends would continue into the future. Although the experts are clearly able to recognise a trend, as illustrated in Tables 1 and 2, these individuals would always have in mind that trends can change in currency series and this together with an academic leaning towards random walk theory may have resulted, in some instances, in an explicit search for randomness in the face of contradictory evidence. That is, these subjects may have had a tendency to view trends, to a larger extent than

Table 2
Mean performance scores for trend strength, direction and expertise

Trend strength and direction	Performance measures				
	$M(c)(\uparrow)$	$U_{\text{RMPS}}(\downarrow)$	$M(r)$	$B(0)$	$M(\pi)(\uparrow)$
<i>Experts</i>					
0.6 — Pos.	0.570	1.261	0.662	0.091	0.003
0.6 — Neg.	0.556	1.267	0.630	0.075	0.002
0.6 — All	(0.563)	(1.264)	(0.646)	(0.083)	(0.003)
0.7 — Pos.	0.619	0.835	0.666	0.047	0.006
0.7 — Neg.	0.596	0.796	0.676	0.080	0.006
0.7 — All	(0.607)	(0.816)	(0.671)	(0.064)	(0.006)
0.8 — Pos.	0.767	0.506	0.730	−0.36	0.017
0.8 — Neg.	0.800	0.361	0.759	−0.41	0.023
0.8 — All	(0.783)	(0.433)	(0.744)	(−0.039)	(0.020)
0.9 — Pos.	0.885	0.361	0.813	−0.072	0.036
0.9 — Neg.	0.900	0.281	0.828	−0.072	0.046
0.9 — All	(0.893)	(0.361)	(0.821)	(−0.072)	(0.041)
All	{0.712}	{0.708}	{0.720}	{0.009}	{0.017}
<i>Novices</i>					
0.6 — Pos.	0.559	1.230	0.636	0.077	0.002
0.6 — Neg.	0.567	1.056	0.624	0.057	0.002
0.6 — All	(0.563)	(1.142)	(0.630)	(0.067)	(0.002)
0.7 — Pos.	0.626	0.834	0.716	0.090	0.009
0.7 — Neg.	0.633	0.588	0.668	0.034	0.008
0.7 — All	(0.630)	(0.711)	(0.692)	(0.062)	(0.009)
0.8 — Pos.	0.789	0.322	0.790	0.001	0.023
0.8 — Neg.	0.800	0.323	0.796	−0.004	0.026
0.8 — All	(0.794)	(0.323)	(0.793)	(−0.002)	(0.025)
0.9 — Pos.	0.885	0.276	0.840	−0.045	0.040
0.9 — Neg.	0.885	0.221	0.875	−0.010	0.051
0.9 — All	(0.885)	(0.248)	(0.858)	(−0.028)	(0.046)
All	{0.718}	{0.606}	{0.743}	{0.025}	{0.020}
<i>Experts and novices</i>					
0.6 — Pos.	0.565	1.245	0.649	0.084	0.003
0.6 — Neg.	0.561	1.161	0.627	0.066	0.002
0.6 — All	(0.563)	(1.203)	(0.638)	(0.075)	(0.002)
0.7 — Pos.	0.622	0.834	0.691	0.069	0.007
0.7 — Neg.	0.615	0.692	0.672	0.057	0.007
0.7 — All	(0.619)	(0.743)	(0.682)	(0.063)	(0.007)
0.8 — Pos.	0.778	0.414	0.760	−0.018	0.020
0.8 — Neg.	0.800	0.342	0.777	−0.023	0.025
0.8 — All	(0.789)	(0.378)	(0.769)	(−0.020)	(0.022)
0.9 — Pos.	0.885	0.318	0.826	−0.058	0.038
0.9 — Neg.	0.893	0.251	0.852	−0.041	0.048
0.9 — All	(0.889)	(0.285)	(0.839)	(−0.050)	(0.043)
All	{0.715}	{0.657}	{0.732}	{0.017}	{0.019}

Note: \uparrow high values are best; \downarrow low values are best; 0 zero is the best value.

novices, as being influenced by stochastic factors rather than deterministic factors (Wilkie-Thomson et al., 1997). The experts may also have been influenced by the more recent changes in the series (Tagaki, 1991) in comparison to changes in the more

distant past, whereas the novices may have put more emphasis on the overall trend.

