ELSEVIER

European Journal of Operational Research 114 (1999) 281–293

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Evaluating predictive performance of judgemental extrapolations from simulated currency series

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

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

Received 3 June 1997; accepted 6 July 1998

Abstract

Judgemental forecasting of exchange rates is critical for financial decision-making. Detailed investigations of the potential effects of time-series characteristics on judgemental currency forecasts demand the use of simulated series where the form of the signal and probability distribution of noise are known. The accuracy measures Mean Absolute Error (MAE) and Mean Squared Error (MSE) are frequently applied quantities in assessing judgemental predictive performance on actual exchange rate data. This paper illustrates that, in applying these measures to simulated series with Normally distributed noise, it may be desirable to use their expected values after standardising the noise variance. A method of calculating the expected values for the MAE and MSE is set out, and an application to financial experts' judgemental currency forecasts is presented.

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

1. Introduction

Uncertainty in exchange rates constitutes a problematic, yet inescapable, component of the decisions made by investors, financial agents, and firms participating in international markets (Lessard and Lightstone, 1986; Bahmani-Oskooee and Ltaifa, 1992; Steil, 1993; Chowdhry, 1995; Goldberg and Frydman, 1996). Accordingly, a plethora of models on exchange rate dynamics have been developed (e.g., Dornbusch, 1976; Kouri, 1976; Mussa, 1976; Frankel, 1979, 1983; Meese and Rogoff, 1983; van Hoek, 1992; Lastrapes, 1992; Nachane and Ray, 1993; Liu et al., 1994; Chinn and Meese, 1995; Kuan and Liu, 1995), with mixed evidence on their predictive performances.

* Corresponding author. Tel: ++44 141 331 3613; fax: ++44 141 331 3608.

0377-2217/99/$ – see front matter
PII: S0377-2217(98)00256-2

Human judgement, on the other hand, is found to play a significant role in currency forecasting practice (Pollock and Wilkie, 1992, 1993, 1996; Pollock et al., 1996; Wilkie and Pollock, 1994). It is not unusual for predictions to be made in an essentially subjective framework, for instance, in the application of chartist techniques. Chartists' extrapolations are claimed to represent major contributors to accurate forecasts of exchange rates (Pilbeam, 1995). Furthermore, recent work has indicated that chartist forecasts (i.e., extrapolations from past data) perform an important role in determining the market participants' foreign exchange positions (Allen and Taylor, 1989, 1990; Frankel and Froot, 1990; Taylor and Allen, 1992). These findings accentuate the importance of: (i) examining potential factors that may affect judgemental forecasting accuracy in the currency domain, and (ii) critically evaluating the accuracy of such forecasts.

It can be argued that characteristics of time series, such as trend, can influence the accuracy of judgemental forecasts (O'Connor et al., 1993; Webby and O'Connor, 1996). In practical situations, these factors can be masked as currency markets are subject to events or "news", that are impossible to foresee, yet have a major impact on the perceived forecasting performances. Hence, experimental settings may be employed to delineate the potential effects of such time-series characteristics. To examine the influence of particular series characteristics on judgement, it is often desirable to control the form of noise generation in a currency series, so as to provide a means for separating the noise and the signal. This can be accomplished via simulated series with known characteristics, which can, in turn, be used for assessing detailed judgemental forecasting performance. In fact, such constructed series are particularly advocated and extensively utilized in past research into judgemental forecasting accuracy (e.g., Ang and O'Connor, 1991; O'Connor and Lawrence, 1992; Lawrence and O'Connor, 1992, 1993; O'Connor et al., 1993; Bolger and Harvey, 1995a; Lim and O'Connor, 1995; Remus et al., 1995; Harvey and Bolger, 1996; Webby and O'Connor, 1996). In short, the use of abstract time series is argued to enable thorough investigations of extrapolative judgement, as this design avoids the potentially confounding effects of environmental cues on predictive accuracy (O'Connor and Lawrence, 1989).

Various accuracy measures can be applied to evaluate the predictive performance of exchange rate forecasts. Commonly used examples include the Mean Absolute Error (MAE) and the Mean Squared Error (MSE). Such measures have been typically employed in assessing the accuracy of weekly currency predictions given by commercial banks (Pollock and Wilkie, 1996; Pollock et al., 1996). Applying these measures to predictions made from simulated series, however, may not yield an appropriate portrayal of forecasting performance. Current research illustrates that, in the application of MAE and MSE to simulated series with Normally distributed noise, it may be preferable to use the expected values of these measures to evaluate forecasting performance. Accordingly, this paper outlines a method for calculating the expected values of MAE and MSE, followed by an application to financial experts' judgemental currency forecasts. The importance of this study stems from the potential implications of the findings for supporting the processes involved in selecting financial forecasters, conducting periodic performance-appraisal, determining training needs, and providing effective feedback mechanisms.

2. The use of simulated data in the context of judgemental currency forecasting

When examining predictive accuracy for judgemental currency forecasts the effects of particular series characteristics on judgement need to be delineated. In particular, trends or drifts in series constitute the key characteristics that currency forecasters attempt to identify, regardless of whether they follow the fundamentalist or the technical analyst (chartist) approach. Specifically, while the fundamentalists rely mostly on judgement to identify variables likely to cause the market to trend, chartists use judgement to make direct extrapolations from the conceivably trended series (based on the assumption that any information that can possibly influence the exchange rate is already incorporated into its value). Hence, the influence of the strength of the trend on judgement is of fundamental importance in evaluating judgemental performance in

currency forecasting. Simulated currency data are essential to examine this issue within the framework of appropriate accuracy measures. For example, we found in judgemental studies using constructed currency data that expert subjects tend to underestimate the strength of strong constant drift and overestimate the strength of weak constant drift (Pollock and Wilkie, 1993).

