Garratt, A. & Lee, K.

Investing under model uncertainty: decision based evaluation of exchange rate forecasts in the US, UK and Japan

Journal of International Money & Finance - 29(3), pp.403-422 (2010)

Investing Under Model Uncertainty: Decision Based Evaluation of Exchange Rate Forecasts in the US, UK and Japan

by

Anthony Garratt*† and Kevin Lee††

†Birkbeck, University of London, Malet Street, London,WC1E 7HX, UK

††University of Leicester, University Road, Leicester,LE1 7RH, UK

Abstract

We evaluate the forecast performance of a range of theory-based and atheoretical models explaining exchange rates in the US, UK and Japan. A decision-making environment is fully described for an investor who optimally allocates portfolio shares to domestic and foreign assets. Methods necessary to compute and use forecasts in this context are proposed, including the means of combining density forecasts to deal with model uncertainty. An out-of-sample forecast evaluation exercise is described using both statistical criteria and decision-based criteria. The theory-based models are found to perform relatively well when their forecasts are judged by their economic value.

Keywords: Exchange rates, Investment Strategies, Forecast Evaluation, Model Averaging

JEL Classifications: C32, C53, E17

*Corresponding author. Telephone: +44(0) 207 631 6410, Fax: +44(0) 207 631 6416, E-mail: a.garratt@bbk.ac.uk.

1 Introduction

Recent years have seen a growing interest in the decision-based approach to the evaluation and comparison of forecasts. Here, forecast accuracy is judged according to its economic value to an individual given an explicitly defined decision-making context. This reflects the recognition that models should be judged according to their purpose and that the statistical criteria used to evaluate models, typically based solely around point forecasts and measured using mean squared forecasting error (MSE), are unlikely to provide information on the economic value of their forecasts.[1] The preponderance of studies employing the decision-based approach to forecast evaluation are in the area of applied finance where the decision-making context is relatively straightforward to describe.[2] But they remain relatively rare even here and model evaluation in the context of the analysis of exchange rates, for example, still focuses primarily on statistical criteria.[3].

In this paper, we consider an illustrative investment scenario where an investor uses exchange rate forecasts in choosing the proportion of her portfolio to be invested in domestic and foreign assets. The forecasts can be based on one of a variety of models or on an aggregation of model forecasts and we evaluate the forecast performance of the various models in the context of our specified investment scenario. The problem studied is similar to those studied by Barberis (2002), West *et al.*(1993) and Abhyankar *et al.* (2005). However, in contrast to West *et al.*(1993), the focus here is on the conditional mean of the exchange rate (as opposed the variance). Also, given there is rarely consensus

[1]See Granger and Pesaran (2000a,b) for an overview of this discussion.

[2]See, for example, , Boothe (1983, 1987), Leitch and Tanner (1991), West *et al.* (1993), Pesaran and Timmerman (1995), Kandel and Stambaugh (1996), Barberis (2000) and Abhyanker *et al.* (2005).

[3]See for example, Meese and Rogoff (1983), Mark (1995), Mark and Sul (2001), Berkowitz and Giorganni (2001), Faust *et al.*(2003) Clarida *et al.*(2004), Killian and Taylor (2003) and Cheung *et al.* (2005) among others.

on the appropriate model(s) to be employed in these contexts, this paper extends the analysis of Abhyankar *et al.* (2005) by focusing on model uncertainty through the use of 'Bayesian-style' model averaging methods. Wright (2003) also employs model averaging methods for a range of exchange rate models, but where forecasts are evaluated using only statistical criteria.

The methods employed here are based on simulation techniques and are straightforward to implement. We apply the methods to a range of theory-based and atheoretical models explaining exchange rates in the US, UK and Japan. The exercise involves calculating predictive density forecasts, combining density forecasts to allow for model averaging, and identifying and implementing the appropriate decision-based criterion with which to judge the models. An out-of-sample forecast evaluation exercise is conducted using both statistical criteria and decision-based criteria. It demonstrates that the conclusions drawn on the basis of the alternative criteria are quite different. We find that atheoretical models, and model averages, perform relatively well when judged by statistical criteria, but that theory-based models typically dominate the atheoretical models when using economic criteria.