Table 1 also reveals significant trend strength by trend direction interactions on the $M(r)$ measure [$F(3,102) = 4.32$, $P < 0.01$], and on $M(\pi)$

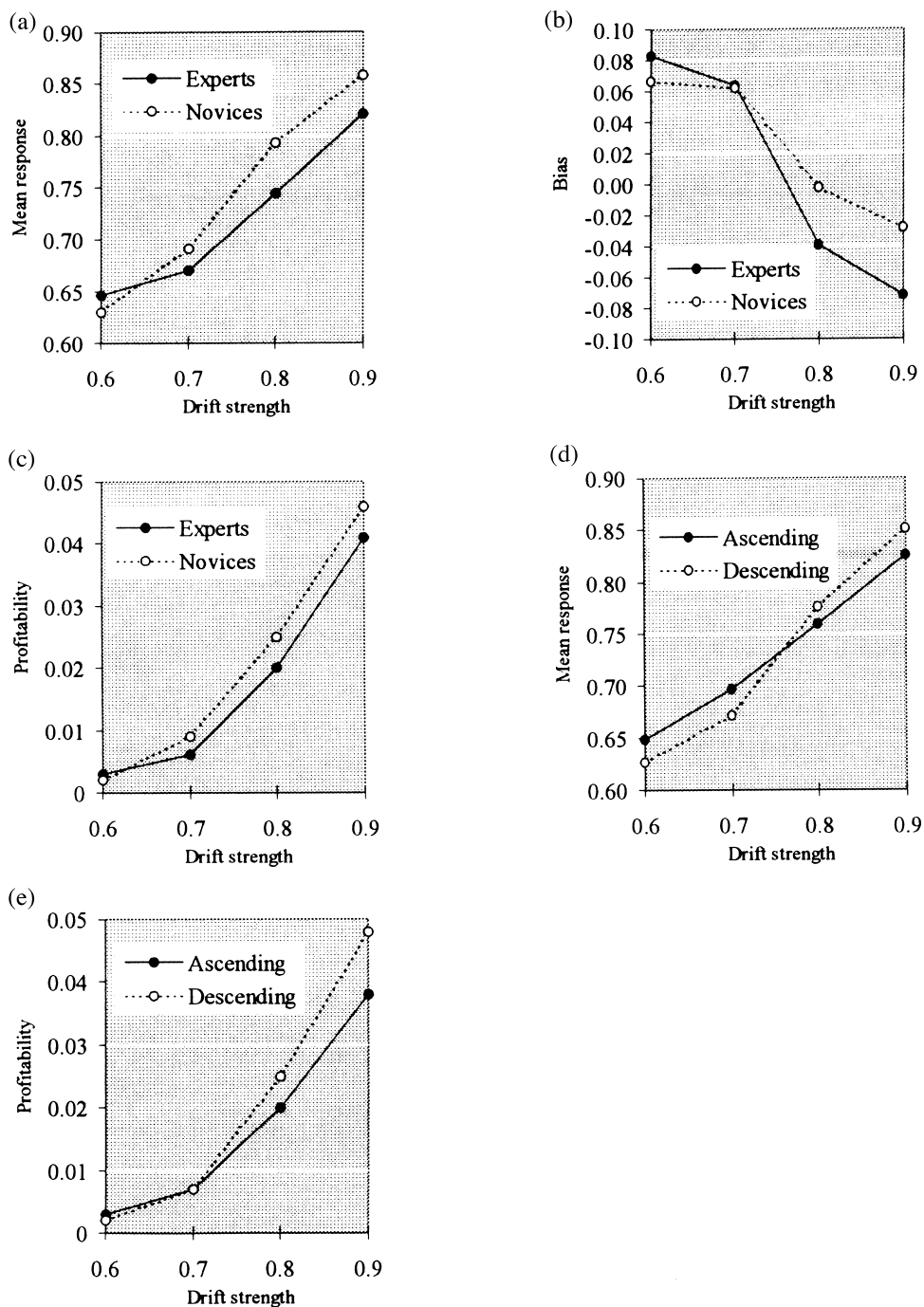


Fig. 1. Interaction plots of mean performance score against trend strength.