The psychological literature on time-series extrapolative judgement has illustrated that the use of simulated series, where subjects are given no information on the method used to construct the data, has considerable advantages over the use of actual series in the analysis of judgement (Goodwin and Wright, 1993). Although in some situations this approach may make the experiment less representative of real-world forecasting practice, O'Connor and Lawrence (1989) have argued that the quality of time-series extrapolative judgement can be effectively investigated only when other influences such as environmental cues are excluded. If such cues are not eliminated, the subject can potentially utilize non-time-series information in addition to the time-series information. Hence, it becomes impossible to attribute poor/good performance to the salient non-time-series information (Tversky and Kahneman, 1973), or to the factors specific to the series (Bolger and Harvey, 1993).

3. Considerations in applying MAE and MSE to simulated series

In applying the MSE or MAE to a set of predictions made from simulated data, it is generally desirable, both on statistical and economic grounds, to use the predictions of the first differences rather than those of the actual values. This stems from the principle that, in general, currency series are not stationary: their autocovariance functions depend on time. In particular, 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 as a difference-stationary process by Nelson and Plosser (1982). These authors distinguish between two different views concerning non-stationarity in macroeconomic time series: trend-stationarity (i.e., stationary fluctuations around a deterministic trend) and difference-stationarity (i.e., non-stationarity arising from the accumulation over time of stationary and invertible first differences). Evidence suggests, however, that trends in exchange rate series, most financial price series and many economic series tend to be associated with high, positive, first-order autocorrelation. Empirical studies, using a wide range of economic series (e.g., Nelson and Plosser, 1982; Perron, 1988; Dejong and Whiteman, 1994), are consistent with the difference-stationary view, particularly for economic data in nominal, as opposed to real or price-adjusted form. Hence, it is asserted that the currency series can be viewed as following a quasi random walk with first differences having a Normal distribution with time varying parameters (Boothe and Glassman, 1987a, b; Friedman and Vandersteel, 1982). [1] Recent applications involving weekly forecasts of the \$/£ and Yen/DM have also suggested that the assumption of Normally distributed first differences with time-varying parameters is appropriate (Pollock and Wilkie, 1996; Pollock et al., 1996). Within this framework, currency series can be made stationary via simple transformations. In particular, taking first differences of a difference-stationary series with a linear trend simultaneously removes the trend and the first order autocorrelation of unity, resulting in a differenced series with constant drift and zero first order autocorrelation.

[1] Earlier studies of the statistical characteristics of exchange rates (e.g., Westerfield, 1977) proposed a Stable Paretian distribution {i.e., a distribution that is more peaked and has fatter tails than the Normal (of which the Normal is a particular class)}. Now, however, it is recognised that observed non-Normality can often be explained by a mixture of Normal distributions with time-varying parameters. Furthermore, the Central Limit Theorem would suggest that, as exchange rate changes between two points in time are essentially the sum of exchange rate changes over shorter horizons, the distribution will tend to Normality, even if the underlying distribution is not Normal, provided this underlying distribution is stable.

The quasi-random walk nature of exchange rate behaviour has implications for the cognitive processes involved in forming judgemental predictions. It can be argued that effective judgemental prediction requires the consideration of the underlying probability distribution on which a series is perceived to be formed (Keren, 1991). Accordingly, it may be desirable for judgemental directional predictions to be based on the assumption of Normally distributed currency movements (Wilkie and Pollock, 1996). Research in this domain is definitely lacking. In particular, much of the previous work examining judgemental accuracy has addressed non-financial trend-stationary series, albeit with low levels of autocorrelation introduced by an Autoregressive Moving Average process and Normally distributed errors (e.g., Bolger and Harvey, 1993, 1995b; Lawrence and O'Connor, 1992). The current study attempts to extend further the judgemental accuracy research to the financial forecasting domain via an application addressing the difference-stationary nature of currency series.

The difference-stationary form of exchange rate series also has implications on the simulation of series: it is more appropriate to generate data using first differences than actual values. The resulting actual changes and predicted changes can then be used to compute the accuracy measures mentioned previously. It should be noted that, since the proposed framework addresses differences (which can be equal to zero) rather than actual values, it prohibits the use of another acclaimed accuracy statistic: the Mean Absolute Percentage Error (For discussions on the choice of error measures, see Armstrong and Collopy (1992), Fildes (1992), Clements and Hendry (1993), Mathews and Diamantopoulos (1994), Armstrong and Fildes (1995)).

A series generated by a difference-stationary process can, in practice, be used in two basic ways. Firstly, *time based* tasks involve consecutive predictions on a single series over a moving period. Secondly, *cross section* based tasks involve predictions from a number of different series. A set of n predictions can be obtained using either of these tasks. In any case the accuracy of predictions can be analysed by comparing the actual change (a_i') with the predicted change (p_i') for $i = 1, 2, \ldots, n$ forecast occasions.

