



Effects of trend strength and direction on performance and consistency in judgmental exchange rate forecasting

Mary E. Thomson^{a,*}, Andrew C. Pollock^{b,1}, M. Sinan Gönül^{c,2}, Dilek Önkal^{d,3}

^a Department of Psychology and Allied Health Sciences, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, Scotland, UK

^b Division of Computer, Communications and Interactive Systems, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, Scotland, UK

^c Department of Business Administration, Middle East Technical University, Üniversiteler Mah. Dumlupınar Blv. No:1, 06800, Çankaya, Ankara, Turkey

^d Faculty of Business Administration, Bilkent University, 06800, Ankara, Turkey

ARTICLE INFO

Keywords:

Judgmental forecasting
Probability forecasting
Trend strength
Trend direction
Consistency
Damping
Exchange rate

ABSTRACT

Using real financial data, this study examines the influence of trend direction and strength on judgmental exchange rate forecasting performance and consistency. Participants generated forecasts for each of 20 series. Half of the participants also answered two additional questions regarding their perceptions about the strength and direction of the trend present in each of the series under consideration. The performance on ascending trends was found to be superior to that on descending trends, and the performance on intermediate trends was found to be superior to that on strong trends. Furthermore, the group whose attention was drawn to the direction and strength of each trend via the additional questions performed better on some aspects of the task than did their “no-additional questions” counterparts. Consistency was generally poor, with ascending trends being perceived as being stronger than descending trends. The results are discussed in terms of their implications for the use and design of forecasting support systems.

© 2012 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

One of the most important factors to be taken into account in the design of forecasting support systems is the persistence of human judgement in the forecasting process. This point has long been recognized by forecasting practitioners, who continue to rely heavily on judgment. For instance, in a survey of 240 US companies, only 11% claimed to use quantitative forecasting methods, and 60% of these firms stated that they regularly used their

judgment to adjust the statistical forecasts (Sanders & Manrodt, 2003). In a more recent study analysing more than 60,000 forecasts gathered from four large supply chain companies, the percentage of judgmentally adjusted forecasts was high, soaring to 91% in one company (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009). In exchange rate forecasting, which is the focus of the present study, professionals who practice “chartist” techniques (a popular branch of technical analysis) rely entirely on human judgment (Allen & Taylor, 1990; Murphy, 1999).

However, academics' acceptance of the value of human judgment in this context is a relatively recent occurrence (Lawrence, Goodwin, O'Connor, & Önkal, 2006). Initially, academics advocated the exclusive use of statistical methods in the forecasting process, and hence in the development of forecasting support systems (FSSs), but this view gradually changed, once comparative studies demonstrated that judgmental forecasts were at least as good as their statistical counterparts. For instance, Armstrong (1985) reviewed earnings forecasts and found

* Corresponding author. Tel.: +44 0 141 331 3855; fax: +44 0 141 331 3005.

E-mail addresses: M.Thomson@gcal.ac.uk (M.E. Thomson), a.c.pollock@gcal.ac.uk (A.C. Pollock), msonul@metu.edu.tr (M.S. Gönül), onkal@bilkent.edu.tr (D. Önkal).

¹ Tel.: +44 0 141 331 3855; fax: +44 0 141 331 3005.

² Tel.: +90 312 210 2034; fax: +90 312 210 7962.

³ Tel.: +90 312 290 1251; fax: +90 312 266 4960.

that judgmental forecasts generally surpassed statistical forecasts. Various studies restricting forecasters to only 'blind' time series data have demonstrated that such findings cannot simply be explained by the fact that judgmental forecasters usually have access to additional information that is not incorporated in the statistical models (e.g., Lawrence, Edmundson, & O'Connor, 1985; Wilkie-Thomson, Önkâl-Atay, & Pollock, 1997). Despite the current acknowledgement of the value of judgment in forecasting, it has also been recognized that various biases tend to occur in these contexts. Therefore, owing to the enduring prevalence of judgmental forecasting in the real world, it is essential to enhance our understanding of the advantages and disadvantages of human judgment in forecasting, so that remedies for the latter may be incorporated into the design of FSSs.

A well-established judgmental bias in this context is the tendency of forecasters to dampen both ascending and descending trends (e.g., Andreassen, 1990; Andreassen & Kraus, 1990; Eggleton, 1982; Keren, 1983; Lawrence & Makridakis, 1989; O'Connor, Remus, & Griggs, 1997). That is, their forecasts are usually situated below ascending trends and above descending trends. The main reason put forward to account for such damping is that the very act of asking people to predict future prices may induce mean-reverting expectations (Andreassen, 1988; Glaser, Langer, Reynders, & Weber, 2007a). As was suggested by Andreassen (1988, p. 373): "when a price rises, some people may attend to the fact that today's price is higher than average, and so think it is likely to fall...". However, the damping bias tends to be more severe with descending trends (Harvey & Bolger, 1996; Lawrence & Makridakis, 1989; O'Connor et al., 1997). Various explanations have been put forward to account for this ascending advantage. For instance, Reimers and Harvey (2011) suggest that it might be due to an optimism bias (Weinstein, 1989), given that forecasters are mostly predicting quantities for which higher values are more desirable than lower values (e.g., sales or profits). Another explanation for the ascending trend advantage is that the labels used in such studies (e.g., sales) may stimulate an assumption on the part of the forecasters that management will eventually intervene to prevent the values from falling further (e.g., Lawrence & Makridakis, 1989).

The quality of judgmental forecasts is also affected by the strength of the trend (Andreassen, 1988; Eggleton, 1982; Lawrence & Makridakis, 1989). For instance, the damping bias has been found to be particularly strong whenever forecasts are extrapolated from ascending deterministic exponential functions. In fact, the observed gross underestimation of exponential growth was not reduced by a knowledge of the growth processes (Wagenaar & Sagaria, 1975), the provision of more data points (Wagenaar & Timmers, 1978), or the presentation of the data as they became available over time (Wagenaar & Timmers, 1979). Interestingly, predictions from exponentially-descending series have been found to be much less biased (Timmers & Wagenaar, 1977). Paradoxically, therefore, these studies suggest that the performance is less likely to be biased on ascending trends, but not if the trends are particularly steep.

The influence of signal strength on judgment has also been demonstrated in non-forecasting contexts. For instance, in relation to the determinants of confidence in judgment, Griffin and Tversky (1992) found that people are overconfident when signals are salient and strong but underconfident when signals are weak. This strength and weight account of the "hard-easy" effect (i.e., where underconfidence is demonstrated with easy tasks and overconfidence is shown with difficult tasks) (e.g., Lichtenstein, Fischhoff, & Phillips, 1982) has subsequently prompted financial economists to develop models to account for both market over- and under-reaction (e.g., Barberis, Shleifer, & Vishny, 1998), which explain belief revisions in a Bayesian manner. However, more recent research has suggested that such revisions may be able to be described more adequately as quasi-Bayesian. Specifically, Massey and Wu (2005) have proposed a "system neglect" hypothesis, whereby people typically overweight strong and salient signals because they are in the foreground but underweight important system parameters that are in the background, such as instability.

Along similar lines, Thomson, Önkâl-Atay, Pollock, and Macaulay (2003) reported thought-provoking results in relation to participants' confidence levels in an exchange rate forecasting study. In this study, system parameters such as noise and stability levels were held constant, but the direction and strength of the trend were controlled simultaneously. An almost mirror image was found, with participants being less confident about descending trends than ascending trends when the trends were weak, but with the reverse being the case when the trends were strong. Therefore, the direction and strength of the trend were suggested as clearly being important factors to consider when forecasting trends from time series information.

It is also important to note that most time series studies have been based on constructed data. It is argued that such data are useful in examining the quality of judgment in this context because they allow the researcher to control important time series characteristics while preventing non-time series information from impacting upon the performance (Goodwin & Wright, 1993; O'Connor & Lawrence, 1989). Nevertheless, there have been queries about the extent to which the results of such studies generalize to forecasting performance in the real world. Indeed, one study which used actual data from the famous M-Competition (Makridakis et al., 1982) failed to find evidence of damping behaviour (Lawrence & O'Connor, 1995). However, according to Harvey (2011, personal communication), since this study utilized data from a whole set of series with differing trend strengths, one might expect there to have been damping for series that were more steeply trended than the average in the set, and 'anti-damping' for those series that were less steeply trended than the average of the series in the set. These two opposite effects might, therefore, have cancelled out to produce the overall results. In addition to this, evidence that real life trends tend to be damped to some extent, and that people have adjusted to this, comes from the finding that the statistical forecasting performance can be enhanced by adding damping terms (e.g., Collopy & Armstrong, 1992; Gardner & McKenzie, 1985).

On the other hand, using real stock price data, *Budescu and Du (2007)* found only slight overconfidence in one study and only modest underconfidence in another. These authors warned against accepting the standard view (e.g., *Barber & Odean, 2000; De Bondt & Thaler, 1995, Chapter 3*) that financial forecasters are overconfident, and questioned whether FSSs should assume overconfidence when predicting behaviour. Conversely, many studies which have used actual stock price data have demonstrated overconfidence (e.g., *Bartos, 1969; Önköl & Muradoglu, 1994; Önköl & Muradoglu, 1996; Önköl, Yates, Şimga-Muşan, & Öztin, 2003; Stael Von Holstein, 1972; Whitecotton, 1996; Yates, McDaniel, & Brown, 1991*). In a more recent study, *Glaser, Langer, and Weber (2007b)* found overconfidence in their participants' underestimation of the volatility of stock returns. In line with studies using constructed data (e.g., *Harvey & Bolger, 1996; Lawrence & Makridakis, 1989; O'Connor et al., 1997*), these authors found that damping was greater for descending trends. They also found that damping was greater when the participants were asked to predict prices than when they were required to predict returns. Therefore, in addition to relevant time series characteristics, task framing is also likely to affect forecasting behaviour.