$\{F(3,102) = 16.58, P < 0.001\}$. These interactions are illustrated in Fig. 1d and e, respectively, and in Table 2. Fig. 1d shows a clear pattern for the subjects' confidence levels in relation to the direction and strength of trend. Participants are less confident (as revealed by lower mean probability responses) about negative trends than about positive trends when trends are weak (i.e., 0.6 and 0.7 trends), but the reverse is true when trends are strong (i.e., 0.8 and 0.9 trends). In terms of profitability, Fig. 1e shows that, although performance is essentially similar with upward and downward sloping series on the weak (i.e., more difficult) trends, profitability scores improve more for negatively-trended series for the strong (i.e., less difficult) trends, with the largest difference occurring on the strongest (i.e., least difficult) trend.

These analyses were further complemented by comparing the groups' mean scores to those of the uniform or random walk forecaster (RW), the random walk with constant drift (RWCD) and autoregressive order one (AR(1)) models, and human composite forecasts. These comparisons are illustrated in Table 3. Comparing the relative performance of judgemental forecasts with those of the hypothetical random walk, Table 3 shows that the subjects performed better on $M(c)$ over all trends and on U_{RMPS} on the three strongest trends (i.e., drift strength of 0.7, 0.8 or 0.9). These results compare favourably to those of previous probabilistic studies in the stock price forecasting domain (e.g., Stael von Holstein, 1972; Yates, McDaniel & Brown, 1991; Önköl & Muradoglu, 1994) whose subjects have tended to perform worse than a random walk. However, the subjects in the present study were generally outperformed by the more sophisticated models (i.e., RWCD and AR(1)) on the various measures over the four trends.

As for composite versus individual judgement, in terms of overall performance, Table 3 shows that there was generally a slight improvement in the composite cases for $M(c)$ and a dramatic improvement for U_{RMPS} . These findings are similar to those reported in Thomson, Pollock, Henriksen and Macaulay (in press).

However, in terms of B , although there was considerable improvement in the composite cases on the weakest trend, performance was generally better

for the individual cases on the stronger trends. Combining individual judgements on the stronger trends, therefore, failed to improve the bias score.

Instances where human composite forecasts surpassed statistical models are worth noting. In terms of U_{RMPS} , on the weakest trend, the novice composite, overall, achieved considerably better scores than both models. On the same measure, the novice composite, overall, surpassed the AR(1) model on the 0.8 trend. In relation to B , the two composites outperformed both models on the weakest trend, and the novice composites surpassed both models on the 0.8 trend.

In relation to profitability, however, as the composite forecasts will always give identical profitability for each forecast series as the group mean, there is no direct benefit in combining forecasts from a profitability viewpoint.

5. Conclusions and directions for future research

As stated by O'Connor et al. (1997), "there appears to be very little known about how people behave with trended time series" (p. 166). The current study has provided an exploratory step in investigating this important issue. Currency forecasting is chosen as an exemplar domain where judgemental forecasts prevail, where their accuracy is of utmost importance, and where differences in trend strength are commonly confronted. Using directional probability forecasts, we found that the strength of trend has an important influence on all aspects of forecasting performance studied. In particular, predictive performance has been shown to improve with increasing trend strength. This finding may be explained in conjunction with the hard–easy effect, such that increasing drift strengths may be viewed as making the forecasting task easier. Accordingly, overconfidence is observed with the more difficult trends (i.e., trends that are harder to detect, with drift strengths of 0.6 and 0.7), while underconfidence is displayed on the comparatively easier trends (i.e., strong trends that are relatively easier to discern, with drift strengths of 0.8 and 0.9). In fact, the probability profiles of both the experts and the novices provide support for the Lichtenstein and