Given that the actual change (a_i') can be viewed as the sum of the signal (s_i') and Normally distributed noise (w_i') {i.e., $a_i' = s_i' + w_i'$}, variations in actual changes are directly related to the size of the standard deviation (σ_i) of the noise term, which can in turn vary across the $i = 1, 2, \ldots, n$ forecast occasions. It is appropriate, therefore, to scale the actual and predicted changes by the standard deviation: the actual change (a_i') is divided by the standard deviation (σ_i) to give a scaled actual change ($a_i = a_i'/\sigma_i$). The signal component ($s_i = s_i'/\sigma_i$) is then measured relative to the standard deviation, and the error term ($w_i = w_i'/\sigma_i$) follows a Standard Normal distribution. It is, of course, also necessary to scale similarly the predicted change (i.e., $p_i = p_i'/\sigma_i$). The scaled actual change (a_i) is, therefore, the sum of the scaled signal (s_i) and noise (w_i) components {i.e., $a_i = s_i + w_i$}. These transformations recognise the fact that large forecast errors are more likely in high noise situations than in low noise situations. Where there exists a mixture of high and low noise series, or a comparison is to be made between them, the above transformations are particularly appropriate. Furthermore, they allow a more straightforward derivation of the expected values of the MSE and MAE.

Once these adjustments have been made, the scaled predicted change (p_i) and actual change (a_i) can be compared for a set of n forecasts. This is accomplished using the mean square error or mean absolute error, calculated from combined signal and noise components, and denoted $\mathrm{MSE_a}$ and $\mathrm{MAE_a}$. They are defined in Eqs. (1a) and (1b), respectively:

$$\mathrm{MSE_a} = \frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)^2, \tag{1a}$$

$$\mathrm{MAE_a} = \frac{1}{n}\sum_{i=1}^{n}|p_i - a_i|. \tag{1b}$$

The problem with these measures is that the random behaviour of the error term (w_i) influences the resulting values of MSE_a and MAE_a. One approach to overcome this problem is to ignore the error term (w_i) in the calculation of the accuracy measures, concentrating only on the signal term (s_i). The mean square error and mean absolute error could be computed, therefore, using only the signal term, giving MSE_s and MAE_s, as defined in Eqs. (2a) and (2b), respectively:

$$\mathrm{MSE_s} = \frac{1}{n}\sum_{i=1}^{n}(p_i - s_i)^2, \tag{2a}$$

$$\mathrm{MAE_s} = \frac{1}{n}\sum_{i=1}^{n}|p_i - s_i|. \tag{2b}$$

The signal term (s_i) excludes the error, so the random behaviour of the error does not influence MSE_s and MAE_s. In simulated series the values of s_i and w_i would, of course, be known. The values of MSE_s and MAE_s will not, however, be comparable with MSE_a and MAE_a: in fact, they will be smaller.

Given the inherent uncertainties in making predictions, it can be asserted that the noise term has a definitive influence in real-life forecasting situations. Accordingly, when simulated data are used in an experimental context, it becomes especially important to reflect the noise term in the calculation and interpretation of accuracy statistics. Pursuing this perspective, it is shown in Appendix that the expected values for MSE_a and MAE_a can be obtained in the form of Eqs. (3a) and (3b):

$$E(\mathrm{MSE_a}) = \mathrm{MSE_s} + 1, \tag{3a}$$

$$E(\mathrm{MAE_a}) = \mathrm{MAE_s} - \frac{1}{n}\sum_{i=1}^{n}|p_i - s_i|\Phi(-|p_i - s_i|) + \frac{1}{n}\sqrt{2/\pi}\sum_{i=1}^{n}\mathrm{e}^{-(p_i - s_i)^2/2}. \tag{3b}$$

In Eq. (3b), Φ denotes the cumulative distribution function of the Standard Normal distribution. Eqs. (3a) and (3b) illustrate that the expected values of MSE_a and MAE_a are generally not equal to MSE_s and MAE_s, respectively. In the case of MSE_a in Eq. (3a), the adjustment only requires the addition of a unity term. MSE_s reflects the part of the MSE under the control of the forecaster, and the unity term reflects the uncontrollable part. In other words, even when predictions are made on a precisely recognised signal (i.e., $MSE_s = 0$), the expected MAE_a has a value of unity due to the uncontrollable noise component, i.e., MSE_s gives a downward bias to the estimate of the expected MSE_a.

The derivation is more complex for the MAE_a. The second term on the right-hand side of Eq. (3b) has a maximum value of zero (when the $|p_i - s_i|$ values are all either zero or infinity) and a minimum value of -0.34 (when the $|p_i - s_i|$ values are all 0.75). The first two terms taken together reflect aspects of MAE_a under the control of the forecaster. The last term on the right-hand side of Eq. (3b) is a term that is not directly under the control of the forecaster. This term has its largest value, approximately 0.8, when the difference between each predicted and signal value is zero. That is, even when predictions are made on a precisely recognised signal, the expected MAE_a has a value of 0.8 reflecting the uncontrollable noise. This term tends to zero, however, when the differences between the predicted and signal values increase. In Appendix A, it is shown that $0 \leqslant E(\mathrm{MAE_a}) - \mathrm{MSE_s} < 0.8$, with equality occurring where $|p_i - s_i| = 0$ for each i. Thus MSE_s also gives a downward bias to the estimate of the expected MAE_a. In short, it can be concluded that both MSE_s and MAE_s underestimate the true error since they are based on signal values alone. Eqs. (3a) and (3b) illustrate, however, that corrections can be made to obtain expected values (viz., $E(\mathrm{MAE_a})$ and $E(\mathrm{MSE_s})$) that also incorporate the noise, hence yielding more representative measures of forecasting accuracy for the series under consideration.