The plan of the paper is as follows. In section 2 we describe the investment decision and the methods required to use and evaluate forecasts from a number of individual models and/or from a model average. Section 3 outlines the candidate set of models for the exchange rate on which the investment decision might be made. Section 4 describes the estimation of the models using US, UK and Japanese data for the period 1981m1-2002m6 (the out of sample evaluation period extends to 2006m6) and evaluates their forecasting performance using statistical criteria. Section 5 describes the decision-based forecast evaluation, judging the models' performance according to the utility derived from the associated investment strategies. Section 6 concludes.

2 The Investment Decision

The decision problem we consider is one in which an investor, with an investment horizon H, chooses at time T what proportion of her portfolio to allocate to a foreign asset (ω) and how much to a domestic asset $(1-\omega)$.[4] The set up is deliberately simple, and restricts attention to a 'buy-and-hold' strategy, where the investor's allocation made at time T applies throughout the decision (forecast) period $T+1$ to $T+H$. We assume that identical domestic and foreign assets are available, both maturing in each period, and their returns measured in local currency at time t are r_t and r_t^* respectively. If we normalise by setting wealth at T equal to unity, $W_T = 1$, then the end-of-decision-period wealth can be expressed as:

$$W_{T+H}(\omega) = (1-\omega)\exp(\sum_{h=0}^{H} r_{T+h}) + \omega \exp(\sum_{h=0}^{H} r_{T+h}^* + \Delta_H\, e_{T+H}), \tag{1}$$

where $e_t = \log(E_t)$ and E_t denotes the spot (end-of-period) nominal bilateral exchange rate describing the domestic price of the foreign currency, and where we use the approximations $\log(1+r) \approx r$ and $\log(E_{T+H}/E_T) \approx e_{T+H} - e_T$. Throughout the paper, we assume the investor chooses the fraction of the portfolio to invest in safe assets at home and abroad, with no dynamic rebalancing, and the returns r_{T+h} and r_{T+h}^*, $h = 1, ..., H$, are assumed known with certainty at time T. Thus we assume that uncertainty on end-of-period wealth arises only from potential movements in the exchange rate. Even in this very straightforward case, however, it will not be possible to obtain a point forecast of W_{T+H}, or make decisions on ω, simply using point forecasts of the e_{T+h}, $h = 1, .., H$. Rather, the non-linearity of (1) means that the investor will need to evaluate the entire

[4] The investment decision problem is similar to that in Kandel and Stambaugh (1996), Barberis (2000) and Abhyankar *et al.* (2005), among others.

joint probability distribution of the forecast values of e_{T+h}, $h = 1, .., H$, to evaluate $\mathrm{E}(W_{T+H} \mid \Omega_T)$.

Extending the exercise to accommodate risk aversion in the investor's decision making, we might assume that the investor derives utility from W_{T+H} according to the standard constant relative risk aversion (CRRA) power utility function,

$$\nu(W_{T+H}) = \frac{W_{T+H}^{1-A}}{1-A}, \tag{2}$$

where A is the coefficient of risk aversion.[5] In this case, the investor's problem at time T can be written as

$$\max_{\omega} \left\{ \mathrm{E} \left[\nu(W_{T+H}(\omega)) \mid \Omega_T \right] \right\}.$$

Given the additional non-linearities involved in (2), evaluating expected utility will again require the investor to use the entire joint probability distribution of the forecast values of e_{T+h} $(h = 1, .., H)$. The investment decision similarly relies on on these joint distributions as it involves first calculating the expected utility for any given portfolio share, and then identifying the optimal portfolio share as that which maximises the expected utility across all portfolio shares.

2.1 The Probability Density Function of the Forecast Values

The key to decision-making here is the probability density function of the forecast values of the exchange rate over the decision horizon. Denoting $\mathbf{z}_t = (z_{1t}, z_{2t}, ..., z_{nt})'$ to be an $n \times 1$ vector of variables of interest (including at least e_t here) and $\mathbf{Z}_{1,T} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T)'$ to be the available observations at the end of period T, we are interested in the probability density function of $\mathbf{Z}_{T+1,T+H} = (\mathbf{z}_{T+1}, \mathbf{z}_{T+2}, ..., \mathbf{z}_{T+H})'$ conditional on $\mathbf{Z}_{1,T}$; that is

[5] Campbell and Viceria (2002) argue in favour of power utility functions as they have the attractive property that absolute risk aversion declines with wealth whilst relative risk aversion remains constant.