Taken together, the evidence discussed so far is rather mixed. On the one hand, a tendency for forecasters to dampen both ascending and descending trends is well established. However, this evidence has tended to come from studies which have used constructed data (with the notable exception of *Glaser et al., 2007b*), thus leading to questions about the ecological validity of such findings. Studies using actual series have also produced inconsistent findings. Furthermore, the observed bias seems to depend on both the context and the task. Therefore, in terms of the design of FSSs, more studies using actual data from real-world contexts are needed to establish the extent to which biases such as those discussed above actually occur, and the particular circumstances in which they are most prominent. The evidence discussed above suggests that such studies should attempt to control important time series characteristics systematically, in order to determine their influence on forecasting behaviour. Accordingly, one aim of the present study is to examine the forecasting performance in relation to the direction and strength of the trend, in a task which uses actual exchange rate data. This particular context is an important application domain, given that judgment is frequently used as the sole method for forecasting price series (e.g., by "chartists"), thus enhancing the ecological validity of the "abstract" forecasting task (i.e., forecasting time series without additional contextual information) which is used in the present study. However, another reason for the selection of the exchange rate forecasting context was the fact that it provides an excellent situation in which to examine performance in relation to the direction of trends, given that exchange rate pairs are completely invertible (in a way that stock prices, for example, clearly are not). That is, if an exchange rate price series graph of the USD/GBP exhibits a descending trend, then a similar graph displaying the GBP/USD rate would exhibit an ascending trend of a similar pattern. This invertible context also provides an

ideal opportunity for examining directional consistency in forecasting performance, another aim of the present study which is discussed later.

In the meantime, a second aim of the present study was to determine whether the act of merely drawing participants' attention to the direction and strength of trend makes a difference to their forecasts. Although it is not normally considered in time series studies, such information is vital to the design of FSSs. One way of directing participants' attention was to explicitly ask them to provide ratings on the perceived trend direction and strength. Through answering such questions, the characteristics related to the trend could become more salient in their minds, and hence, they might focus more on these two features when generating predictions. In this way, the damping behaviour might be accentuated relative to the case where there is no explicit highlighting of the trend characteristics. On the other hand, if the decision makers are already incorporating information on the trend direction and strength in their mental forecast generation process, answering such questions would make no difference to the predictions produced.

A final aim of the study was to examine forecasting consistency. The concept of consistency, as used in this study, relates to the relationships between the probability forecasts of paired exchange rate series that are identical in every respect except for the direction of the trend. For instance, if a perfectly consistent forecaster predicted a 65% probability of a rise in the exchange rate when the series was displayed in an ascending fashion, then he would predict a 65% probability of a fall when the same series was exhibited in a descending manner. In the present study, a consistency analysis in this form is applied to the paired series in relation to trend direction and strength of the series, and to the condition as to whether or not the attention of the participants was drawn to these characteristics. Despite its obvious importance, consistency has rarely been examined in time series forecasting contexts, and when it has, the focus has been on comparing the consistency between probability estimates and forecasts of future values (e.g., *Budescu & Du, 2007; Glaser et al., 2007b*). To the best of our knowledge, unlike other aspects of performance, consistency between related paired series has never previously been examined in relation to specific time series characteristics, such as trend direction. As was pointed out above, the invertible nature of exchange rate series enabled this to be done in the present study. Accordingly, statistical measures of consistency, which are described in detail later in the methodology section, were developed for this purpose. In order to proceed along these lines, however, it is first necessary to describe the exchange rate series which were used in the present study.

2. Methodology

The exchange rate between two currencies is the price at which one currency is exchanged for another. Exchange rates are determined by the interaction of demand and supply on the global foreign exchange markets, which had a combined turnover of almost 4 trillion USD in

2010 (Bank for International Settlements (BIS), 2010). High frequency movements in exchange rates frequently occur at intervals of less than a second through the activities of currency trading, largely by major financial institutions. Weekly exchange rate movements reflect the sum of these high frequency changes over the entire week. Despite the complexity and scale of the foreign exchange market, graphical information is used by *chartists* (a branch of *technical analysis*) and plays a fundamental role in the forecasting of exchange rate movements. *Technical analysts* consider that the ‘market discounts everything’, such that all information is quickly incorporated in the price, and hence contextual information is of little use (Murphy, 1999). Survey data has supported this widespread use of *technical analysis* by foreign exchange market participants (Allen & Taylor, 1990; Cheung & Chinn, 2001). In practice, foreign exchange market participants undertake activities to maximise their returns, but base their buy and sell action decisions on signals from the actual series. In this study, the series are presented to participants in a modified version of their original form. This gives substantive ecological validity to the methodology used in this study, as set out in this section.

2.1. The data

The data for the study were obtained from the *Bank of England, Statistical Interactive Database* (from the website: <http://www.bankofengland.co.uk>). Specifically, the *Interest and Exchange Rate, Spot Exchange Rate* section was used. Daily exchange rates were obtained for seven exchange rates against the USD (namely AUD/USD, CAD/USD, CHF/USD, EUR/USD, GBP/USD, JPY/USD and NZD/USD). The data period extended from 2 January 1975 to 31 December 2009. This gave a total of 9131 days, including 280 days of UK bank holidays but excluding weekends. The Bank of England daily exchange rates indicate middle market (the average of spot buying and selling) rates, as observed by the Bank’s Foreign Exchange Desk in the London Interbank Market at around 16.00 h UK time. See Appendix B for further information about these data.

From these USD exchange rate series, related exchange rates were obtained directly for non-USD exchange rates (e.g., GBP/EUR = (GBP/USD)/(EUR/USD)). This gave a total of 28 exchange rate series for all combinations of AUD, CAD, CHF, EUR, GBP, JPY, NZD and USD. From these rates, a further 28 inverted rates (e.g., EUR/GBP from GBP/EUR) were also obtained.

To obtain the weekly data values, only the prices for Friday were used, thus reducing the number of values to 1826. The weekly period, therefore, extended from 16.00 Friday to 16.00 the following Friday (except, of course, when Friday was a holiday, in which case the value at 16.00 on the previous non-holiday weekday was used).

2.1.1. Data transformation

Exchange rate series are not statistically stationary. That is, the variance tends to increase over time, and first order serial correlation occurs with a value close to unity. In other words, the series tend to follow what is

described by Nelson and Plosser (1982) as a difference-stationary process. Trends in exchange rate series are associated with a first order positive serial correlation of unity. Using first differences, however, removes a linear trend, replacing it with constant, usually referred to as *drift*. It also removes serial correlation of unity, replacing it with serial correlation of zero. In addition, as the magnitude of the changes in exchange rates tends to be related to their levels, taking logarithms before differencing usually removes this problem. Defining weekly values of an exchange rate series, X_t , at week t , the logarithms (to base 10) of the series x_t were obtained, i.e. $x_t = \log_{10} X_t$. After this, weekly first differences were taken, Δx_t (where $\Delta x_t = x_t - x_{t-1}$), which is often termed *log returns*. The inverted exchange rate, Y_t , is given by $Y_t = 1/X_t$, and in first differences in log form, $\Delta y_t = -\Delta x_t$.

2.1.2. Selection of series of data used in the task

Trends are generally viewed as being the most important characteristics of actual exchange rate series; and hence, this study selected series that showed intermediate and strong trends. To do this, weekly first differences in logs of the 28 exchange rate series were used to obtain 10 series, denoted j ($j = 1, 2, \dots, 9, 10$), of 30 weeks (plus the associated 10 series of inverted exchange rates). Of these 10 series, five were chosen to represent intermediate drifts (intermediate trends in the non-differenced series) and five to represent strong drifts (strong trends in the non-differenced series). Exponential trends in the actual (non-differenced) series reflect linear trends in the log series. However, given that major exchange rates have been used and the length of the data series has been limited to 30 values, the differences in the visual interpretation of trends would be relatively minimal.

The initial data series obtained were constrained to have a number of characteristics. The criteria used to choose the drift characteristics (associated with intermediate or strong trends) were based on the value of the *empirical probability* (EP), calculated over the 30 weekly values differenced in logs. Empirical probabilities are outlined below and discussed in more detail by Pollock, Macaulay, Thomson, and Önköl (2005, 2008) and Pollock, Macaulay, Thomson, Önköl, and Gönül (2010). EPs use the Student- t distribution with the estimated mean and standard deviation over the 30-week period to give an estimated directional probability. Values of 0.5 indicate no change, values under 0.5 indicate a fall, and values above 0.5 a rise. The procedure for obtaining EPs for each series is discussed below. The EPs are not affected when the series are adjusted to a common standard deviation.

The first stage in obtaining EPs for a series j is to calculate the drift, which can be identified by the mean, m_j , of the weekly changes in logarithms in series j , $x_{j,t}$, as defined in Eq. (1):

$$m_j = \frac{1}{30} \sum_{t=1}^{30} \Delta x_{j,t}. \quad (1)$$

The second stage in obtaining EPs for a series j is to calculate the standard deviation, s_j , of changes in the

logarithms of the data over the same interval, as defined in Eq. (2):

$$s_j = \sqrt{\frac{1}{29} \sum_{t=1}^{30} (\Delta x_{j,t} - m_j)^2}. \tag{2}$$

The next stage in the calculation of EPs is to obtain a t -value. The ratio of the mean (m_j) to the standard deviation (s_j) is multiplied by the square root of 30 to give a quantity (t_j), as defined in Eq. (3):

$$t_j = \sqrt{30} * (m_j/s_j). \tag{3}$$

Since the EPs are obtained on the assumption that the first differences of logarithms are normally distributed, t_j has the Student- t distribution with 29 degrees of freedom. The cumulative probability or EP, e_j , for series j , is calculated in Eq. (4):

$$e_j = F(t_j) = F \left\{ \sqrt{30} * (m_j/s_j) \right\}, \tag{4}$$

where F is the cumulative distribution function of the Student- t distribution with 29 degrees of freedom. The EPs reveal the combined influences of drift and volatility over time.

In this study, the intermediate drift (trends) used in the study had EP values from 0.1 to 0.35 or 0.65 to 0.90 and the strong drift (trends) had values from 0 to below 0.1 or above 0.9 to 1.

For EPs to be valid, it is important to check for non-normality and serial correlation, which can violate the assumptions on which they are based. The validity of the normality assumption for each series was examined using the Anderson–Darling (A–D) test (Anderson & Darling, 1954). To examine the independence of observations, Bartlett’s autocorrelation test for first-order serial correlation was applied (Bartlett, 1946). In situations where either of these tests was significant at the 5% level, the series was excluded.

The 10 series of the differences in logarithms of the exchange rates all had differing variation. As differences in variation were not the focus of the study, the series were adjusted to give all series similar variations with a standard deviation, s_j^* , of 0.004. This was achieved by obtaining an adjusted set of the 30 series, values, $\Delta z_{j,t}$, for $t = 1, \dots, 30$, as follows:

$$\Delta z_{j,t} = (\Delta x_{j,t}/s_j) * 0.004.$$

An adjusted series mean, m_j^* (where $m_j^* = (m_j/s_j) * 0.004$), was also obtained. However, this adjustment did not affect the EP values for each series.