4. An application of the framework

The application of the above framework is illustrated using a set of judgemental predictions on a cross section based task designed to simulate monthly currency series. The judgemental predictions were obtained from ten members of the EURO-Working Group on Financial Modelling. The sample comprised academics and practitioners from a number of different countries. All individuals who took part in the inquiry had considerable expertise in the field of finance and working knowledge of currency markets.

Simulated data for the time paths of 36 series were presented numerically and graphically to the participants. The participants were not told how the data were formulated, only that they were obtained through a statistical procedure to simulate currency series. These series were presented for a 60-month period (months were numbered from 1 to 60) and indexed with the initial value in month 0 set at 1000. The data were based on six randomly generated series from a Standard Normal distribution. Cumulative values of the series were then formed with a starting value of 1000. Constant drifts of varying size were added to the six resulting series. Specifically, these drifts could be categorized as:

(i) zero – which gave a probability of 0.5 for increase/decrease;
(ii) mild (±.2533) – which gave a probability of 0.6 for increase/decrease;
(iii) medium (±.5244) – which gave a probability of 0.7 for increase/decrease;
(iv) strong (±.8416) – which gave a probability of 0.8 for increase/decrease;
(v) very strong (±1.2816) – which gave a probability of 0.9 for increase/decrease;
(vi) dominant (±3.0902) – which gave a probability of almost 0.999 for increase/decrease.

For each series, three positive and three negative forms of drift were used. This resulted in 36 series, of which six were random walks and 30 were random walks with varying degrees of constant drift (15 positive and 15 negative). The data were rounded to the nearest whole number and presented to the subjects in a random fashion.

Simulated random walk series with varying degrees of drift were chosen for two reasons. Firstly, random walk series with varying degrees of drift reasonably approximate monthly financial time-series behaviour. For example, Pollock and Wilkie (1992) found on a time-series probabilistic forecasting task with actual monthly currency series that the random walk with drift model performed relatively well in comparison to more complex models and much better than the time-series extrapolations of a group of professional forecasters. Secondly, these series contain only one signal (drift) that individuals need to identify, easing the cognitive load on the forecasters.

The series were presented numerically and graphically to the participants together with an instruction sheet and a booklet to indicate predictions, which they were requested to complete independently of other subjects. Given a 60-month period for each series, the participants were required to make judgemental point predictions for month 61 of each of the 36 series. The subjects were requested to make their predictions independently of the other subjects and at their own pace and convenience.

To compare predictions with the optimal, it is necessary to obtain the theoretical expected point values for the one-month-ahead forecasts (i.e., for month 61). Denoting the exchange rate at time t as y_t the expected one-step-ahead change in the exchange rate {i.e., $E(\Delta y_{t+1})$} can be viewed as the signal term (μ) and is given in Eq. (4):

$$E(\Delta y_{t+1}) = \mu. \tag{4}$$

The actual change (Δy_{t+1}) consists, however, of the signal (μ) and noise (ε_{t+1}), as given in Eq. (5):

$$\Delta y_{t+1} = \mu + \varepsilon_{t+1}. \tag{5}$$

Using the theoretically attained point values outlined above, accuracy measures were computed for the judgemental point forecasts provided by the participants. Table 1 shows the following measures for each of the 10 participants:

Table 1
Results from the performance analysis

	Accuracy measure					
Subject	$E(\text{MAE}_a)$	MAE_s	MAE_a	$E(\text{MSE}_a)$	MSE_s	MSE_a
1	0.811(1)	0.129(1)	0.937(1)	1.034(1)	0.034(1)	1.507(1)
2	1.050(6)	0.574(5)	1.151(6)	2.107(9)	1.107(9)	2.299(9)
3	1.130(9)	0.781(8)	1.062(2)	2.000(8)	1.000(8)	2.174(5)
4	0.921(2)	0.420(3)	1.157(8)	1.336(2)	0.336(2)	2.104(3)
5	0.974(4)	0.511(4)	1.135(5)	1.582(4)	0.582(4)	2.251(7)
6	1.013(5)	0.628(6)	1.172(9)	1.607(5)	0.607(5)	2.265(8)
7	1.086(7)	0.770(7)	1.293(10)	1.821(6)	0.821(6)	2.619(10)
8	1.102(8)	0.827(9)	1.126(4)	1.845(7)	0.845(7)	2.083(2)
9	1.358(10)	1.069(10)	1.087(3)	3.178(10)	2.178(10)	2.109(4)
10	0.947(3)	0.397(2)	1.151(7)	1.533(3)	0.533(3)	2.239(6)
Mean	1.039	0.611	1.127	1.804	0.804	2.165
Random walk	1.194	0.810	1.237	2.412	1.412	2.413

Ranks in descending order of performance are given in brackets.

(i) $E(\text{MAE}_a)$ and $E(\text{MSE}_a)$,
(ii) MAE_s and MSE_s,
(iii) MAE_a and MSE_a.

For each measure the rank orderings (1–10) are given in brackets. For comparison the corresponding measures for the random walk forecaster are also tabulated. The random walk forecaster is a hypothetical subject who always gives the predicted change as zero. Subjects would, generally, be expected to have performance measures that were below those of the random walk forecaster.

Table 2 shows the rank correlation matrix for the six performance measures across the 10 participants, with values significantly different from zero highlighted (given in boldface).