$\Pr(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T})$, sometimes termed the "predictive density function". The decision problem can then be written as

$$\max_{\omega} \left\{ \int \nu(W_{T+H}(\omega)) \Pr\left(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T}\right) d\mathbf{Z}_{T+1,T+H} \right\}. \tag{3}$$

The form of the density function $\Pr(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T})$ depends on the types of uncertainty that surround the forecast and the approach taken to characterising and estimating the function. The types of uncertainty that might influence the forecasts include: the *stochastic uncertainty* associated with the innovations impacting on a model; the *parameter uncertainty* associated with estimated model parameters; and the *model uncertainty* surrounding the choice of model itself. The first two of these are routinely taken into account in forecasting, but model uncertainty is less frequently considered. This is despite the fact that this latter source of uncertainty is potentially more important in decision-making if there is little consensus on how the variables are determined (as is the case with international investment decisions, for example, where there is little agreement on the processes underlying exchange rate determination).

The approach taken to characterising and estimating the density function varies according to judgements on the role of economic theory in econometric modelling and pragmatic decisions on the use of prior knowledge. Draper (1995) and Hoeting *et al.* (1999), for example, describe the "Bayesian Model Averaging" approach which elegantly accommodates all three forms of uncertainty described above in a comprehensive, fully Bayesian approach to estimating $\Pr(\mathbf{Z}_{T+1,\mathbf{T}+H} \mid \mathbf{Z}_{1,T})$. At the same time, as is well known, there may be practical difficulties involved in the choice of priors for models, or in the choice of priors for the parameters of any given model, in the context of forecasting that involves high-dimensional models. In this paper, we choose to use approximations to certain probabilities in an approach that adopts a classical stance in a Bayesian framework (following Garratt *et al.,* (2003) [GLPS]).

To be more specific, if there are m different models, denoted M_i, $i = 1, ..., m$, each characterized by a probability density function of $\mathbf{z}_t$ defined over the estimation period $t = 1, ..., T$, as well as the forecast period $t = T+1, ..., T+H$, then the rules of conditional probability imply

$$\Pr\left(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T}\right) = \sum_{i=1}^{m} \Pr\left(M_i \mid \mathbf{Z}_{1,T}\right) \Pr(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T}, M_i). \quad (4)$$

Thus, inference about $\mathbf{Z}_{T+1,T+H}$ involves a weighted average of the models' density functions, with weights being the model probabilities. This is "Bayesian model averaging" (BMA). To implement this here, we follow Burnham and Anderson (1998) who suggest the use of the familiar Akaike information criterion to obtain model weights w_{iT}:

$$\Pr\left(M_i \mid \mathbf{Z}_{1,T}\right) = \frac{\exp(AIC_{iT}^{*})}{\sum_{j=1}^{m} \exp(AIC_{jT}^{*})}. \quad (5)$$

where $AIC_{iT}^{*} = AIC_{iT} - \max_j(AIC_{jT})$, $AIC_{iT} = LL_{iT} - k_i$ is the Akaike information criterion, and LL_{iT} is the maximized value of the log-likelihood function for model M_i calculated on the basis of the sample running to period T.[6] We also approximate the densities $\Pr(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T}, M_i)$ for each model using maximum likelihood estimates of the models. These assumptions allow $\Pr\left(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T}\right)$ to be estimated straightforwardly using (4) based on ML estimation of the candidate models. Further, if the models are sufficiently simple, integrals of the form given in (3) can also be readily evaluated through simulation methods under these assumptions. (See Appendix A for details when the models are in VAR form).

[6] Alternatively, Draper (1995) suggests using Schwarz Bayesian information criterion weights, using SBC in place of AIC in (5). The SBC weights are asymptotically optimal if the data generation process lies in the set of models under consideration, but the AIC weights are likely to perform better when the models represent approximations to a complex data generation process. Fernandez *et al.*(2001) note that the choice of uninformed priors implies Bayes factors which behave asymtotically like SBC.