The selected series reflected intermediate and strong trends, and therefore it was considered appropriate to have a log return in week 31 for series j ($\Delta z_{j,31}$) which was very close to the average log return over the previous 30 week period. Specifically, the differences between the actual and mean changes in the series used were less than 0.00002 in absolute terms. That is, other signals present in the selected series were considered to have a neutral effect on the exchange rate value in week 31.

If it is considered that the adjusted values of the exchange rate in week 31 for series j follow a normal

distribution with mean μ_j and standard deviation σ_j , with parameters approximated by m_j^* and s_j^* respectively, this allows an *outcome probability* to be obtained. This outcome probability, v_j , can be defined as $v_j = 1 - \Phi(\mu_j/\sigma_j)$, where $\Phi(\mu_j/\sigma_j)$ denotes the probability of a standard normal random variable less than μ_j/σ_j . This probability will be closer to 0.5 than the empirical probability for the series, as the value uses only one week rather than the 30 weeks to obtain the empirical probability.

The participant’s *directional probability prediction* for week 31 for series j can be compared directly with the outcome probability, v_j , to allow an analysis of the performance. However, this performance analysis would only measure the performance in relation to trend recognition in the appropriate direction. Participants who use other forecasting strategies or look for other non-trend signals would, of course, perform poorly on these criteria. The outcome probability, however, has no direct effect on the consistency analysis.

2.1.3. Presentation of the selected series

To ensure that the presentation of the series was representative of the way in which exchange rate data are often used in practice, it was deemed desirable to present the series in a modified version of their original form, rather than in logged or differenced form. As was pointed out above, foreign exchange market practitioners frequently use *chartist* techniques, or more broadly, *technical analysis* techniques, using the actual exchange rate series, although their activities centre on maximising their (log) returns. The log return series were therefore transformed into actual exchange rate series. Specifically, these adjustments involved:

- (i) A starting value set to zero for week zero ($t = 0$) assigned to each series j .
- (ii) From this starting value of zero, at $t = 0$, cumulative values of the changes were obtained to give resulting values, $q_{j,t}$, for $t = 1, \dots, 30$, where:

$$q_{j,t} = \sum_{i=0}^{t-1} \Delta z_{j,t-i}.$$

- (iii) Anti-logarithms of base 10 were taken to give a series of implied values, $Q_{j,t}$, where:

$$Q_{j,t} = 10^{q_{j,t}}.$$

Note that the starting value for $q_{j,t}$ of zero, at $t = 0$, would result in a starting value for $Q_{j,t}$ of unity, at $t = 0$, for each series j .

The graphical presentations obtained for these implied actual exchange rates, $Q_{j,t}$, for $t = 0$ to 30 for each series j , have the desirable property that they contain the inherited characteristics of the actual rates from which the series was obtained. A further refinement was made to exclude series that showed characteristics that could have distorted judgemental interpretations of the resulting graph of the series. Unexplained variations or patterns in exchange rate data can arise from economic and political events and cause atypical weekly movements that could be misleading when the time series profile of the series is viewed in a judgemental context.

Table 1
The selected series.

Series no.	Exchange rate	Series no. Invert	Exchange rate Invert	Series period		Prediction week w/e	EP			Value $\Delta z_{j,31}$	Mean m_j^*	Diff. $\Delta z_{j,31} - m_j^*$
				Start w/e	End w/e		Fall	Rise				
1	GBP/JPY	14	JPY/GBP	29/07/05	24/02/06	03/03/06	0.3086	0.6914	-0.00037	-0.00038	0.00001	
2	GBP/EUR	13	EUR/GBP	04/11/05	02/06/06	09/06/06	0.3400	0.6600	0.00030	0.00032	-0.00002	
3	CAD/USD	11	USD/CAD	20/04/79	02/11/79	09/11/79	0.1319	0.8681	0.00083	0.00083	0.00000	
4	GBP/CHF	20	CHF/GBP	08/12/78	06/07/79	13/07/79	0.0494	0.9506	-0.00125	-0.00125	0.00000	
5	CHF/USD	19	USD/CHF	13/04/07	09/11/07	16/11/07	0.0515	0.9485	-0.00123	-0.00122	-0.00001	
6	EUR/CHF	16	CHF/EUR	02/02/01	31/08/01	07/09/01	0.2972	0.7028	0.00039	0.00039	0.00000	
7	NZD/CHF	12	CHF/NZD	08/01/82	06/08/82	13/08/82	0.3267	0.6733	-0.00033	-0.00033	0.00000	
8	EUR/CHF	15	CHF/EUR	22/11/02	20/06/03	27/06/03	0.0318	0.9682	-0.00141	-0.00141	0.00000	
9	USD/NZD	18	NZD/USD	08/12/95	05/07/96	12/07/96	0.0585	0.9415	0.00117	0.00118	-0.00001	
10	GBP/NZD	17	NZD/GBP	22/09/95	19/04/96	26/04/96	0.0543	0.9457	0.00121	0.00121	0.00000	

The resulting series were presented on graphs of the same scale with a minimum value of 0.9 and maximum value of 1.1 on the vertical axis for each of the series presented to the participants. There were a total of 20 series used, with the following characteristics:

- (i) 5 intermediate trends, descending;
- (ii) 5 intermediate trends, ascending (i.e., (i) inverted);
- (iii) 5 strong trends, descending;
- (iv) 5 strong trends, ascending (i.e., (iii) inverted).

The series were presented in essentially a random order, although care was taken to avoid related series (in inverted and non-inverted forms) being positioned close together in the ordering. The specific exchange rates, periods and characteristics of the 20 series are set out in Table 1.

2.2. Participants

A total of 92 participants took part in the study, of which 86 submitted 'usable' questionnaires. The participants were final year undergraduate finance students and M.Sc. finance students from Glasgow Caledonian University who had studied currency forecasting as part of their respective courses.

2.3. Procedure

Half of the participants were assigned to Group A (42 usable questionnaires were obtained for this group). Each of these participants was provided with a questionnaire containing 20 time series graphs of exchange rate series, with each graph showing 30 data points. The details of the exchange rates and the time periods used were kept confidential, in order to prevent any potential biases or extraneous information effects. The participants were instructed to study each 30-week time series graph and indicate, for each one, whether they thought that the series would rise or fall in the 31st week, and to provide a percentage probability between 50% and 100% to represent how confident they were that their stated direction was correct. The participants were directed toward considering the trend strength and direction in the instructions. They were then asked to provide a point forecast (i.e., an actual value) for week 31.

The other half of the participants were assigned to Group B (44 usable questionnaires were obtained for this

group). This group followed the same procedure as the first group, but were required to do one additional task for each series. After making the directional and point forecasts, they were required to study the time series graphs under consideration once again and assess the perceived strength and direction of the overall trend by answering two particular questions. This manipulation was designed to test whether or not merely drawing participants' attention to the direction and strength of trends would make a difference to the forecasts produced in relation to these characteristics. Please see Appendix A for the participants' instructions and examples of the currency series graphs.

3. Results for the performance analysis

The first two aims of the present study were to investigate the quality of judgmental forecasting performances in relation to (i) the strength and direction of the trend, and (ii) whether or not drawing participants' attention to the direction and strength of the trend would be sufficient to make a difference to the assessed forecasts. These issues were investigated for both directional and point forecasts.

3.1. Directional forecasts

In terms of directional forecasts, Table 2 displays the percentage of movement directions that were predicted in the same direction as the actual trend for Week 31 by the two groups of participants.

The primary factors that seem to influence the accurate prediction of directional trend movement in the exchange rate series are the strength of the trend ($F_{1252} = 6.76$, $p = 0.01$) and the direction of the trend ($F_{1252} = 23.43$, $p < 0.0001$), as indicated through an ANOVA run on the data. Although, as Table 2 shows, the simple act of drawing participants' attention to these characteristics achieved higher percentages of correct predictions, these differences were not significant, and none of the interaction effects among the factors attained statistical significance.

The analysis shows that the direction of the trend in the exchange rate for Week 31 is forecast more accurately when the trend strength is 'intermediate' rather than 'strong'. The mean percentages of directions predicted correctly are 64.5% and 58.9% for the series with intermediate and strong trends, respectively. This difference is statistically significant ($t_{85} = 2.27$, $p = 0.026$). For the trend

Table 2

Percentage of up/down directions predicted correctly for the value of the 31st week (the number of data points in each category is given in parentheses).

	Group A		Group B	
	Trend strength 'intermediate'	Trend strength 'strong'	Trend strength 'intermediate'	Trend strength 'strong'
Trend direction "negative"	58.10% (42)	51.43% (42)	61.82% (44)	54.55% (44)
Trend direction "positive"	66.67% (42)	60.48% (42)	71.36% (44)	69.09% (44)

Table 3

Mean absolute differences between the assessed probabilities and the corresponding outcome probabilities calculated (the number of data points in each category is given in parentheses).

	Group A		Group B	
	Trend strength 'intermediate'	Trend strength 'strong'	Trend strength 'intermediate'	Trend strength 'strong'
Trend direction "negative"	0.173 (42)	0.204 (42)	0.176 (44)	0.233 (44)
Trend direction "positive"	0.174 (42)	0.210 (42)	0.190 (44)	0.217 (44)

Table 4

Mean absolute percentage error (MAPE) for the generated point forecasts (the number of data points in each category is given in parentheses).

	Group A		Group B	
	Trend strength 'intermediate'	Trend strength 'strong'	Trend strength 'intermediate'	Trend strength 'strong'
Trend direction "negative"	0.65% (42)	0.89% (42)	0.75% (44)	0.91% (44)
Trend direction "positive"	0.66% (42)	0.80% (42)	0.63% (44)	0.74% (44)

direction, the percentage of correct predictions is higher for the series with ascending trends (67.0%) than for those with descending ones (56.5%). This difference is also statistically significant ($t_{85} = 4.90, p < 0.0001$).