Table 3

Performance scores for expertise, trend strength and direction: performance benchmarks and composite forecasts

	Performance measures				
	$M(c)(\uparrow)$	$U_{RMPs}(\downarrow)$	$M(r)$	$B(0)$	$M(\pi)(\uparrow)$
<i>Trend strength = 0.6</i>					
Experts — Pos.	0.570	1.261	0.662	0.091	0.003
Experts — Neg.	0.556	1.267	0.630	0.075	0.002
Novices — Pos.	0.559	1.230	0.636	0.077	0.002
Novices — Neg.	0.567	1.056	0.624	0.057	0.002
Uniform	0.500	1.000	0.500	0.000	0.000
RWCD	0.600	0.494	0.574	−0.026	0.002
AR(1)	0.600	0.488	0.560	−0.040	0.002
Comp. Exp. — Pos.	0.600	0.281	0.607	0.007	0.003
Comp. Exp. — Neg.	0.600	0.549	0.585	−0.015	0.002
Comp. Nov. — Pos.	0.600	0.371	0.604	0.004	0.002
Comp. Nov. — Neg.	0.600	0.305	0.599	−0.001	0.002
<i>Trend strength = 0.7</i>					
Experts — Pos.	0.619	0.835	0.666	0.047	0.006
Experts — Neg.	0.596	0.796	0.676	0.080	0.006
Novices — Pos.	0.526	0.834	0.716	0.090	0.009
Novices — Neg.	0.533	0.588	0.668	0.034	0.008
Uniform	0.500	1.000	0.500	0.000	0.000
RWCD	0.700	0.129	0.698	−0.002	0.010
AR(1)	0.700	0.077	0.697	−0.003	0.010
Comp. Exp. — Pos.	0.700	0.730	0.587	0.020	0.006
Comp. Exp. — Neg.	0.567	0.426	0.616	−0.084	0.006
Comp. Nov. — Pos.	0.700	0.687	0.672	0.105	0.009
Comp. Nov. — Neg.	0.700	0.332	0.650	−0.050	0.008
<i>Trend strength = 0.8</i>					
Experts — Pos.	0.767	0.506	0.730	−0.036	0.017
Experts — Neg.	0.800	0.361	0.759	−0.041	0.023
Novices — Pos.	0.769	0.322	0.798	0.001	0.023
Novices — Neg.	0.800	0.323	0.796	−0.004	0.026
Uniform	0.500	1.000	0.500	0.000	0.000
RWCD	0.800	0.091	0.782	−0.018	0.024
AR(1)	0.800	0.129	0.782	−0.018	0.024
Comp. Exp. — Pos.	0.800	0.327	0.713	−0.087	0.017
Comp. Exp. — Neg.	0.800	0.160	0.760	−0.040	0.023
Comp. Nov. — Pos.	0.800	0.102	0.793	−0.007	0.023
Comp. Nov. — Neg.	0.800	0.121	0.809	0.009	0.027
<i>Trend strength = 0.9</i>					
Experts — Pos.	0.885	0.361	0.083	−0.072	0.036
Experts — Neg.	0.900	0.281	0.828	−0.072	0.046
Novices — Pos.	0.885	0.276	0.840	−0.045	0.040
Novices — Neg.	0.885	0.221	0.875	−0.010	0.051
Uniform	0.500	1.000	0.500	0.000	0.000
RWCD	0.900	0.800	0.878	−0.022	0.049
AR(1)	0.900	0.085	0.878	−0.022	0.049
Comp. Exp. — Pos.	0.900	0.254	0.808	−0.092	0.036
Comp. Exp. — Neg.	0.900	0.192	0.824	−0.076	0.046
Comp. Nov. — Pos.	0.900	0.200	0.827	−0.073	0.040
Comp. Nov. — Neg.	0.900	0.068	0.874	0.026	0.051

Note: \uparrow high values are best; \downarrow low values are best; 0 zero is the best value.