2.2 Decision-Based Forecast Evaluation

The above discussion shows that there might be a variety of alternative predictive densities available to a decision-maker, including model-specific densities, $\Pr(\mathbf{Z}_{T+1,H} \mid \mathbf{Z}_T, M_i)$, $i = 1, ..., m$, and densities obtained through model averaging. Pesaran and Skouras (2000) suggest a decision-based criterion function for the evaluation of a predictive density function which, in the context of (3), is given by

$$\begin{aligned} \Psi &= \mathrm{E}_P \left[\nu(W_{T+H}(\omega^\dagger) \mid \Omega_T \right] \\ &= \int \nu(W_{T+H}(\omega^\dagger)) P\left(\mathbf{Z}_{T+1,H}\right) d\mathbf{Z}_{T+1,H}, \end{aligned} \tag{6}$$

where $\omega^\dagger$ is the chosen optimal value of ω for the given predictive density and $\mathrm{E}_P[.]$ is the expectations operator with respect to $P\left(\mathbf{Z}_{T+1,H}\right)$, the "true" probability density function of $\mathbf{Z}_{T+1,H}$ conditional on Ω_T. This can be viewed as the average utility obtained using the given predictive density function when large samples of forecasts and realisations are available. The criterion function for the evaluation of the predictive density function clearly depends on the decision-making context, as captured by the utility function $\nu(.)$. Pesaran and Skouras show that the form of this criterion function is independent of the parameters of the underlying utility function only in the special case of the "LQ problem" involving a single decision variable (where the utility function is quadratic and constraints (if they exist) are linear). In that special case, the criterion is proportional to the MSE so that the purely statistical measure is appropriate. However, even the multivariate version of the LQ problem involves the parameters of the utility function so that, generally, statistical and decision-based forecast evaluation criteria are markedly different.

In evaluating the alternative prediction densities, based on alternative models or model averages, the sample counterpart of the criterion function in (6) is

$$\overline{\Psi} = \frac{1}{N}\sum_{s=1}^{N} \nu(W_{T+H+s}(\omega^{\dagger})), \tag{7}$$

calculated recursively for $s = 1, .., N$ for each predictive density (with associated optimal share $\omega^{\dagger}$) and over the out-of-sample forecast evaluation period $T + s$,.., $T + H + s$. This provides an estimate of the realised utility to the decision-maker of using the predictive distribution function. In practice an absolute standard for forecast evaluation is not available because the true probability density function of the forecast variable is not known. But calculating loss differentials, comparing the economic value of outcomes based on alternative predictive distributions, is straightforward and a choice between the two can simply depend on whether the differential is positive or negative. If one predictive distribution function is given the status of a 'null', then the choice can be cast in terms of whether the loss differential is significantly greater than zero. The asymptotic distribution of the loss differential can be derived in the case of the LQ problem (see Diebold and Mariano, 1995) but the nature of the test needs to be investigated on a case-by-case basis for other problems. This is relatively straightforward in the linear VAR case discussed here, however, since the distributional properties of the criterion function under the null can also be obtained through simulation.

3 The Candidate Set of Models

The exercise described above requires that we forecast the exchange rate. In what follows, we consider four alternative models on which forecasts of e_t can be based. In the modelling exercises, we assume the variables of interest in $\mathbf{z}_t$, including e_t, are $I(1)$ so that the candidate set of models can be written in the vector error correction (VECM) form:

$$\Delta \mathbf{z}_t = \mathbf{a}+\sum_{i=1}^{p} \Gamma_i \Delta \mathbf{z}_{t-i} + \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{z}_{t-1} + \mathbf{u}_t, \tag{8}$$

using alternative cointegrating vectors $\boldsymbol{\beta}$ as suggested by the theory associated with the alternative models. Each model that we consider therefore represents a restricted version of the levels form in Appendix A, chosen to reflect a particular view on exchange rate determination.

The set of models that we consider for predicting exchange rates is:

- M_E : Efficient Market Hypothesis [EMH]
- M_M : Monetary Fundamentals model [MF]
- M_P : Purchasing Power Parity [PPP]
- M_A : Autoregressive model of e_t in differences [AR(p)]

In the EMH, we define $\mathbf{z}_t = (e_t,\ f_t)'$ where f_t is the logarithm of the forward (end-of-period) nominal bilateral exchange rate and we assume the cointegrating vector is given by $\boldsymbol{\beta}' = (1, -1)$. This model relates to the literature on foreign exchange market efficiency which tests whether the forward rate is an optimal predictor of the future spot exchange rate. Although the empirical evidence is mixed regarding the optimality of the forward rate as a predictor of the spot rate, there is evidence that some information is contained in the term structure of the forward rate; see Clarida and Taylor (1997), for example. Moreover the EMH specification we adopt does not require efficient markets to hold at all points in time.