The probabilities provided by the participants were evaluated using the Mean Absolute Difference between the Assessed Probabilities and the Outcome Probabilities (MADP), using the formula:

MADP

$$= \begin{cases} |\text{provided prob.} - \text{outcome prob.}| & \text{if the predicted direction for week 31 is "up"} \\ |(1-\text{provided prob.}) - \text{outcome prob.}| & \text{if the predicted direction for week 31 is "down"}. \end{cases}$$

With this measure, a smaller MADP score indicates that the probabilities provided by the participants are more accurate (i.e., they are closer to the outcome probabilities theoretically calculated for those series). Table 3 presents the average absolute differences between the assessed probabilities and the corresponding outcome probabilities.

Full factorial ANOVA reveals that the only factor that seems to have an influence on the absolute difference between the provided probabilities and the outcome probabilities is the *trend strength* factor ($F_{1252} = 36.87, p < 0.0001$). Neither the main effects for the *group* and *trend direction* factors nor any of their interactions seem to have a significant impact on the MADP scores.

For series with intermediate strength trends, the assessed and outcome probabilities seem to be more similar, resulting in a MADP score of 0.178, which is significantly smaller than the score (0.216) for series with strong trends ($t_{85} = -5.07, p < 0.0001$). An intermediate trend strength seems to lead the participants to provide probabilities that are closer to the theoretically calculated ones.

3.2. Point forecasts

Point forecasts were evaluated via the Mean Absolute Percentage Error (MAPE) measure, using the formula:

$$\text{MAPE} = \frac{|\text{point forecast} - \text{actual value for the 31st week}|}{\text{actual value for the 31st week}} \times 100.$$

The average MAPE scores for the point forecasts are exhibited in Table 4. Full factorial ANOVA demonstrates that the "within-subject" factors *trend direction* and *trend strength* both have significant main effects on the trend accuracy of the point forecasts generated ($F_{1252} = 14.54, p < 0.0001$; and $F_{1252} = 4.83, p = 0.029$, respectively). None of the interaction effects among the factors were found to have a significant influence on the MAPE scores, nor was the main effect of the *group* factor.

The results show that the accuracy of the assessed point forecasts is higher when the trend has an intermediate

strength than when it is strong. The MAPE for the series with intermediate trend strengths is 0.67%, while it is 0.84% for the series with strong trends. This difference in MAPEs is significant ($t_{85} = 3.39, p = 0.001$). For the trend direction effect, the participants perform better for ascending series than for descending ones, resulting in smaller mean absolute errors. There is a significant difference ($t_{85} = 2.69, p = 0.009$) between the MAPE scores for series with ascending and descending trends (0.71% and 0.80%, respectively.)

4. Consistency methodology

An important aim of the study was to examine the influence that the trend characteristics of the exchange rate series had on participants' consistency. The participants' directional and half range probability forecasts were partitioned into matched pair descending and ascending trended series. Their directional forecasts for descending and ascending trends are denoted g_{1ij} and g_{2ij} respectively, where i denotes the subject ($i = 1, 2, \dots, n$) and j denotes the matched series ($j = 1, 2, \dots, k$). A value of 1 indicates a predicted fall and a value of 2 a predicted rise. The directional probability forecast (which was obtained by dividing the percentage probability forecast by 100) is denoted r_{1ij} and r_{2ij} for descending and ascending trends respectively. In this study, 10 paired series were used (i.e., $k = 10$), with a total of 86 participants (i.e., $n = 86$), partitioned into two groups, with Group A having 42 participants (i.e., $n_a = 42$) and Group B having 44 (i.e., $n_b = 44$).

To undertake a consistency analysis, it was necessary to convert the half-range probabilities to full-range probabilities, which give the probability of a rise in the exchange rate on a scale from zero to unity. Values below 0.5 indicate a predicted fall and values above 0.5 a predicted rise. To convert the half-range probabilities, g_{1ij} , r_{1ij} , g_{2ij} and r_{2ij} , to full-range probabilities, h_{1ij} and h_{2ij} , the absolute value expressions $h_{1ij} = |g_{1ij} - 2 + r_{1ij}|$ and $h_{2ij} = |g_{2ij} - 2 + r_{2ij}|$ are used. For example, if a full-range probability prediction of 0.73 is made, then the half-range probability would be 0.73 with a rise predicted. If a full-range probability of 0.24 is made, the half-range probability would be 0.76 with a fall predicted. The full-range 0.5 probability, i.e., no change prediction, could be assigned arbitrarily as a rise or fall with a half-range probability equal to 0.5.

To examine consistency, a further simple adjustment was made. For the descending series, the full-range probability was subtracted from one to give the adjusted full-range probability: $f_{1ij} = 1 - h_{1ij}$. For the ascending series, the same full-range probability value was used to give the adjusted full-range probability: $f_{2ij} = h_{2ij}$.

4.1. Statistical measures for consistency

The matched paired adjusted full-range values, f_{1ij} and f_{2ij} , were used in the consistency analysis. When predictions are totally consistent, these probabilities should be equal, i.e., $f_{1ij} = f_{2ij}$, so that situations where the values are not equal reflect some degree of inconsistency. Consistency was examined in three forms: *Form 1*, where

each participant i ($i = 1, 2, \dots, n$) is considered across all series j ($j = 1, 2, \dots, k$); *Form 2*, where each series j is considered across all participants i ; and *Form 3*, where all series j are considered across all participants i .

There are two overall consistency measures that can be used.

The *Mean Squared Consistency Score (MSCS)* for *Form 1* was computed using the adjusted probability responses, f_{1ij} and f_{2ij} . The *MSCS* for each individual i , over the j series, is defined in Eq. (5):

$$MSCS_i = \frac{1}{k} \sum_j (f_{1ij} - f_{2ij})^2. \quad (5)$$

Eq. (5) for *Form 2* (and the other equations below) only requires the subscript i for the measure to be replaced by j ($MSCS_j$), the divisor k to be replaced by n (or n_a and n_b) and the summation to be over i rather than j . Eq. (5) for *Form 3* (and the other equations below) would only require the omission of the subscript i ($MSCS$), the replacement of the divisor k by kn (or kn_a and kn_b), and the change of the summation over j to a double summation over i and j .

The *Mean Absolute Consistency Score (MACS)* can also be obtained. The *MACS* for each individual, i , is defined in Eq. (6):

$$MACS_i = \frac{1}{k} \sum_j |f_{1ij} - f_{2ij}|. \quad (6)$$

A value of zero on both these measures implies that an individual has made perfectly consistent predictions across the descending and ascending trended series. The *MSCS*, however, penalises large inconsistencies or differences between f_{1ij} and f_{2ij} more heavily than the *MACS*.

When analysing consistency, it is desirable to consider hypothetical forecasters. The *random walk* forecaster would make all predictions with a half range probability of 0.5 in an arbitrary direction. Hence, $f_{1ij} = f_{2ij} = 0.5$ for all i and j . In this case, the forecasts would be perfectly consistent, although of limited use in a practical context, and $MACS_i = MSCS_i = 0$.

The *perfect* forecaster would make probability forecasts that are precisely in line with the *adjusted outcome probability*, which is equal to unity minus the *outcome probability* for descending trends and the *outcome probability* for ascending trends. Hence, $f_{1ij} = f_{2ij}$ for all i and j , with the values being identical to the adjusted outcome probabilities (i.e., for the series used in this study, these were: 0.5299, 0.5329, 0.5369, 0.5388, 0.5822, 0.6160, 0.6189, 0.6208, 0.6227, 0.6378). In this case, the forecasts would, of course, also be perfectly consistent, with $MACS_i = MSCS_i = 0$.

The *uniform* forecaster would give probability predictions arbitrarily, following a continuous *random uniform* distribution with a lower limit of zero and an upper limit of unity. For this forecaster, the expected $MACS_i = 0.3333$ and the $MSCS_i = 0.1666$.

The *MSCS* is an overall consistency measure whose decompositions identify specific aspects of the consistency performance. The *MSCS* can be decomposed in a number of ways. The decomposition proposed here follows the lines used by Wilkie and Pollock (1996) in relation to the

evaluation of directional probability performance. This is presented in Eq. (7):

$$MSCS_i = V_i(f_1) + V_i(f_2) - 2C_i(f_1, f_2) + [M_i(f_1) - M_i(f_2)]^2, \quad (7)$$

where M denotes the mean, such that:

$$M_i(f_1) = \frac{1}{k} \sum_j f_{1ij} \quad \text{and} \quad M_i(f_2) = \frac{1}{k} \sum_j f_{2ij};$$

V denotes the variance, such that:

$$V_i(f_1) = \left(\frac{1}{k} \sum_j f_{1ij}^2 \right) - M_i(f_1)^2 \quad \text{and}$$

$$V_i(f_2) = \left(\frac{1}{k} \sum_j f_{2ij}^2 \right) - M_i(f_2)^2;$$

and C denotes the covariance, such that:

$$C_i(f_1, f_2) = \left(\frac{1}{k} \sum_j f_{1ij}f_{2ij} \right) - M_i(f_1)M_i(f_2).$$

The first two terms on the right hand side of Eq. (7) imply that the measure is affected by the sum of the variation in the predictions for the descending and ascending series for an individual i . In the case where a subject views all of the series as a *random walk*, the variance terms will be zero. For the *perfect forecaster*, the variance terms would each equal 0.0018, or a total for both of 0.0036 for the series presented. For the *uniform forecaster*, the variance terms would have expected values for each of 0.0833 or a total for both of 0.1666.

The third term of Eq. (7) reflects covariation between the adjusted probability predictions for the paired descending and ascending series for an individual i . For a *random walk* forecaster, the covariation will be zero. For the *perfect* forecaster, the value for the presented series of the covariance would be 0.0018 (the same as the variance); hence, for this forecaster, the first three terms of Eq. (7) sum to zero. That is, the variation presented by a *perfect* forecaster across the descending and ascending trended series is explained fully by the perfect correlation. For the *uniform* forecaster, the covariance would be zero.

The last term of Eq. (7) reflects the squared difference between the means of the adjusted probability predictions for the descending and ascending trended series for an individual i , and is termed the *squared bias*. A value of zero for this measure indicates an absence of bias. *Bias* (without squaring) can be used as a measure to indicate whether an individual tends to give higher adjusted probabilities for ascending trends than for descending trends, or vice versa. In the case of the *random walk* forecaster, both mean terms will be 0.5, and hence the difference will be zero. In the case of the *perfect* forecaster, for the presented series, the means will be 0.5838, with the bias being zero. The *uniform* forecaster would have an expected value of 0.5 for each mean, and hence a zero bias.