Fischhoff's (1977) and Suantak, Bolger and Ferrell (1996) findings in that overconfidence is reduced as mean weighted outcome index is increased from 50% to 75%, with underconfidence emerging when this index exceeds the 75% level.

In general, the subjects were more overconfident on the weaker (more difficult) trends than they were underconfident on the stronger (easier) trends, and this resulted in a low level of overconfidence overall. This finding disagrees with Bolger and Harvey's

(1995) supposition that directional probabilistic forecasting tasks would produce underconfidence, rather than the overconfidence that is normally observed in confidence interval tasks (e.g., Lawrence & Makridakis, 1989; Lawrence & O'Connor, 1993; O'Connor & Lawrence, 1989, 1992). Bolger and Harvey hypothesised that this would be the case since subjects in the directional situation are likely to anchor and adjust (Tversky & Kahneman, 1974) from the centre of the probability scale (Poulton, 1989, 1994), and insufficient adjustment away from this anchor would lead to hypoprecision or underconfidence. Conversely, when fractile judgements are required, the subject's best single estimate tends to be used as an anchor, and adjusting away from this point is likely to produce hyperprecision (Pitz, 1974), resulting in the overconfidence that is usually observed. Given that the participants in the present study attained a low level of overconfidence with their directional forecasts, current findings appear to contradict Bolger and Harvey's assumption while supporting Seaver von Winterfeldt and Edwards' (1978) results that their subjects' judgements were "not too flat; they were about right, though not quite flat enough" (p. 384). The discrepancy between these findings and those of Bolger and Harvey could be attributed to contextual factors, such as the labelling of time series. For instance, unlike the 'sales' label employed in the Bolger and Harvey study, our experts and novices were predicting under the 'currency' label. It has been suggested that the subjects' expectations about the behaviour of the series may be affected by the particular labels used (Goodwin & Wright, 1994), which may in turn alter the participants' reactions to trend. Future work to systematically delineate the effects of contextual information on predictive accuracy in domains where differing trend strengths predominate remains vital for the users and the providers of such forecasts.

Regarding the effects of expertise, this study has found the experts' forecasts to yield relatively low accuracy and profitability scores in comparison to the novices. The novice group has also been found to use higher probabilities on average than the expert group. These results, complemented by significant expertise by trend-strength interactions, may be viewed as reflecting the experts' resistance to accept strong trends. Firmly grounded on random walk

theory, their academic base may have induced these experts to search for randomness to correct any 'misperceptions' of strong trends. This supports van Hoek's (1992) proposition that "... analysts appear to expect some reversal in recent exchange rate movements or a return to some long-run 'normal' value" (p. 467). Frankel and Froot (1990) and Tagaki (1991) also report similar findings. Hence, an intriguing extension of current work may entail investigating the experts' reactions to detailed task and/or performance feedback (Benson & Önköl, 1992; Bolger & Wright, 1993, 1994; Muradoglu & Önköl, 1994; Wright, Rowe, Bolger and Gammack, 1994; Önköl & Muradoglu, 1994, 1995, 1996; Önköl-Atay, 1998).

Another interesting result is given by the influence of trend direction on relative accuracy and profitability measures. The inferior performance on positive trends reported in the present study supports the findings of Bolger and Harvey (1995) and Timmers and Wagenaar (1977), while disagreeing with the Lawrence and Makridakis (1989) results on prediction intervals. Why is this the case? Basic task differences provide one potential answer. Task difficulty may offer another viable explanation. In the present study, participants' forecasts reflected a sensitivity to differing growth rates, with the greatest increasing-trend disadvantage occurring on the strongest trend. Focusing only on upward-sloping trends, Eggleton (1982) noted that the underestimation bias seemed to increase more than proportionately as growth rates increase. When growth is exponential, this underestimation bias on increasing trends appears to be substantial (Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1979). However, predictions from exponentially-decaying series have been found to be much closer to those prescribed by the mathematical relationships (Timmers & Wagenaar, 1977). Therefore, the strength and the direction of the slope appear to be important determinants of performance, as highlighted by the significant interactions of trend strength with trend direction in the current study. Further work detailing these potential effects remains crucial for designing forecast support systems to enhance predictive accuracy under differing trend conditions for diverse application domains.