In the MF model, $\mathbf{z}_t = (e_t,\ x_t)'$, where x_t represents a 'fundamentals' term, given by $x_t = (m_t - m_t^*) - (y_t - y_t^*)$, and m_t and y_t denote the log-levels of the domestic money supply and real income respectively, the '*' superscript indicates the corresponding foreign variable, and $\boldsymbol{\beta}^{'} = (1, -1)$. This specification has a long tradition in the analysis of exchange rate determination (Frenkel, 1976; Mussa, 1976, 1979; Frenkel and Johnson, 1978), and has recently been the subject of much debate (as in Mark, 1995; Mark and Sul, 2001; Berkowitz and Giorgianni, 2001, for example).

In the PPP model, $\mathbf{z}_t = (e_t,\, p_t - p_t^*)^{'}$, where p_t and p_t^* denote the logarithm of the domestic and foreign price level respectively and $\boldsymbol{\beta}^{'} = (1, -1)$ so that the real exchange rate is stationary. Like the EMH, this theory is often viewed as an arbitrage condition in international goods and is considered to be an integral part to many open economy views of the world. The literature considering the empirical validity of PPP is well developed and the conclusions are mixed, but there is some recent evidence that it may hold in the long-run (see Garratt *et al.,* 2006, for example).

Finally the simplest model that we consider is an autoregressive model of order p in the change in exchange rates, so that $\mathbf{z}_t = (e_t)$ and $\boldsymbol{\beta}^{'} = 0$. This type of specification is widely used and constitutes an atheoretical alternative against which to judge the other three, more structural, models of exchange rate determination.[7]

4 Exchange Rate Models for the US, UK and Japan

4.1 Data

In our empirical work, we use monthly data for the US, UK and Japan over the period 1981m1-2006m6 (306 observations) and consider two separate exercises based on the decision to invest in the US or UK, and the decision to invest in the US or Japan. Variables employed in the analysis include short term 3-month nominal interest rates (r_t

[7] In the empirical section, all four models are assessed relative to the standard random walk model without drift.

and r_t^*), money supply (m_t and m_t^*), industrial production (y_t and y_t^*) and consumer prices (p_t and p_t^*) in the three countries. We also consider the one month spot- and forward- nominal exchange rates (denoted by e_t and f_t respectively) for Sterling-Dollar and Yen-Dollar. All the data used in the analysis are in natural logarithms and the precise definitions, sources and transformations are described in the Data Appendix. The main sample period used in estimation is 1981m1-2002m6, but we also consider data up to four years later for out-of-sample model evaluation.

Figures 1-8 plot the levels and first differences of the exchange rates, the level of short term interest rates and their differentials, plus the excess returns computed as $\Delta e_{t+1} - (r_t - r_t^*)$. Figures 1-4 show the exchange rates to be volatile and suggest that they are non-stationary in levels (confirmed by unit root tests). For the out-of-sample forecasting period 1990m1 onwards, the Pound-Dollar rate shows no clear pattern, first depreciating but then appreciating back to the levels observed at the beginning of the period. The Yen-Dollar exchange rate also shows an appreciation which is then mostly reversed. Figures 5-6 suggest non-stationarity of the interest rates for the sample period, with similar looking downward trends in all three rates demonstrating some co-movement. The differentials are mostly positive for the US-Japan case and negative for the US-UK case. The differentials also look downward trended in the first half of the sample. The excess returns in Figures 7 and 8 are volatile and do not exhibit any clear patterns.

[INSERT FIGURES 1-8 HERE]

4.2 Estimation

Our empirical analysis began by testing the assumption in Section 3 that all variables are $I(1)$. In every case, we failed to reject the null of a unit root in levels but rejected the null in first differences. We therefore proceeded in the analysis assuming all variables are

$I(1)$.[8] Next, we selected the lag length to be used in our forecasting models by estimating a sequence of unrestricted VAR(p), $p = 0, 1, 2, ...12$ for each set of variables employed in the models M_E to M_A for both the US-UK and US-Japan data sets and over the sample period 1981m1-2002m6. The lag selection criteria used was the likelihood ratio test and the lag length chosen was twelve for all models in both data sets.[9]