This decomposition of the *MSCS* illustrates the various different aspects of consistency. In the evaluation of consistency, it is not sufficient to be interested only

in a high correlation; one must also consider the bias in the average responses. For instance, do individuals tend to overestimate probabilities for ascending trends but underestimate (adjusted) probabilities for descending trends? In addition, differences in variation can also be considered. For instance, do individuals have lower levels of variation in their ascending trend predictions than in their descending trend predictions? Therefore, the examination of consistency requires a consideration of correlation, as well as of differences in location and variation.

The *MSCS* (and *MACS*) is, of course, related to the degree of variation in the predictions. For instance, it is more difficult to achieve a low value on the *MSCS* where a considerable degree of variation in probability predictions is exhibited, as this high level of variation needs to be explained by a higher covariance or correlation. For instance, both the *random walk* and *perfect* forecasters show perfect consistency on the *MSCS*. However, an *Adjusted MSCS* measure, *AMSCS*, can be used to partly offset this problem. This is given in Eq. (8):

$$AMSCS_i = AC_i + ABS_i, \quad (8)$$

where

$$AC_i = 1 - \{C_i(f_1, f_2) / \sqrt{[V_i(f_1)V_i(f_2)]}\}$$

and

$$ABS_i = \{[M_i(f_1) - M_i(f_2)]^2\} / \{2\sqrt{[V_i(f_1)V_i(f_2)]}\}.$$

In Eq. (8), the first term is *Adjusted Correlation (AC)*, which involves a unity term less the correlation value between the adjusted probabilities for ascending and descending trending series. The lower the value of this measure, the better the consistency. Here, the *perfect* forecaster has a value of zero and the *uniform* forecaster has a value of unity. The second term, *Adjusted Bias Squared (ABS)*, is the squared bias term divided by two and the square root of the sum of the variances of the adjusted probabilities. This value would also be zero for the *perfect* forecaster. Therefore, the *AMSCS* has a value of zero for the *perfect* forecaster and a value of unity for the *uniform* forecaster. In the case of the *random walk* forecaster, there would be an undefined value, as the divisor is zero. Values above unity indicate an extremely poor consistency performance, as this reflects poor or negative correlation and bias.

4.2. Statistical tests on the consistency measures

To examine specific aspects of consistency, statistical tests were applied to the component measures. It is not practical to apply statistical tests to the overall measures, as the distribution is not symmetric around the best possible value of zero, and a value of zero would be obtained for both the *random walk* and *perfect* forecasters. In addition, due to potential non-normality of the distribution of probability responses, consistency measures were examined using both parametric and non-parametric tests.

To consider whether subjects showed a *consistency bias* (i.e., whether differences occurred in the paired adjusted

probabilities for each series j (and all series) across all individuals (n , n_a and n_b), the Wilcoxon signed rank and paired samples t -tests were applied. To examine whether the subjects showed some degree of *consistency correlation* (i.e., positive correlation), the non-parametric Spearman rank correlation tests and the parametric Pearson product moment correlation tests were applied and presented, together with coefficient values.

In addition, to consider whether the subjects exhibited a *performance bias* (i.e., whether adjusted probability predictions underestimated or overestimated trends), a comparison was made with the adjusted outcome probability using one-sample t -tests and sign tests. This allowed the integration of the performance biases in trend recognition with the consistency analysis (correlation and consistency biases).

5. Consistency analysis results

The analysis was undertaken using a range of consistency statistics for the adjusted probabilities, including MACS, MSCS, AMSCS, ABS, Pearson and Spearman correlation coefficients, and the adjusted probability means and variances, together with p -values for test measures based on correlation (Pearson and Spearman) and location (t -test and Wilcoxon test). In addition, the analysis of performance bias was undertaken by using one-sample t -tests and sign tests.

5.1. Consistency results for participants across all series

Table 5 provides a range of summary measures and statistical test results for the paired adjusted probabilities of the individuals in each group. Just over half the participants overall (23 from A and 26 from B) had AMSCS values of less than one (i.e., better than the *uniform* forecaster). For the Pearson correlation, 27 participants from Group A and 35 from Group B gave positive values, but only 12 and 14, respectively, reached significance at the 10% level. For the Spearman correlation, 20 participants from Group A and 33 from Group B gave positive values, but only 9 were significant at the 10% level for Group A and 10 for Group B.

The other statistical test measures reflect the bias aspects of inconsistency. The paired sample t -test and the Wilcoxon test both showed 9 participants in each group to be significant at the 10% level. They also showed that these participants gave higher adjusted probabilities for the ascending series than for the descending series, in line with the forecasting performance analyses presented in Section 3. An analysis of the means of the significant paired sample t -test adjusted probabilities for the 10% significance case also provided interesting results. 11 participants from the 18 significant cases gave adjusted probabilities for ascending trends above 0.5 and descending below 0.5 (with group means of 0.433, 0.668), illustrating that these participants tended to identify the direction of ascending trends correctly, but identified descending trends incorrectly as ascending ones. This could

indicate extreme damping of descending trends, or, alternatively, could suggest that these participants used one approach for ascending trends (e.g., trend damping), and a different forecasting approach for descending trends (e.g., sequential dependence, see Harvey, 2011, personal communication).

Some individuals with this extreme bias came from each group. Therefore, the basic group manipulation of the present study does not appear to be related to this surprising finding.

The mean adjusted probability responses, however, showed a significant difference between the two groups for descending series ($p < 0.001$), with Group A having a mean value of 0.522, compared to Group B's value of 0.559. Again, this fits in with the accuracy analyses above, which demonstrated significant differences in probability performance, which could, in turn, indicate extreme dampening of descending trends, particularly for the subjects in Group A, who were not asked the additional trend-related questions.

5.2. Results for the paired series across the two groups of participants

The results showed substantial variations in consistency between the 10 paired series across the two groups. Tables 6a and 6b provide a range of consistency statistics for Groups A and B respectively for each of the 10 paired series (identified at the top of the columns) and all series. Relevant values for the actual series/*perfect* forecaster are also presented. The adjusted outcome probability (Adj. Out. Prob.) for each paired series is presented in the second row, with values below 0.59 (the first five series) viewed as intermediate trends and values above 0.61 (the next five series) viewed as strong trends.

The AMSCS had values of less than unity in all but three cases (i.e., paired series 13, 2 for both groups and paired series 7, 12 for Group A), with the values for Group A varying between 0.568 and 1.028 and those for Group B varying between 0.430 and 1.144. The Pearson correlation coefficients for both groups and the Spearman correlation coefficients for Group B were both positive for all paired series except one (i.e., 13, 2). The Pearson correlation showed that Group A had four significant values (at the 5% level) while Group B had six. The Spearman test statistics showed that Group A had four significant values and Group B had six. Therefore, Group B showed a slightly better correlation consistency. The paired sample t -test statistics for Group A showed three significant values, while Group B had five. The Wilcoxon test statistics for Group A showed three significant values and those for Group B showed five. Therefore, the participants from Group B showed a higher level of inconsistency on the adjusted mean probability responses than those in Group A. The one-sample t -tests and sign tests gave four significant values for both tests for Group A on descending trend series, with trend dampening being illustrated in all four cases. There were also four significant t -test values and three sign test values for Group A on ascending trend series, which illustrated both dampening and anti-dampening. For Group B, on downward trending series,

Table 5
Group comparison across all series.

Measure/group	A (42)			B (44)				
	Number	Significant at			Number	Significant at		
		10%	5%	1%		10%	5%	1%
AMSCS < 1	23				26			
Pearson's correlation (positive values)	(27)	12	8	4	(35)	14	9	5
Spearman's correlation (positive values)	(20)	9	5	4	(33)	10	7	4
<i>t</i> -test ([] denotes number with difference in means negative)		9 [7]	5 [4]	2 [2]		9 [9]	7 [7]	2 [2]
Wilcoxon signed rank test ([] denotes number with median negative)		9 [8]	4 [3]	0 [0]		9 [9]	6 [6]	1 [1]

Table 6a
Results for the 10 paired series and all series across all subjects.

Measures	Group A											Actual, perfect forecast
	Series ^a											
	13, 2 Adj. Out. Prob.	7, 12	1, 14	16, 6	11, 3	18, 9	17, 10	5, 19	4, 20	8, 15	All	
	0.530	0.533	0.537	0.539	0.582	0.616	0.619	0.621	0.623	0.638	0.584	
MACS	0.210	0.196	0.191	0.198	0.188	0.158	0.176	0.151	0.218	0.188	0.187	0.000
MSCS	0.085	0.080	0.083	0.070	0.068	0.053	0.061	0.058	0.101	0.101	0.076	0.000
AMSCS	1.028	1.028	0.936	0.929	0.724	0.617	0.633	0.568	0.935	0.868	0.746	0.000
PCorr	-0.020	0.123	0.065	0.149	0.300	0.512	0.440	0.442	0.065	0.134	0.268	1.000
SCorr	0.022	0.040	0.133	0.176	0.312	0.455	0.449	0.445	0.110	0.165	0.276	1.000
PCorr (<i>p</i> -value)	0.550	0.220	0.341	0.172	0.027	0.000	0.002	0.002	0.341	0.198	0.000	
SCorr (<i>p</i> -value)	0.495	0.401	0.201	0.132	0.022	0.001	0.001	0.002	0.244	0.148	0.000	
Mean (<i>f</i> ₁)	0.547	0.503	0.440	0.563	0.575	0.472	0.535	0.622	0.557	0.406	0.522	0.584
One S. <i>t</i> (<i>p</i> -value)	0.588	0.404	0.006	0.453	0.824	0.000	0.013	0.961	0.075	0.000	0.000	
Sign (<i>p</i> -value)	1.000	0.164	0.008	0.644	0.644	0.000	0.000	0.644	0.044	0.000	0.003	
Mean (<i>f</i> ₂)	0.522	0.609	0.430	0.640	0.623	0.574	0.619	0.591	0.563	0.423	0.559	0.584
One S. <i>t</i> (<i>p</i> -value)	0.801	0.005	0.002	0.001	0.257	0.246	0.993	0.443	0.110	0.000	0.031	
Sign (<i>p</i> -value)	0.441	0.164	0.008	0.020	0.164	0.280	0.280	0.644	0.044	0.000	0.665	
Paired <i>t</i> (<i>p</i> -value)	0.518	0.014	0.829	0.061	0.239	0.003	0.026	0.413	0.901	0.742	0.006	
Wilcoxon (<i>p</i> -value)	0.537	0.024	0.668	0.063	0.116	0.006	0.013	0.304	0.758	0.811	0.000	
ABS	0.008	0.151	0.001	0.078	0.025	0.129	0.073	0.010	0.000	0.002	0.014	0.000
Var(<i>f</i> ₁)	0.040	0.050	0.047	0.043	0.042	0.031	0.043	0.043	0.053	0.055	0.049	0.002
Var(<i>f</i> ₂)	0.043	0.027	0.041	0.032	0.051	0.053	0.053	0.059	0.055	0.062	0.053	0.002

^a Note: The first number relates to the downward trending series and the second to the upward trending series.

there were five and three significant values, respectively, again split between the dampening and anti-dampening of trends; while on upward trending series, there were six significant values for each test, with the majority of these values indicating anti-dampening. This tendency to dampen descending trends relative to ascending trends not only caused performance bias, but also contributed to consistency bias.