Other important findings to emerge from the

present study include the observed advantage of composite forecasts over individual forecasts and contexts enabling composite human judgement to surpass the performance of the statistical models. It is also worth noting that combining individual judgements, on stronger trends, failed to reduce bias. Composite forecasts and persistence of bias represent future research venues of interest.

The accuracy of exchange rate forecasts is critical for both the users and the producers of financial information. Accordingly, profiling the effects of factors like trend strength that effectively alter predictive performance promises an important research area worth pursuing.

Appendix A. Subjects' instruction sheet

Instructions for making the forecasts

For each series we would like you to indicate, with a tick, the direction of movement of the series in month 61 (1 month ahead forecast) relative to month 60.

Then we would like you to indicate how confident you are in your choice by writing down a percentage of 50% to 100%. A value of 50% would mean that you are equally likely to be right or wrong — that is, your answer is completely a guess. A value of, say 60%, would indicate a greater degree of confidence in your response. A value of 100% indicates complete certainty that the series will move either up or down. You should not, however, use values of less than 50% as this would indicate that you are less likely to be right than wrong, in which case the alternative direction should be indicated as more likely.

Three examples of how a hypothetical participant might respond are indicated below:

In which direction do you expect the series to move?			Probability of being correct (%)
Series X	(a) rise	✓	90%
	(b) fall	—	
Series Y	(a) rise	—	70%
	(b) fall	✓	
Series Z	(a) rise	✓	50%
	(b) fall	—	

Appendix B. Model details

Each of the performance measures was analysed as a single factor (expertise) fixed effects design using 4 (strength of trend) by 2 (direction of trend) repeated measures. Specifically, the model assumed was

$$Y_{ijkm} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \delta_m + (\alpha\delta)_{jm} + (\beta\delta)_{km} + (\alpha\beta\delta)_{jkm} + \gamma_{i(j)} + (\beta\gamma)_{ki(j)} + (\delta\gamma)_{mi(j)} + \varepsilon_{(ijkm)}$$

where Y_{ijkm} is the mean measured score (i.e., averaged over the 24 series) of the i th subject, expertise level j , trend k , direction m ; α_j is the 'expertise' effect, $j = 1$ for *experts*, $j = 2$ for *novices*; β_k is the 'trend' effect, $k = 1, 2, 3, 4$ for mild, medium, strong and very strong drift, respectively; δ_m is the 'direction' effect, $m = 1, 2$ for upward trends (positive drift) and downward trends (negative drift), respectively; $(\alpha\beta)_{jk}$ is the expertise*trend interaction; $(\alpha\delta)_{jm}$ is the expertise*direction interaction; $(\beta\delta)_{km}$ is the trend*direction interaction; $(\alpha\beta\delta)_{jkm}$ is the expertise*trend*direction interaction; $\gamma_{i(j)}$ is the 'subject' effect (subjects nested within expertise), $i = 1, 2, 3, \dots, 18$; $(\beta\gamma)_{ki(j)}$ is the trend*subject interaction; $(\delta\gamma)_{mi(j)}$ is the direction*subject interaction; μ is a constant (overall mean); α_j 's are constants such that $\sum \alpha_j = 0$; β_k 's are constants such that $\sum \beta_k = 0$; δ_m 's are constants such that $\sum \delta_m = 0$; $(\alpha\beta)_{jk}$'s are constants such that $\sum \sum (\alpha\beta)_{jk} = 0$; $(\alpha\delta)_{jm}$'s are constants such that $\sum \sum (\alpha\delta)_{jm} = 0$; $(\beta\delta)_{km}$'s are constants such that $\sum \sum (\beta\delta)_{km} = 0$; $(\alpha\beta\delta)_{jkm}$'s are constants such that $\sum \sum \sum (\alpha\beta\delta)_{jkm} = 0$; $\gamma_{i(j)}$'s are constants such that $\sum \sum \gamma_{i(j)} = 0$; $(\beta\gamma)_{ki(j)}$'s are constants such that $\sum \sum \sum (\beta\gamma)_{ki(j)} = 0$; and $\varepsilon_{(ijkm)} \sim N(0, \sigma^2)$.