Analysis of the long-run relationships that exist among the variables provides good evidence to support the pairwise cointegration of exchange rates with the various explanatory variables in our candidate set of models (i.e. with f_t, x_t and $p_t - p_t^*$ respectively). There is, however, weaker evidence to support the one-to-one relationships suggested by the theories.[10] Nevertheless, estimating the models outlined in (8) over the full sample, assuming the long-run restrictions suggested by the various theories hold, can provide the basis of exchange rate forecasts. For reasons of parsimony, we do not report the full set of results for each model here, but Table 1 documents some basic diagnostics for the Δe_t equations of each of the models, along with those from a 'random walk' model, referred to as model M_{RW} and estimated as a reference against which to compare the performance of the four models. The results indicates that the diagnostic tests for the exchange rate equations are reasonable although their explanatory power is low (in line with most empirical findings in the literature). Like much of the literature, then, we are faced with a trade off between empirical fit and a form which reflects known theories of exchange rate determination. The fact that none of the empirical models seems entirely satisfactory on

[8] The results are available from the authors on request.

[9] Here we do not investigate uncertainty with respect to the lag length.

[10] Johansen's trace test indicates the presence of a cointegrating relationship in all cases except the PPP relationship in the US-Japan dataset. Formal tests of the one-to-one relationships suggested by theory were rejected in all cases except the efficient markets and fundamental relationships in the US-Japan case. Full details of the results are available on request.

purely statistical grounds (particularly with respect to their long run properties), and that no one model unambiguously dominates the others in terms of model diagnostics, lies at the heart of the model uncertainty experienced in decision-making.

[INSERT TABLE 1 HERE]

Probabilistic statements on the likely relevance of models over the estimation period can be made on the basis of the weights given in (5). Table 2 and Figures 9-10 report on the model weights, w_{iT}, based on the AIC statistics and calculated according to formula (5) for $T = 1989m12$, $1990m3, ...$, $2002m6$.[11] To obtain these statistics, and in anticipation of the forecasting exercise below, the four models were each estimated over the period 1981m1-1989m12 and then recursively, at three month intervals, through to 1981m1-2002m6 (making 51 recursions in total). Table 2 shows that, averaging over all 51 recursions, there is reasonably strong support for the AR model M_A, some support for the efficient markets model M_E and the monetary fundamentals model M_M and little support for the PPP model M_P in both the US-UK and US-Japan exercise based on these weights. However, the figures illustrate that the average statistics hide some considerable time variation in the weights, with the efficient markets model performing reasonably well in both exercises in the early periods and the monetary fundamentals model also showing with significant weights in the US-UK case at that time. These figures reflect

[11] As our candidate set of models are not nested, system-based criteria are not directly comparable. Instead, the reported AIC statistics of Table 3 are based on the equation explaining Δe_t, in each model, taking the equation for this series in isolation from the system in which it is embedded. For example, the criteria are based on equation log likelihoods calculated by $LL = \frac{-n}{2}\left\{1 + \log(2\Pi\widetilde{\sigma}^2)\right\}$ where $\widetilde{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n}$ and $\mathbf{e}$ are the equation residuals. Such a decomposition of a system's likelihood effectively assumes the covariances between the variables of interest and the other variables in the system are negligible. While this is unlikely to be true in practice, these approximations allow model comparison across alternative systems.

the fact that the weights can be quite sensitive to even relatively small movements in the values of the equation likelihoods over time. But they reinforce again the view that it is difficult to choose between the models on purely statistical criteria.

[INSERT TABLE 2 HERE]

[INSERT FIGURES 9 AND 10 HERE]

4.3 Statistical Evaluation of Forecasting Performance

The out-of-sample forecasting performance of the models can be evaluated statistically by calculating the root mean squared error (RMSE) relating to the forecasts of the (cumulative) exchange rate change, defined for forecast horizon H at time T as $c_T(H) = \sum_{h=1}^{H} \Delta e_{T+h}$. The RMSE are calculated for each model and their ratios, relative to a random walk, are reported in Table 3 for forecast horizons $H = 1, 3, 6, 12, 24, 36$ and 48. The table also reports the ratios of RMSEs obtained using a weighted average of forecasts from all the models, with equal weights (i.e. $\frac{1}{4}$) and weights based on AIC as in (5). The reported statistics in the table are again averages based on the RMSEs obtained in 51 recursions covering the evaluation period T = 1989m12, 1990m3, ..., 2002m6 at three-monthly intervals.