On the AMSCS, there were generally only limited differences between Groups A and B. The differences were greater than 0.1, in absolute terms, for five paired series (i.e., 13, 2; 7, 12; 18, 9; 5, 19; and 4, 20). The discussion here concentrates mainly on these five paired series. However, the results for all series are presented in Tables 6a and 6b.

The largest absolute difference (i.e., 0.328) occurred for paired series (7, 12), with Group B showing the best consistency. This series showed a general movement against the underlying directional trend to week 8, followed by a general directional trend to week 23, then a flattening out to week 27, a movement against the underlying trend in week 28, and the last two movements following the direction of the overall trend.

These contradictory movements towards the end of the series may have favoured Group B, who would have been more likely to view the overall trend rather than the recent values. For correlation, both the Pearson and Spearman coefficients were highly significant for Group B, but neither was significant for Group A. The paired samples *t*-test and Wilcoxon test were both highly significant for Group B and significant for Group A, with the ABS being relatively high for both groups. Both groups showed considerably higher adjusted probabilities for the ascending trend series, with the one-sample *t*-tests being highly significant for both groups and the sign test highly significant for Group B, indicating that the participants anti-dampened the trend. The better AMSCS value for Group B was explained by the higher correlation.

The second highest absolute difference in AMSCS (i.e., 0.160) occurred for paired series (18, 9), with a better consistency illustrated by Group A. This series showed a relatively well defined, directional trend, with the last four weekly movements towards the end of the series following the direction of the underlying trend. However, a marked contradictory movement against the directional

Table 6b

Results for the 10 paired series and all series across all subjects.

Measures	Group B											Actual, perfect forecast
	Series ^a											
	13, 2	7, 12	1, 14	16, 6	11, 3	18, 9	17, 10	5, 19	4, 20	8, 15	All	
	Adj. Out. Prob.											
	0.530	0.533	0.537	0.539	0.582	0.616	0.619	0.621	0.623	0.638	0.584	0.584
MACS	0.253	0.183	0.189	0.160	0.180	0.208	0.160	0.163	0.229	0.243	0.197	0.000
MSCS	0.102	0.059	0.080	0.050	0.058	0.079	0.062	0.060	0.107	0.113	0.077	0.000
AMSCS	1.144	0.700	0.900	0.887	0.680	0.777	0.549	0.430	0.818	0.953	0.668	0.000
PCorr	-0.085	0.427	0.108	0.205	0.367	0.362	0.514	0.602	0.249	0.269	0.388	1.000
SCorr	-0.156	0.366	0.191	0.251	0.467	0.353	0.587	0.646	0.233	0.195	0.399	1.000
PCorr (p-value)	0.707	0.002	0.242	0.091	0.007	0.008	0.000	0.000	0.052	0.039	0.000	
SCorr (p-value)	0.844	0.007	0.107	0.050	0.000	0.009	0.000	0.000	0.064	0.102	0.000	
Mean (f_1)	0.468	0.556	0.480	0.604	0.647	0.518	0.590	0.643	0.582	0.352	0.544	0.584
One S. t (p-value)	0.045	0.497	0.091	0.038	0.033	0.004	0.460	0.588	0.335	0.000	0.001	
Sign (p-value)	0.174	0.451	0.291	0.010	0.096	0.049	0.451	0.291	0.880	0.000	0.886	
Mean(f_2)	0.540	0.659	0.452	0.673	0.710	0.636	0.673	0.710	0.674	0.511	0.624	0.584
One S. t (p-value)	0.763	0.000	0.009	0.000	0.000	0.579	0.112	0.027	0.155	0.004	0.000	
Sign (p-value)	1.000	0.000	0.049	0.000	0.001	0.880	0.096	0.023	0.291	0.049	0.000	
Paired t (p-value)	0.132	0.004	0.524	0.041	0.081	0.004	0.025	0.070	0.060	0.001	0.000	
Wilcoxon (p-value)	0.231	0.009	0.335	0.054	0.044	0.007	0.037	0.179	0.120	0.003	0.000	
ABS	0.060	0.127	0.009	0.092	0.048	0.139	0.062	0.032	0.066	0.222	0.056	0.000
Var(f_1)	0.039	0.047	0.047	0.040	0.037	0.045	0.064	0.074	0.076	0.044	0.059	0.002
Var(f_2)	0.050	0.037	0.042	0.016	0.048	0.056	0.048	0.065	0.054	0.074	0.056	0.002

^a Note: The first number relates to the downward trending series and the second to the upward trending series.

trend occurred between weeks 23 and 26. In this case, Group A's better performance on the AMSCS measure could reflect the fact that this group could have been more likely to concentrate on extrapolating the recent values, which would have been more straightforward than making a prediction based on the underlying trend. This is supported by results which show that Group A had higher Pearson and Spearman coefficients than Group B, but the coefficients for both groups were highly significant for both measures. The correlation values were higher for Group A, and all correlation test values were highly significant. The means were all below the adjusted outcome index, except for Group B in the case of ascending trends. The one-sample t -tests for both groups were highly significant for descending trends, and the sign test was also highly significant for descending trends for Group A and significant for Group B. The paired sample t -tests and Wilcoxon tests were both highly significant for both groups. The ABS values were relatively high for both groups. The relatively better consistency for Group A is explained by the better correlation.

The third highest absolute difference in AMSCS (i.e., 0.138) occurred for paired series (5, 19), with better performance for Group B. This series showed a relatively well defined, directional trend, with the last four weekly movements following the direction of the underlying trend and with the last value being particularly pronounced. This very sharp movement may have favoured Group B and resulted in more consistent predictions, based on the underlying trend against extrapolations based on recent values. The correlation values were higher for Group B, but all correlation test values were highly significant. For ascending trends, the group B mean was above the adjusted outcome probability and significant for the one-sample t -test and sign test. The paired sample t -tests and Wilcoxon tests were non-significant. The ABS values were

low for both groups. The relatively good consistency arose from a relatively good correlation with low bias.

The fourth highest absolute difference in AMSCS (i.e., 0.117) occurred for paired series (4, 20), with a better performance illustrated for Group B. This series showed a relatively clear overall directional trend to week 18, after which the trend flattened out to 28, with a clear movement in the direction of the trend in the last two weeks. The marked movement in weeks 29 and 30 may have resulted in trend confirmation, such that the more consistent predictions for Group B were based on the underlying trend, as opposed to extrapolations based on recent values. Both correlation test values were non-significant. The means were all below the adjusted outcome probability, except for Group B in the case of ascending trends. Only the sign tests showed significance, and this was for Group A on descending and ascending trends. The paired samples t -tests and Wilcoxon tests were non-significant. The ABS values were very low for Group A but higher for Group B. The relatively poor consistency can be explained largely by the poor correlation.

The fifth highest absolute difference in AMSCS (i.e., 0.117) occurred for paired series (13, 2), with a better performance illustrated for Group A. This was clearly the poorest series for consistency performance, with both groups having AMSCS values less than unity. This series showed a clear directional trend from weeks 6 to 21, but then a movement against the trend until week 28, followed by a movement with the trend in week 29; week 30 showed little change. It appears that the subjects found it very difficult to achieve consistency for this series. Group A showed slightly better correlation coefficient values than Group B, but only the Spearman value for Group A was positive. The paired sample t -test and Wilcoxon test were non-significant. The one-sample t -tests and sign tests were all non-significant, except in the case of the t -test for Group

B for descending trends, which was significant, with the mean below 0.5. The *ABS* value was low for Group A and relatively average for Group B. The very poor consistency appears to have been caused largely by an extremely poor correlation.

In general, the results show that better consistency occurs with reasonably strong trends which have clear overall directional patterns, and movements toward the end of the series that are in a similar direction to the trend. In this case, there is a tendency for the probability of a trend movement to be overestimated (particularly for Group B). A poorer consistency tends to be associated with weaker trends, and, to a lesser extent, stronger trends which show periods of movements against the overall trend or movements towards the end of the series that occur in the opposite direction to the trend. In this case, there was a tendency for descending trends to be dampened (especially for Group A) and ascending trends to be anti-dampened (especially for Group B).

5.3. Combined results for all series across all participants

The combined results for all participants and all series (also Tables 6a and 6b) showed *MACS* and *MSCS* values better than the *uniform* forecaster, with *AMSCS* values of 0.746 for Group A and 0.668 for Group B. Both the Pearson and Spearman correlation coefficient values were highly significant for both groups, with values of 0.268 and 0.276 for Group A and 0.388 and 0.399 for Group B respectively. The paired *t*-test and Wilcoxon test were highly significant for both groups, with the *ABS* being much higher for Group B (0.056) than for Group A (0.014), which was more than offset by the higher correlation of Group B. In general, there was a marked underestimation of descending trends, compared with a value for the adjusted outcome probability (0.584), with the participants from Group A having a lower adjusted probability (0.522) than those from Group B (0.544). On ascending trends, the Group A (0.559) value also indicated an underestimation of the trend, but Group B (0.624) showed an overestimation. The independent samples *t*-test and the Mann–Whitney test for the difference between the Group A and Group B values on ascending trends were highly significant. The combined results did not show any clear differences in variation, but the variation does appear to be considerably greater than the variance for the *perfect* forecaster. The results tend to indicate that, in general, combining all of the individual probability forecasts gives more consistent predictions than those obtained from individual subjects considered in Section 5.1. However, there was clear evidence from the combined forecasts that the participants in Group A dampened trends, especially for descending series. There was less damping on descending trends for participants in Group B (who were asked the additional trend-related questions), with anti-damping emerging on ascending trends.