References

- Ang, S., & O'Connor, M. (1991). The effect of group interaction processes on performance in time series extrapolation. *International Journal of Forecasting*, 7, 141–149.
- Benson, P. G., & Önköl, D. (1992). The effects of feedback and training on the performance of probability forecasters. *International Journal of Forecasting*, 8, 559–573.
- Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *The Quarterly Journal of Experimental Psychology*, 46A, 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., & Harvey, N. (1996). Graphs versus tables: effects of data presentation format on judgmental forecasting. *International Journal of Forecasting*, 12, 119–137.
- Bolger, F., & Wright, G. (1993). Coherence and calibration in expert probability judgment. *Omega*, 21, 629–644.
- Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: 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.
- 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: Chicago Press.
- Eggleson, I. R. C. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research*, 20, 68–102.
- Frankel, J., & Froot, K. (1990). The rationality of the foreign exchange rate: chartists, fundamentalists, and trading in the foreign exchange market. *AEA Papers and Proceedings*, 80, 181–185.
- 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. *Omega*, 22, 553–568.
- Harvey, N. (1995). Why are judgments 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, 119–137.
- Lawrence, M., & Makridakis, S. (1989). Factors affecting judgemental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, 42, 172–187.
- Lawrence, M., & O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, 8, 15–26.
- Lawrence, M., & O'Connor, M. (1993). Scale, variability and the calibration of judgmental prediction intervals. *Organizational Behavior and Human Decision Processes*, 56, 441–458.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know. *Organizational Behavior and Human Performance*, 20, 159–183.
- Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: its effectiveness and biases. *Journal of Behavioral Decision Making*, 8, 149–168.
- Muradoglu, G., & Önkale, 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. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association*, 79, 489–500.
- Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, 8, 139–162.
- 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, 165–176.
- Officer, L. H. (1982). *Purchasing power parity and exchange rates: theory, evidence and relevance*, Contemporary Studies in Economic and Financial Analysis, Vol. 35. London: JAI.
- Önkale, D., & Muradoglu, G. (1994). Evaluating probabilistic forecasts of stock prices in a developing stock market. *European Journal of Operational Research*, 74, 350–358.
- Önkale, D., & Muradoglu, G. (1995). Effects of feedback on probabilistic forecasts of stock prices. *International Journal of Forecasting*, 11, 307–319.
- Önkale, D., & Muradoglu, G. (1996). Effects of task format on probabilistic forecasting of stock prices. *International Journal of Forecasting*, 12, 9–24.
- Önkale-Atay, D. (1998). Financial forecasting with judgment. In Wright, G., & Goodwin, P. (Eds.), *Forecasting with Judgment*. Chichester: Wiley, pp. 139–167.
- Önkale-Atay, D., Wilkie-Thomson, M. E., & Pollock, A. C. Judgemental forecasting. In: Clements, M. P., & Hendry, D. (Eds.), *Companion to Economic Forecasting*, Blackwell, in press.
- Pitz, G. F. (1974). Subjective probability distributions for imperfectly known quantities. In Gregg, L. W. (Ed.), *Knowledge and Cognition*. New York: Wiley.
- 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., Macaulay, A., Önkale-Atay, D., & Wilkie-Thomson, M. E. (1999). Evaluating predictive performance of judgmental extrapolations from simulated currency series. *European Journal of Operational Research*, 114, 281–293.
- Pollock, A. C., & Wilkie, M. E. (1992). Currency forecasting: human judgment or models. *VBA-Journal*, 3, 21–29.