Table 3 indicates that, as a rule, our exchange rate models' forecasts perform poorly relative to those of a random walk model when judged using a statistical measures such as RMSE. This is generally true for both data sets, although the finding is stronger for the US-UK data set where the random walk model's forecast performance dominates all the others at virtually every horizon.[12] For the US-Japan data set, the ratios are typically greater than unity but not so large, and model M_p outperforms the random

[12]Note, however, that the Diebold-Mariano (1995) tests (not reported) suggest these differences are not statistically different from zero.

walk model M_{RW} for $H \geq 6$. These results are of course consistent with the literature and are not too surprising given that the models other than M_{RW} are heavily parameterised; as shown in Clements and Hendry (2005), using RMSE as a criterion penalises models for including variables with low associated t-values even if the model is misspecified by their exclusion.[13]

[INSERT TABLE 3 HERE]

Interestingly, the forecasts generated by the equal-weights average model dominates those of the M_{RW} in both sets of results and at nearly all horizons according to this criteria (see for example Timmerman, 2006). This finding is in line with the findings in the literature, described in the review of Clemen (1989) and more recently by Harvey and Newbold (2005) for example, that combinations of forecasts typically perform well in a statistical sense and can outperform the forecasts of a single model even if this is the true (but estimated) data generating process. Moreover, the performance of the AIC-weighted average model also shows relatively well. While its forecasts are dominated by those of M_{RW} at the shorter horizons, the ratio in Table 3 drops below unity at longer horizons and the ratios are well-below the mean of the individual models;' ratios (indicating that averaging and exploiting the time-variation of the weights serves to improve the $RMSE$ performance).

In brief, then, a statistical evaluation of the models in terms of their diagnostic statistics or in-sample fit provides relatively little guidance on the appropriateness of the various models for use in investment decisions or on the gains to be made from the various models. In terms of forecasting performance measured by RMSE, the theory-

[13] More precisely, Clements and Hendry (2005) show that forecasting stationary processes using a model that retains all variables with an expected $(t - value)^2 > 2$ will dominate in terms of one-step ahead forecast accuracy measured by RMSE.

based models do not perform well although an equally weighted model-averaging is useful (if there is ambiguity over the true model). It remains to be seen whether a more clear-cut picture emerges on the usefulness of these models' forecasts when they are judged more directly in the context of the objectives of the investment decision.

5 Forecast Evaluation by US Investors

Section 2 described the decision made by a buy-and-hold investor with a given horizon to be one of solving the problem in (3) to choose the proportion of her/his portfolio that should be devoted to domestic and foreign assets. This choice requires the implementation of the simulation-based procedure described in Appendix A to obtain the probability distribution of the future values of $\mathbf{Z}$, $\Pr(\mathbf{Z}_{T+1,\ T+H}\ |\mathbf{Z}_{1,T}, M_i, \boldsymbol{\theta}_i)$, with which to evaluate (and then maximise) expected future utility. A description of the algorithm used to compute the optimal portfolio shares in also provided in Appendix A.

Having computed the portfolio shares we are then able to conduct what is the main focus of this paper; namely, an ex-post forecasting exercise which uses the optimal portfolio shares and evaluates, given observed outcomes for exchange rates and interest rates, the end-of-period wealth and utility (for each investment horizon) obtained from the buy and hold strategy. The evaluation takes the form of comparing the utility ratios of the models and the model-averages relative to a benchmark strategy which allocates wealth using the random walk model M_{RW} to forecast the exchange rate.

Note the exercise reported here considers the average values of the optimal portfolio shares and maximised utility ratios calculated across the 51 recursions. The reported results illustrate the outcome of the investment strategy if it had been repeated at quarterly intervals throughout the nineties. The analysis covers a range of observed exchange rate and interest rate paths, therefore, and mitigates against the possibility that our results are period-specific.

5.1 Portfolio Weights

Tables 4a and 4b report the optimal portfolio share allocated by a US investor, averaged through the nineties and early 2000's, over the investment horizons $H = 1, 3, 6, 12, 24, 36, 48$ and for three different values of the coefficient of risk aversion, $A = 2$, 5, and 10. The Tables report the shares that would have been chosen if forecasts were obtained employing the four alternative models, the equal-weight and AIC-based average models and the random walk model of exchange rate determination. Table 4a relates to the choice between US and UK assets and Table 4b relates to the US-Japan choice. The statistics are again generated in the recursive manner described above. Hence, the models are estimated first for the period 1981m1-1989m12 and the optimal portfolio shares decided based on the forecasts obtained from the various models. The process is then repeated moving forward three months, recomputing the model weights (for the average models) and wealth and utility forecasts to obtain new optimal shares for each model. This process is repeated for each recursion until we have results for 51 recursions covering the evaluation period $T = 1989m12$, $1990m3, ...,$ $2002m6$ at three-monthly intervals. The statistics reported in Table 4 show the average portfolio share across the 51 recursions.