The results show considerable diversity between the participants reflecting a high degree of heterogeneity of the subjects in their judgemental point predictions. The results illustrate that the signal-only statistic values (i.e., MSE_s) provide a similar ordering in performance to $E(\text{MAE}_a)$. The ordering for the MSE_s is, of course, identical to that for $E(\text{MAE}_a)$. The values for the MAE_s and MSE_s are, however, much smaller than the $E(\text{MAE}_a)$ and $E(\text{MSE}_a)$, respectively (as reflected by the values in Table 1). These findings may be viewed as suggesting that using the signal alone (and neglecting the noise) may lead to unrealistic comparisons of performance when simulated versus actual data are employed in investigations of forecasting accuracy.

The results also illustrate that including the values of the random error term in the calculation of accuracy statistics (as done via MAE_a and MSE_a) may yield noticeable changes in performance ordering as compared to the rankings given by $E(\text{MAE}_a)$ and $E(\text{MSE}_a)$. This is reflected in the correlations of Table 2

Table 2
Spearman rank correlations

	$E(\text{MAE}_a)$	MAE_s	MAE_a	$E(\text{MSE}_a)$	MSE_s
MAE_s	**0.964****				
MAE_a	−0.164	−0.091			
$E(\text{MSE}_a)$	**0.927****	**0.855****	−0.139		
MSE_s	**0.927****	**0.855****	−0.139	(1.000)	
MSE_a	0.188	0.103	**0.673***	0.309	0.309

Significant values: $^{*}p < 0.05$; $^{**}p < 0.01$.

and the rankings presented in Table 1. For example, the performance of Subject 9 falls from third place on MAE_a to tenth place on the $E(MAE_a)$ and from fourth place on the MAE_a to tenth on the $E(MAE_a)$. It is conceivable that these differences would be much greater in other situations where the simulated data display extreme chance error fluctuations. As error behaviour is inherently unpredictable, it may be expedient to exclude such random error from formal performance assessments, as proposed by the $E(MAE_a)$ and $E(MSE_a)$ measures.

5. Conclusion

Exchange rates are viewed as indispensable inputs to the decision-making processes of firms involved in international trade and markets, hence accentuating the need for accurate forecasts (Stockman, 1987; Gerlow and Irwin, 1991). This paper has focused on evaluating the accuracy of judgemental currency forecasts given by financial experts. Simulated series were used to delineate the effects of various characteristics of time series on predictive performance. It has been illustrated that corrections can be made to obtain expected values for the MAE and the MSE measures that incorporate the noise in simulated series which follow a difference-stationary framework and where the error terms are Normally distributed. The resulting performance statistics were found to have values comparable with the statistics based on actual (non-simulated) data, with the additional advantage of not being influenced by atypical values caused by random variation.

The main conclusion of the paper is that when using simulated currency series it is advisable to use the expected values of the MSE_a and MAE_a {i.e., $E(MSE_a)$ and $E(MAE_a)$} formulations derived in the paper. The resulting values can then be compared with the hypothetical random walk forecaster. The formulation also allows separation from the $E(MSE_a)$ and $E(MAE_a)$ of the part under the control of the forecaster from the part outside his/her control. This can be used to give an indication of an individual's ability to separate the signal from the noise in a series. It has also been shown, however, that if the main concern is with the ranking of the performance then the mean absolute deviation based on the signal values (i.e., MAE_s) provides a reasonable approximation of the $E(MAE_a)$ rankings and the mean square error based on the signal values (i.e., MSE_s) provides the same ordering as the $E(MSE_a)$.

The work has provided an initial attempt to apply the proposed measures of accuracy to judgemental forecasts given for simulated currency series. Further applications may involve many financial and economic series that follow difference-stationarity. In addition, the analysis only needs minor modifications to deal with trend-stationary processes and can easily be extended to handle other error generation that is non-Normal, for example, noise generated by a uniform distribution. The importance of the work hinges on expanding its applicability so as to build a framework of tested relationships for a variety of series. Consequently, future extensions promise to entail investigations of judgemental forecasting via a plethora of critical variables such as interest rates, earnings, etc.. Hence, even though the current findings may constitute a preliminary step in exploring the proposed measures, profound implications of this research for both the providers and users of financial and economic forecasts become apparent when viewed in this wider context.

The procedure outlined has important implications for analysing time-series extrapolative judgement in currency forecasting practice. Given that the identification of trend is crucial to the chartists' extrapolations, which, in turn, play a central role in the market positions assumed by the financial agents, the proposed framework can be utilized to assess forecasters' skills in accurately recognizing trends in simulated currency series. Accordingly, these measures may support the decision processes involved in selecting financial forecasters and conducting performance-appraisals. The proposed measures could also be used as effective feedback and training tools (Benson and Önkal, 1992; Bolger and Wright, 1994; Önkal and Muradoglu, 1995; Harvey and Bolger, 1996).

The framework can also be employed to explore potential biases in judgemental forecasting that may stem from different series-specific characteristics (e.g., noise and trend). Results from such analyses with simulated data may also help to identify conditions amenable to enhanced judgemental revisions of statistical forecasts. This issue is particularly critical, since it has repeatedly been argued that, even when quantitative techniques are used in forecasting practice, the resulting predictions are often combined with human judgement, yielding final forecasts which are a mixture of both quantitative and subjective analyses (Lim and O'Connor, 1996; Winklhofer et al., 1996).

A related direction for future research involves combining exchange rate forecasts (MacDonald and Marsh, 1994). It has been asserted that the relative accuracy of composite forecasts versus individual forecasts demands further work (Guerard, 1989), and the measures suggested by current research could provide a starting point for such evaluations.