6. Discussion

This study examines the performance of judgmental exchange-rate forecasts and the influence of two important time series characteristics: trend direction and trend

strength. Its contributions to the extant literature are original at various levels. First, the study uses actual data and systematically controls these two important time series characteristics simultaneously in order to examine their combined impact on forecasting behaviour. While previous work has only studied trend strength and direction concurrently with the use of constructed data (e.g., Thomson et al., 2003), the current study has enabled us to address the issue of whether or not the biases in relation to these important time series characteristics that have been evident in studies using constructed data actually apply to real-world phenomena. Second, the study aimed to determine whether or not the simple act of drawing participants' attention to the direction and strength of the trend would be sufficient to make a difference to the forecasts assessed. This is important, in that it contrasts considerably with past work (which, for example, provided feedback regarding the influence of certain time series characteristics, e.g. Sanders, 1997), thus complementing previous results in order to enhance our understanding of the forecasting practitioners' needs for forecast support and training. Finally, the study aimed to examine probability forecasting consistency closely in relation to these time series characteristics. Consistency is an important aspect of the time series forecasting performance which has rarely been considered, and never with respect to time series features. The implications of the findings with respect to both these issues and the use and design of FSSs are discussed next.

In relation to the directional and point forecasts, both the direction and strength of the trend were found to have a significant influence on performance. Specifically, the performance was found to be better for ascending trends than for descending trends, and better for trends of intermediate strength than for strong trends. With respect to these time series characteristics, our findings support most of the previous research using constructed data (e.g., Harvey & Bolger, 1996; Lawrence & Makridakis, 1989; O'Connor et al., 1997), which has found an ascending advantage. In addition, our current finding of a better performance for intermediate trends supports the previous findings of Wagenaar and colleagues that the forecasting bias is more pronounced on stronger trends, as well as other research both within (e.g., Glaser et al., 2007b) and without (e.g., Griffin & Tversky, 1992; Massey & Wu, 2005) the time series forecasting context. A plausible explanation for this intermediate advantage is that the forecasts were not adjusted enough given the information provided, thus leading to an inferior performance on stronger trends. However, as was suggested above, these effects may be more apparent with particularly strong trends, as in the present case. In fact, in a situation where a variety of trend strengths (very weak, weak, medium, strong, very strong) are compared in terms of their influence on the forecasting performance, it might be reasonable to expect an inverted U-shaped function. In any case, some tentative implications supporting exchange rate forecasting practice may be drawn. For instance, given that there is a choice of presenting the same exchange rate series to forecasters in two different forms, inverted or non-inverted (which essentially reverses the upward/downward movements),

it may improve the forecasting accuracy to make this choice based on the ascending frame. In addition, as was pointed out by Goodwin (2011, personal communication), our intermediate trend performance advantage might imply that a graphical toolbox could be very effective in presenting series so that the trend is perceived to be *more moderate* (as opposed to *strong*). However, given our findings and their relevance to the design of FSSs, more systematic research is needed to determine just how directional advantages are influenced by particular trend strengths.

The results in relation to the provision of additional questions to one group of participants are not straightforward. Although both the mean scores for the percentage of movement directions predicted correctly and the accuracy of point forecasts suggest that the group whose attention was drawn to the task characteristics performed better than their no-additional-information counterparts, these differences were not statistically significant. In this respect, the current task failed to provide substantial evidence on the matter. Nevertheless, significant differences on the consistency aspect of performance emerged. In particular, there was less damping for the “additional questions group” on descending trends, with anti-damping emerging on ascending trends. Such findings have implications for the design and use of FSSs. Specifically, where certain time series characteristics are found to influence the forecasting performance, simply drawing attention to these characteristics may be enough to improve important aspects of the performance. Thus, creating an environment of awareness needs to be a fundamental feature of forecast support systems.

The examination of consistency in judgmental forecasting is something that has rarely been considered, but is likely to be critical to the design and use of FSSs. This is because superior consistency is the main advantage of statistical methods over judgment; and therefore, to improve judgment overall, it makes sense to encourage forecasters to be more consistent, especially since judgment is a commonly used forecasting method in practice.

Various important results have emerged from the present consistency analyses. The most striking finding was the low level of consistency observed in participants. Indeed, as correlation is a key aspect in consistency performance, the observation that less than 75% of participants had correlation values with the correct sign, together with the low numbers of significant values overall, suggests that the participants had a considerable degree of difficulty in making consistent probability predictions.

Other measures were employed to identify inconsistencies arising from biases. In terms of the performance of participants across all series, an interesting finding is that some participants identified ascending trends correctly, but identified descending trends incorrectly as ascending trends, although the group mean was reasonably close to 0.5. As was suggested above, this finding may indicate that the situation is more complex than simply damping descending trends, and that such trends may be inviting a host of different forecasting biases to occur. Participants clearly tended to show much more variation when forecasting descending series. As such, this provides a further

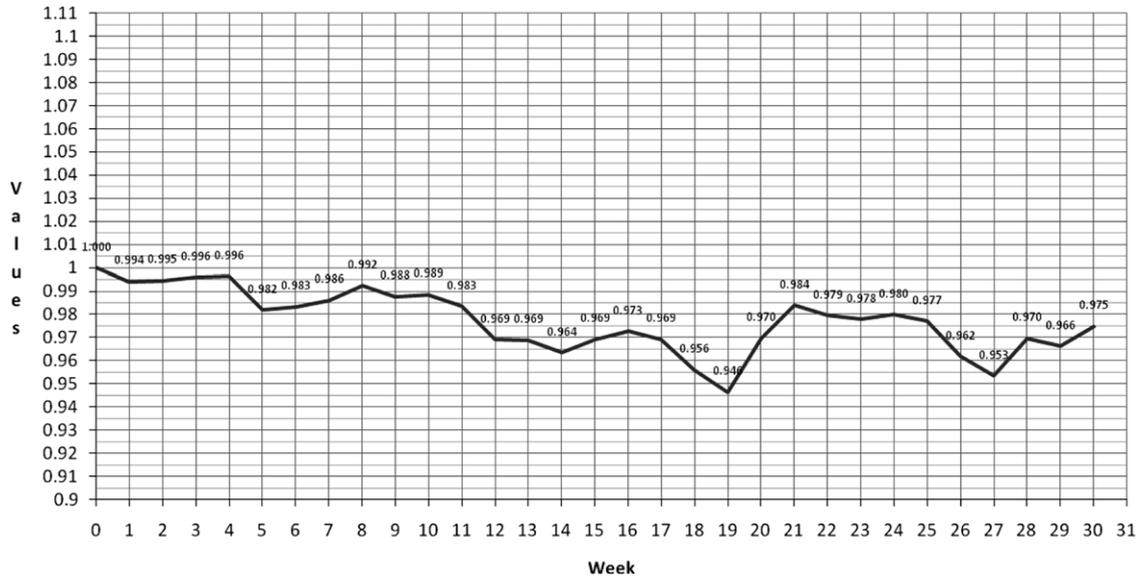
indication that exchange rate forecasting practice might benefit from a careful presentation of trends so as to remove the bias against excessive optimism (reflected via a tendency to impose ‘upward’ trends).

As was pointed out earlier, this inconsistency was not affected by encouraging participants to focus on the trend strength and direction. However, as with the general accuracy analysis, there was a significant difference between the groups’ mean adjusted probability performance on descending series, demonstrating that there was less damping in this respect from the participants whose attention was drawn to the time series characteristics under examination. In terms of FSSs design, this provides a further indication that simply drawing forecasters’ attention to important time series characteristics might be sufficient to enhance forecasting performance.

The consistency analyses for the paired series across participants illustrated interesting observations regarding the basic manipulation of highlighting the time series characteristics. Specifically, the analyses indicate that the basic group manipulation may have caused differences in forecasting behaviour in terms of whether the participants viewed the overall trend or only the most recent values. Hence, whether the particular forecasting strategy in question improves the performance or not appears to depend upon the particular pattern of the series. If contradictory movements to the underlying trend occur near the end of a series, viewing the overall trend appears to be advantageous to performance. However, the reverse seems to be true if contradictory movements to the underlying trend occur nearer to the middle of a series. These findings support the “system neglect” hypothesis of Massey and Wu (2005). Although more work needs to be carried out in relation to the effects of highlighting these and other critical time series characteristics, the current findings suggest the importance of FSSs displaying controls like highlighting either the overall trend or certain parts of the trend in obtaining improved forecasts.

The final part of the consistency analyses (which concerns the combination of the forecasts for all series across all participants), also provides notable findings which are relevant to the design of FSSs. The most obvious, albeit not altogether surprising, of these is that combining forecasts in this context improves the consistency. As groups, the participants seem to have performed quite well. Indeed, both correlation values were highly significant (with the participants whose attentions were drawn to the strength and direction of the trend performing best). The combined analyses also illustrated the damping of descending trends to be more pronounced in the group who had to acknowledge trend strength and direction explicitly; but for both groups, there was a greater damping on descending series. Accordingly, it would appear that damping is not only a real-world phenomenon, but one which requires carefully designed FSSs in order for it to be overcome, especially when forecasters are faced with descending trends. This is likely to be more prominent in other contexts where the trend cannot simply be reversed.

Series 1



Please indicate whether the exchange rate will be going up or down in week 31 : UP / DOWN

Please provide a probability indication how certain you are that your directional forecast is correct (between 50% – 100%) :

What do you think the actual value in week 31 will be :

Please state your perceived level of strength of the TREND in the given period by ticking one of the three options below :
 Weak / low Medium Strong / high

Please state your perceived TREND direction in the given period by ticking one of the three options below :
 Negative No trend Positive

7. Conclusion

Overall, the results in relation to the various manipulations of the present study contribute to the existing time series forecasting literature in several important ways. With respect to damping, the present findings support the evidence from other studies that a variety of factors are likely to affect its severity, such as the direction and strength of the trend (e.g., Thomson et al., 2003; Timmers & Wagenaar, 1977). The consistency analyses add to the existing knowledge in this context by suggesting that the effect of these time series characteristics on the damping behaviour is likely to vary as a function of merely drawing the forecaster’s attention to them in the first place, and by the specific position of contradictory movements to the underlying trend in the series, which can cause anti-damping to emerge in some circumstances. The present findings also suggest that when it comes to descending trends, the situation on damping behaviour may be much more complex than it has previously been considered. In view of these findings and the recent results of Glaser et al. (2007b), an interesting extension might be to measure returns rather than prices, to find out whether the damping behaviour related to the current manipulations still hold in these circumstances.