- Pollock, A. C., & Wilkie, M. E. (1993). Directional judgemental financial forecasting: trends and random walks. In Flavell, R. (Ed.), *Modelling Reality and Personal Modelling*. Heidelberg: Physica, pp. 253–271.
- 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.
- Poulton, E. C. (1989). *Bias in Quantifying Judgements*. Hove and London: Lawrence Erlbaum Associates.
- Poulton, E. C. (1994). *Behavioural Decision Theory: A New Approach*. Cambridge: Cambridge University Press.
- Remus, W., O'Connor, M., & Griggs, K. (1995). Does reliable information improve the accuracy of judgmental forecasts? *International Journal of Forecasting*, 11, 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.
- Sanders, N. R. (1992). Accuracy of judgmental forecasts: a comparison. *Omega*, 20, 353–364.
- Seaver, D. A., von Winterfeldt, D., & Edwards, W. (1978). Eliciting subjective probability distributions on continuous variables. *Organizational Behavior and Human Performance*, 21, 379–391.
- Stael von Holstein, C. A. S. (1972). Probability forecasting: an experiment related to the stock market. *Organizational Behavior and Human Decision Performance*, 8, 139–158.
- Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The hard–easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes*, 67, 201–221.
- Tagaki, S. (1991). Exchange rate expectations: a survey of survey studies. *IMF Staff Papers*, 38, 156–183.
- Timmers, H., & Wagenaar, W. A. (1977). Inverse statistics and misperception of exponential growth. *Perception and Psychophysics*, 21, 558–562.
- Thomson, M. E., Pollock, A. C., Henriksen, K. B., & Macaulay, A. The influence of the forecast horizon on the currency predictions of experts, novices and statistical models. *European Journal of Finance*.
- Tvede, L. (1990). *The Psychology of Finance*. Norway: Norwegian University Press.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science*, 1127–1131.
- van Hoek, T. H. (1992). Explaining mark/dollar and yen/dollar exchange rates in the 1980s. *Economics Letters*, 38, 467–472.
- Wagenaar, W. A., & Sagaria, S. D. (1975). Misconceptions of exponential growth. *Perception and Psychophysics*, 18, 416–422.
- 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. (1996). Judgemental and statistical time series forecasting: a review of the literature. *International Journal of Forecasting*, 12, 91–118.
- Wilkie-Thomson, M. E., Önköl-Atay, D., & Pollock, A. C. (1997). Currency forecasting: an investigation of extrapolative judgement. *International Journal of Forecasting*, 13, 509–526.
- Wilkie, M. E., & Pollock, A. C. (1994). Directional currency forecasting: an investigation into probability judgement accuracy. In Peccati, L., & Virén, M. (Eds.), *Financial Modelling*. Heidelberg: Physica, pp. 354–364.
- 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., Rowe, G., Bolger, F., & Gammack, J. (1994). Coherence, calibration, and expertise in judgmental probability forecasting. *Organizational Behavior and Human Decision Processes*, 57, 1–25.
- Yates, J. F. (1982). External correspondence: decompositions of the mean probability score. *Organizational Behavior and Human Performance*, 30, 132–156.
- 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 Performance*, 49, 60–79.

Biographies: Mary E. THOMSON is a Reader in the Division of Risk, Glasgow Caledonian Business School. She completed a Ph.D. on judgement in currency forecasting and has published articles and papers in a variety of books and journals in this area. 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 and Associate Dean of the Faculty of Business Administration 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 Behavioral 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.

Alex MACAULAY is Senior Lecturer in Statistics at Glasgow Caledonian University. He has a B.Sc. in Mathematics and an M.Sc. in Statistics (Stochastic Processes). He has a particular research interest in the forecasting of financial price series and in the evaluation of predictive performance.