The study has focused on judgemental point forecasts only. This emphasis is in line with previous financial forecasting research (Önkal-Atay, 1998). However, it may be argued that the predictions presented in point format are limited in their information content. In particular, interval and/or probabilistic format may be viewed as providing more detailed information to the users of financial forecasts with regard to the forecaster's uncertainties (Muradoglu and Önkal, 1994; Önkal and Muradoglu, 1994–1996). As emphasized by Bunn and Wright (1991), such communication of uncertainty is of paramount importance for the preparers and users of forecasts. Furthermore, users may focus on accuracy dimensions that are different than the aspects stressed by researchers (Yates et al., 1996). In summary, there is a definitive need for future research on currency forecasting to focus on the user aspect and to explore issues of forecast communication and evaluation from a broader perspective.

Appendix A

A.1. The expected value of MSE_a

The $\mathrm{MSE_a}$ is defined in Eq. (A.1):

$$\begin{aligned}\mathrm{MSE_a} &= \frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\{p_i - s_i - (a_i - s_i)\}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(p_i - s_i)^2 - \frac{2}{n}\sum_{i=1}^{n}(p_i - s_i)(a_i - s_i) + \frac{1}{n}\sum_{i=1}^{n}(a_i - s_i)^2. \end{aligned} \tag{A.1}$$

Given that the first term on the right-hand side of Eq. (A.1) is constant (i.e., p_i is considered to be fixed), it can be easily shown that the expected value of $\mathrm{MSE_a}$ is in the form of Eq. (A.2):

$$E(\mathrm{MSE_a}) = \mathrm{MSE_s} + \frac{1}{n}\sum_{i=1}^{n}E(w_i^2) = \mathrm{MSE_s} + 1, \tag{A.2}$$

where

$$\mathrm{MSE_s} = \frac{1}{n}\sum_{i=1}^{n}(p_i - s_i)^2$$

and

$$w_i \equiv a_i - s_i\ ,\ i = 1,\ 2, \ldots, n.$$

Note that to derive expression (A.2), it has been assumed that the noise term (w_i) is independent of both the signal {i.e., $E(s_i w_i) = 0$} and the predicted value {i.e., $E(p_i w_i) = 0$}. In addition, as w_i is Normally distributed with zero mean and unit variance, w_i^2 follows a chi-squared distribution with 1 degree of freedom so that $E(w_i^2) = 1$.

A.2. The expected value of MAE_a

The derivation of the expected value of the MAE_a is more complex. It is first necessary to obtain $E\{|p_i - s_i - w_i|\}$. If ϕ and Φ respectively denote the probability density function and cumulative distribution function of the Standard Normal distribution, then:

$$\begin{aligned} E\{|p_i - s_i - w_i|\} &= \int_{-\infty}^{p_i - s_i} (p_i - s_i - w_i)\phi(w_i)\, \mathrm{d}w_i - \int_{p_i - s_i}^{\infty} (p_i - s_i - w_i)\phi(w_i)\, \mathrm{d}w_i \\ &= \int_{-\infty}^{\infty} (p_i - s_i - w_i)\phi(w_i)\, \mathrm{d}w_i - 2\int_{p_i - s_i}^{\infty} (p_i - s_i - w_i)\phi(w_i)\, \mathrm{d}w_i \\ &= p_i - s_i - 2(p_i - s_i)\left[1 - \Phi(p_i - s_i) + \sqrt{2/\pi}\, \mathrm{e}^{-(p_i - s_i)^2/2}\right] \end{aligned}$$

since

$$\int_{-\infty}^{\infty} \phi(w_i)\, \mathrm{d}w_i = 1$$

and

$$\int_{-\infty}^{\infty} w_i \phi(w_i)\, \mathrm{d}w_i = E(w_i) = 0.$$

Noting that $\Phi(x) = 1 - \Phi(-x)$ for all x, it follows that

$$E\{|p_i - s_i - w_i|\} = |p_i - s_i| - 2|p_i - s_i|\Phi(-|p_i - s_i|)\sqrt{2/\pi}\, \mathrm{e}^{-(p_i - s_i)^2/2}. \tag{A.3}$$

The expected value of the MAE_a is then obtained by averaging Eq. (A.3) over $i = 1, 2, \ldots, n$, giving equation Eq. (A.4):

$$E(\mathrm{MAE_a}) = \mathrm{MAE_s} - \frac{2}{n}\sum_{i=1}^{n} |p_i - s_i|(-|p_i - s_i|) + \frac{1}{n}\sqrt{2/\pi}\sum_{i=1}^{n} \mathrm{e}^{-(p_i - s_i)^2/2}, \tag{A.4}$$

where

$$\mathrm{MAE_s} = \frac{1}{n}\sum_{i=1}^{n} |p_i - s_i|.$$

To investigate the behaviour of $E(\mathrm{MSE_a}) - \mathrm{MSE_s}$, write Eq. (A.4) in the form

$$E(\mathrm{MAE_a}) = \mathrm{MAE_s} + \frac{1}{n}\sum_{i=1}^{n}\{2\phi(|p_i - s_i|) - 2|p_i - s_i|\Phi(-|p_i - s_i|)\}.$$

Consider the function

$$g(x) = 2\phi(x) - 2x\Phi(-x), \quad x > 0.$$

Clearly

$$g(0) = 2\phi(0) = 0.798 \quad \text{and} \quad g(\infty) = 0.$$

Now

$$g'(x) = 2\phi'(x) - 2x\Phi(-x) + 2x\phi(-x) = -2\Phi(-x),$$

since $\phi'(x) = -x\phi(-x)$ and $\phi(x) = \phi(-x) < 0$ for all x.