Acknowledgments

We are extremely grateful for the invaluable help and advice received from Nigel Harvey, Paul Goodwin and two anonymous reviewers.

Appendix A. Participants’ instructions and an example exchange rate series form

Dear Participant,

Going through the current global financial turmoil, making good and reliable exchange rate forecasts becomes even more essential for international trade and banking activities. The current study investigates the quality of judgmental directional and point forecasts generated for various real exchange rate series involving different trend structures.

Your task is the following:

20 time-series plots for exchange rates will be provided for a 30-week period. The labels of the exchange rates and the time periods used are kept confidential to prevent any potential biases or extraneous information effects.

For each time series:

- Please study the exchange rate series *VERY CAREFULLY!*
- Please provide *YOUR* judgmental *directional probabilistic forecast* for the exchange rate of 31st week.
- Please provide *YOUR* judgmental *point forecast* for exchange rate of 31st week.
- Please assess your perceived *level of trend strength* for each series.
- Please assess your perceived *trend direction* for each series.

THANK YOU FOR YOUR TIME & COOPERATION!

Appendix B. Exchange rate data information

The UK foreign exchange markets were closed on the UK Bank Holidays, which included Christmas Day, Boxing Day, New Year's Day, Good Friday, Easter Monday, May Day, Spring Holiday and Late Summer Holiday, as well as some other occasional days. These 280 Bank Holiday days were assigned an exchange rate value for that day which was the same as the previous weekday's value.

Prior to January 1999, synthetic Euro exchange rates which are provided by the Bank of England were used for the Euro exchange rates. These are calculated by geometrically weighting the bilateral exchange rates of the eleven Euro area countries, prior to 1999, using 'internal weights' based on the country shares of extra-Euro area trade. These weights are provided on the Bank of England's website.

References

- Allen, H., & Taylor, M. P. (1990). Charts, noise and fundamentals in the London foreign exchange market. *The Economic Journal*, 100, 49–59.
- Anderson, T. W., & Darling, D. A. (1954). A test of goodness-of-fit. *Journal of the American Statistical Association*, 49, 765–769.
- Andreassen, P. B. (1988). Explaining the price–volume relationship: the difference between price changes and changing prices. *Organizational Behavior and Human Decision Processes*, 41, 371–389.
- Andreassen, P. B. (1990). Judgmental extrapolation and market overreaction: on the use and disuse of news. *Journal of Behavioral Decision Making*, 3, 153–174.
- Andreassen, P. B., & Kraus, S. J. (1990). Judgmental extrapolation and the salience of change. *Journal of Forecasting*, 9(4), 347–372.
- Armstrong, J. S. (1985). *Long range forecasting: from crystal ball to computer* (2nd ed.). NY: Wiley.
- Bank for International Settlements (BIS) (2010). In *Triennial central bank survey, Report on global foreign exchange market activity*. Bank of International Settlements, Basel, Switzerland. Available on web-site: www.bis.org.
- Barber, D., & Odean, T. (2000). Trading is hazardous to your wealth: the common stock investment performance of individual investors. *Journal of Finance*, 55, 773–806.
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307–343.
- Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society, Series B*, 8, 27–41.
- Bartos, J. A. (1969). *The assessment of probability distributions for future security prices*. Unpublished Ph.D. thesis, Indiana University, Graduate Business School.
- Budescu, D., & Du, N. (2007). The coherence and consistency of investors' probability judgments. *Management Science*, 53(11), 1731–1744.
- Cheung, Y.-W., & Chinn, M. D. (2001). Currency traders and exchange rate dynamics: a survey of the market. *Journal of International Money and Finance*, 20, 439–471.
- Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, 38(10), 1394–1414.
- De Bondt, W. F. M., & Thaler, R. (1995). Financial decision making in markets and firms: a behavioral perspective. In R. A. Jarrow, V. Maksimovic, & W. T. Ziemba (Eds.), *Finance, handbooks in operations research and management science: Vol. 9* (pp. 385–410). Amsterdam, The Netherlands: North-Holland.
- Eggleton, I. R. C. (1982). Intuitive time series extrapolation. *Journal of Accounting Research*, 20, 68–102.
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25, 3–23.
- Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, 31(10), 1237–1246.
- Glaser, M., Langer, T., Reynders, J., & Weber, M. (2007a). Framing effects in stock market forecasts: the difference between asking for prices and asking for returns. *Review of Finance*, 11, 325–357.
- Glaser, M., Langer, T., & Weber, M. (2007b). On trend recognition and forecasting ability of professional traders. *Decision Analysis*, 4(4), 176–193.
- 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.
- Griffin, D., & Tversky, A. (1992). The weighting of evidence and the determinants of confidence. *Cognitive Psychology*, 24(3), 411–435.
- Harvey, N., & Bolger, F. (1996). Graphs versus tables: effects of data presentation format on judgmental forecasting. *International Journal of Forecasting*, 12(1), 119–137.
- Keren, G. (1983). Cultural-differences in the misperception of exponential-growth. *Perception and Psychophysics*, 34(3), 289–293.
- Lawrence, M., Edmundson, R. H., & O'Connor, M. (1985). An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting*, 1, 25–36.
- Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: a review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), 493–518.
- Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, 43, 172–187.
- Lawrence, M., & O'Connor, M. (1995). The anchoring and adjustment heuristic in time series forecasting. *Journal of Forecasting*, 14, 443–451.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge University Press.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The accuracy of extrapolation methods: results of a forecasting competition. *Journal of Forecasting*, 1(2), 111–153.
- Massey, C., & Wu, G. (2005). Detecting regime shifts: the causes of under- and overreaction. *Management Science*, 55(6), 932–947.
- Murphy, J. J. (1999). *The technical analysis of financial markets*. Paramus, New Jersey: New York Institute of Finance.
- 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'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., Remus, W., & Griggs, K. (1997). Going up–going down: how good are people at forecasting trends and changes in trends? *Journal of Forecasting*, 16(3), 165–176.
- Önkal, D., & Muradoglu, G. (1994). Evaluating probabilistic forecasts of stock prices in a developing stock market. *European Journal of Operational Research*, 74, 350–358.
- Önkal, D., & Muradoglu, G. (1996). Effects of task format on probabilistic forecasting of stock prices. *International Journal of Forecasting*, 12, 9–24.
- Önkal, D., Yates, J. F., Şimga-Muşan, C., & Öztin, Ş (2003). Professional vs. amateur judgment accuracy: the case of foreign exchange rates. *Organizational Behavior and Human Decision Processes*, 91, 169–185.
- Pollock, A. C., Macaulay, A., Thomson, M. E., & Önkal, D. (2005). Performance evaluation of judgemental directional exchange rate predictions. *International Journal of Forecasting*, 21, 473–489.
- Pollock, A. C., Macaulay, A., Thomson, M. E., & Önkal, D. (2008). Using weekly empirical probabilities in currency analysis and forecasting. *Frontiers in Finance and Economics*, 5, 26–55.
- Pollock, A. C., Macaulay, A., Thomson, M. E., Önkal, D., & Gönül, M. S. (2010). Evaluating strategic directional probability predictions of exchange rates. *International Journal of Applied Management Science*, 3, 282–304.

- Reimers, S., & Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting*, 27, 1196–1214.
- Sanders, N. R. (1997). The impact of task properties feedback on time series judgmental forecasting tasks. *Omega: The International Journal of Management*, 25(2), 135–144.
- Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega: The International Journal of Management*, 31, 511–522.
- Stael Von Holstein, C. A. S. (1972). Probability forecasting: an experiment related to the stock market. *Organizational Behaviour and Human Performance*, 8, 139–158.
- Thomson, M. E., Önkal-Atay, D., Pollock, A. C., & Macaulay, A. (2003). The influence of trend strength on directional probabilistic currency predictions. *International Journal of Forecasting*, 19, 241–256.
- Timmers, H., & Wagenaar, W. A. (1977). Inverse statistics and misperception of exponential-growth. *Perception and Psychophysics*, 21, 558–562.
- Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception 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(2), 182–184.
- Wagenaar, W. A., & Timmers, H. (1979). The pond-and-duckweed problem: three experiments on the misperception of exponential growth. *Acta Psychologica*, 43, 239–251.
- Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science*, 246, 1232–1233.
- Whitecotton, S. M. (1996). The effects of experience and a decision aid on the slope, scatter, and bias of earnings forecasts. *Organizational Behavior and Human Decision Processes*, 66, 111–121.
- Wilkie, M. E., & Pollock, A. C. (1996). An application of probability judgement accuracy measures to currency forecasting. *International Journal of Forecasting*, 12, 25–40.
- Wilkie-Thomson, M. E., Önkal-Atay, D., & Pollock, A. C. (1997). Currency forecasting: an investigation of extrapolative judgement. *International Journal of Forecasting*, 13(4), 509–526.
- 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.
- Mary E. Thomson** is a Reader of Decision Sciences from The Department of Psychology, Glasgow Caledonian University. She received a Ph.D. in 1999, and has since published in a wide variety of books and journals. Her research focuses on judgmental currency forecasting, forecasting support systems, the quality of expert judgment and risk perception and communication.
- Andrew C. Pollock** is a Reader at Glasgow Caledonian University. He completed a Ph.D. on exchange rates and has published articles and papers in a variety of books and journals in the financial area. His particular research interest is the application of analytical techniques to financial forecasting, particularly in relation to currency markets, involving both judgemental and statistical approaches.
- M. Sinan Gönül** is an assistant professor in the Department of Business Administration at Middle East Technical University in Ankara in Turkey. He has carried out research in the areas of judgmental forecasting and forecasting support systems. He has published in various journals, including *Journal of Behavioral Decision Making*, *Decision Sciences*, *Decision Support Systems* and the *Journal of Forecasting*.
- Dilek Önkal** is the Dean of the Faculty of Economics, Administrative and Social Sciences at Bilkent University, Turkey. Her research interests focus on judgmental forecasting, forecasting support systems, risk perception and risk communication. Her work has appeared in several book chapters, as well as journals such as *Organizational Behavior and Human Decision Processes*, *Decision Sciences Journal*, *Risk Analysis*, *Decision Support Systems*, *International Journal of Forecasting*, *Journal of Behavioral Decision Making*, *Journal of Forecasting*, *Omega: The International Journal of Management Science* and the *European Journal of Operational Research*.