Hence $g(x)$ is monotonically decreasing from about 0.8 to 0, i.e., $0 \leqslant E(\mathrm{MAE_a}) - \mathrm{MSE_s} < 0.8$, with equality occurring when $|p_i - s_i| = 0$ for each i, (i.e., when the predicted values coincide exactly with the signal values).

References

Allen, H.L., Taylor, M.P., 1989. Charts and fundamentals in the foreign exchange market. Bank of England Discussion Paper, No. 40.

Allen, H.L., Taylor, M.P., 1990. Charts, noise and fundamentals in the foreign exchange market. Economic Journal 100, 49–59.

Ang, S., O'Connor, M., 1991. The effect of group interaction processes on performance in time series extrapolation. International Journal of Forecasting 7, 141–149.

Armstrong, J.S., Collopy, F., 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting 8, 69–80.

Armstrong, J.S., Fildes, R., 1995. Correspondence on the selection of error measures for comparisons among forecasting methods. Journal of Forecasting 14, 67–71.

Bahmani-Oskooee, M., Ltaifa, N., 1992. Effects of exchange rate risk on exports: Cross-country analysis. World Development 20, 1173–1181.

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

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

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

Bolger, F., Harvey, N., 1995b. Rationality of expectations in judgemental extrapolations of time series. In: Caverni, J.P., Bar-Hillel, M., Barron, F.H., Jungermann, H. (Eds.), Contributions to Decision Making – 1, Elsevier, Amsterdam.

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

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

Boothe, P., Glassman, D., 1987b. The statistical distribution of exchange rates: Empirical evidence and economic implications. Journal of International Economics 2, 297–319.

Bunn, D.W., Wright, G., 1991. Interaction of judgmental and statistical forecasting methods: Issues and analysis. Management Science 37, 501–518.

Chinn, M.D., Meese, R.A., 1995. Banking on currency forecasts: How predictable is change in money? Journal of International Economics 38, 161–178.

Chowdhry, B., 1995. Corporate hedging of exchange risk when foreign currency cash flow is uncertain. Management Science 41, 1083–1090.

Clements, M.P., Hendry, D.F., 1993. On the limitations of comparing mean square forecast errors. Journal of Forecasting 12, 617–637.

Dejong, D.N., Whiteman, C.H., 1994. The forecasting attributes of trend- and difference-stationary representations for macroeconomic time series. Journal of Forecasting 13, 279–297.
Dornbusch, R., 1976. Expectations and exchange rate dynamics. Journal of Political Economy 84, 1161–1176.
Fildes, R., 1992. The evaluation of extrapolative forecasting methods. International Journal of Forecasting 8, 88–98.
Frankel, J.A., 1979. On the mark: A theory of floating exchange rates based on real interest differentials. American Economic Review 69, 610–622.
Frankel, J.A., 1983. Monetary and portfolio balance models of exchange rate determination. In: Bhandari, J.S., Putnam, B.H. (Eds.), Economic Interdependence and Flexible Exchange Rates. MIT Press, Cambridge, MA.
Frankel, J.A., Froot, K.A., 1990. Chartists, fundamentalists and the demand for dollars. In: Courakis, A., Taylor, M. (Eds.), Policy Issues for Interdependent Economies. Clarendon Press, Oxford.
Friedman, D., Vandersteel, S., 1982. Short run fluctuations in foreign exchange rates: Evidence from the data, 1973–79. Journal of International Economics 13, 171–186.
Gerlow, M.E., Irwin, S.H., 1991. The performance of exchange rate forecasting models: An economic evaluation. Applied Economics 23, 133–142.
Goldberg, M.D., Frydman, R., 1996. Imperfect knowledge and behaviour in the foreign exchange market. The Economic Journal 106, 869–893.
Goodwin, P., Wright, G., 1993. Improving judgemental time series forecasting: A review of the guidance provided by research. International Journal of Forecasting 9, 147–161.
Guerard, J.B., 1989. Composite model building for foreign exchange rates. Journal of Forecasting 8, 315–329.
Harvey, N., Bolger, F., 1996. Graphs versus tables: Effects of data presentation format on judgemental forecasting. International Journal of Forecasting 12, 119–137.
Keren, G., 1991. Calibration and probability judgements: Conceptual and methodological issues. Acta Psychologica 77, 213–217.
Kouri, P.J.K., 1976. The exchange rate and the balance of payments in the short run and the long run: A monetary approach. Scandinavian Journal of Economics 78, 280–304.
Kuan, C.-M., Liu, T., 1995. Forecasting exchange rates using feedforward and recurrent neural networks. Journal of Applied Econometrics 10, 347–364.
Lastrapes, W.D., 1992. Sources of fluctuations in real and nominal exchange rates. Review of Economics and Statistics 74, 530–539.
Lawrence, M., O'Connor, M., 1992. Exploring judgemental forecasting. International Journal of Forecasting 8, 15–26.
Lawrence, M., O'Connor, M., 1993. Scale, variability and the calibration of prediction intervals. Organizational Behavior and Human Decision Processes 56, 441–458.
Lessard, D.R., Lightstone, J.B., 1986. Volatile exchange rates can put operations at risk. Harvard Business Review 64, 107–114.
Lim, J.S., O'Connor, M., 1995. Judgemental adjustment of initial forecasts: Its effectiveness and biases. Journal of Behavioral Decision Making 8, 149–168.
Lim, J.S., O'Connor, M., 1996. Judgmental forecasting with time series and causal information. International Journal of Forecasting 12, 139–153.
Liu, T.-R., Gerlow, M.E., Irwin, S.H., 1994. The performance of alternative VAR models in forecasting exchange rates. International Journal of Forecasting 10, 419–433.
MacDonald, R., Marsh, I.W., 1994. Combining exchange rate forecasts: What is the optimal consensus measure? Journal of Forecasting 13, 313–332.
Mathews, B.P., Diamantopoulos, A., 1994. Towards a taxonomy of forecast error measures. Journal of Forecasting 13, 409–416.
Meese, R., Rogoff, K., 1983. The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? In: Frenkel, J. (Ed.), Exchange Rates and International Macroeconomics. University of Chicago Press, Chicago.
Muradoglu, G., Önkal, D., 1994. An exploratory analysis of the portfolio managers' probabilistic forecasts of stock prices. Journal of Forecasting 13, 565–578.
Mussa, M., 1976. The exchange rate, the balance of payments, and monetary and fiscal policy under a regime of controlled floating. Scandinavian Journal of Economics 78, 229–248.
Nachane, D.M., Ray, D., 1993. Modelling exchange rate dynamics: New perspectives from the frequency domain. Journal of Forecasting 12, 379–394.
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 judgemental confidence intervals in time series forecasting. Journal of Forecasting 8, 141–155.
O'Connor, M., Lawrence, M., 1992. Time series characteristics and the widths of judgmental confidence intervals. International Journal of Forecasting 7, 413–420.
O'Connor, M., Remus, W., Griggs, K., 1993. Judgemental forecasting in times of change. Journal of Forecasting 9, 163–172.
Önkal, D., Muradoglu, G., 1994. Evaluating probabilistic forecasts of stock prices in a developing stock market. European Journal of Operational Research 74, 350–358.

Önkal, D., Muradoglu, G., 1995. Effects of feedback on probabilistic forecasts of stock prices. International Journal of Forecasting 11, 307–319.
Önkal, D., Muradoglu, G., 1996. Effects of task format on probabilistic forecasting of stock prices. International Journal of Forecasting 12, 9–24.
Önkal-Atay, D., 1998. Financial forecasting with judgment. In: Wright, G., Goodwin, P. (Eds.), Forecasting with Judgment. Wiley, Chichester, UK.
Perron, P., 1988. Trends and random walks in macroeconomic time series. Journal of Economic Dynamics and Control 12, 297–332.
Pilbeam, K., 1995. Exchange rate models and exchange rate expectations: An empirical investigation. Applied Economics 27, 1009–1015.
Pollock, A.C., Wilkie, M.E., 1992. Currency forecasting: human judgment or models. VBA-Journal 3, 21–29.
Pollock, A.C., Wilkie, M.E., 1993. Directional judgemental financial forecasting: trends and random walks. In: Flavell, R. (Ed.), Modelling Reality and Personal Modelling. Physica, Heidelberg.
Pollock, A.C., Wilkie, M.E., 1996. The quality of bank forecasts: The dollar-pound exchange rate, 1990–1993. European Journal of Operational Research 91, 304–306.
Pollock, A.C., Wilkie, M.E., Pollock, J.F., 1996. A framework for examining the quality of bank currency forecasts: the yen–DM, 1990–1993. Paper presented at the International Symposium of Forecasting, Istanbul, Turkey, June.
Remus, W., O'Connor, M., Griggs, K., 1995. Does reliable information improve the accuracy of judgmental forecasts? International Journal of Forecasting 11, 285–293.
Steil, B., 1993. Corporate foreign exchange risk management: A study in decision making under uncertainty. Journal of Behavioral Decision Making 6, 1–31.
Stockman, A.C., 1987. Economic theory and exchange rate forecasts. International Journal of Forecasting 3, 3–15.
Taylor, M.P., Allen, H.L., 1992. The use of technical analysis in the foreign exchange market. Journal of International Money and Finance 11, 304–314.
Tversky, A., Kahneman, D., 1973. Availability: A heuristic for judging frequency and probability. Cognitive Psychology 5, 207–232.
van Hoek, T.H., 1992. Explaining mark/dollar and yen/dollar exchange rates the 1980s. Economics Letters 38, 467–472.
Webby, R., O'Connor, M., 1996. Judgemental and statistical time series forecasting: a review of the literature. International Journal of Forecasting 12, 91–118.
Westerfield, J.M., 1977. An examination of foreign exchange risk under fixed and flexible regimes. Journal of International Economics 7, 181–200.
Wilkie, M.E., Pollock, A.C., 1994. Currency forecasting: An investigation into probability judgement accuracy. In: Peccati, L., Virén, M. (Eds.), Financial Modelling. Physica, Heidelberg.
Wilkie, M.E., Pollock, A.C., 1996. An application of probability judgement accuracy measures to currency forecasting. International Journal of Forecasting 12, 25–40.
Winklhofer, H., Diamantopoulos, A., Witt, S.F., 1996. Forecasting practice: A review of the empirical literature and an agenda for future research. International Journal of Forecasting 12, 193–221.
Yates, J.F., Price, P.C., Lee, J.W., Ramirez, J., 1996. Good probabilistic forecasters: The 'consumer's' perspective. International Journal of Forecasting 12, 41–56.