There are a number of interesting features of the statistics reported in the tables. As expected, given the uncertainties associated with the exchange rate, the proportion of wealth in foreign assets falls as the risk aversion parameter rises and as the investment horizon increases. So, for example, if we simply average the figures in the columns of Table 4a for the models M_E, M_M, M_P and M_A i.e. over all investment horizons, the share allocated to UK assets falls from 29% when $A = 2$ to 22% when $A = 5$ and to 15% when $A = 10$. Taking the average of the rows for horizons $H = 1$, 12 and 48, again across models M_E, M_M, M_P and M_A, the share allocated to UK assets falls from 35% when $H = 1$ to 19% when $H = 12$ and to 16% when $H = 48$.

The average results accommodate considerable heterogeneity in outcome across the various models, however, as shown in the table. Hence, for example, again taking the averages across the investment horizons and with $A = 5$, model M_E suggests a holding of 30%, M_M 38%, M_P 7% and M_A 11%. The model M_{RW} suggests 26% holdings for $A = 5$ and this reflects the general pattern whereby M_E and M_M broadly suggest high UK investment holdings relative to M_{RW} while M_P and M_A suggest lower UK holdings.

Similar patterns are observed in Table 4b in terms of lower Japanese holdings as A or H rise, although the Japanese investment appears to be generally more attractive than UK ones. Here, averaging across the columns in the table, the shares of Japanese assets are 39%, 27% and 17% for $A = 2$, 5 and 10 respectively. Averaging across rows or investment horizons gives figures of 38%, 27% and 18% respectively for $H = 1$, 12 and 48. In the Japanese case, all the models suggest larger Japanese investments than M_{RW} which suggests very low Japanese holdings for all A and H.

[INSERT TABLES 4A AND 4B HERE]

5.2 Economic Evaluation

Table 5 provides an economic evaluation of the forecast performance of the various models from the perspective of an investor with risk-aversion parameters of $A = 2$, 5 and 10. The evaluation addresses the question of how well an investor would have done in terms of utility outcomes if she had used the optimal weights suggested by the various models and summarised in Table 4. Hence, the table describes the (average) end-of-period utilities that would have been obtained over the period 1989m12-2002m6 if the investor had chosen the optimal portfolio shares suggested by each of the four models, or the averaged models, in real time. The utilities are expressed as a ratio to the utility that would have been achieved if the investor had followed an investment strategy based on the random walk model forecasts. As before, the statistics reported in the table relate to the average outcome over 51 recursions in the evaluation period (i.e. setting $N = 51$ in (7)).

The results of Table 5a show that, generally speaking, the economic models outperform the random walk model for the UK-US case with utility ratios in excess of unity in most cases for all A and H.[14] This, of course, is in direct contrast with the comparisons based on RMSE's where the random walk model appeared to perform best. Models M_E and M_M, which encouraged investors to hold more UK investments than M_{RW}, perform particularly well with utility more than in excess of 15% higher than M_{RW} utility in some cases. The model averages also typically produce figures in excess of unity although, again in contrast with the comparison based on RMSE's, the average models do not outperform the individual models.[15] Qualitatively similar results are also found in the US-Japanese results reported in Table 5b. These also show end-of-period utility ratios based on the economic models that are systematically higher then those based on M_{RW}, again contrasting with the RMSE results. (The exception being model M_M in this case). And the AIC-based and equal-weight average model again (marginally) outperform the M_{RW} over most horizons and risk parameters but, as in the US-UK case and again in contrast to the conclusion drawn using RMSE's, the average models are outperformed by the economic models when the economic criterion is used.

[14] The model incorporating PPP is the exception.

[15] In addition to the utility outcomes we also compute the Sharpe Ratio as a second measure of economic performance. The results are broadly consistent with the utility ratios, in that they suggest the random walk model does well as compared to the worst performing model being M_P, with the other models between these two.

Of course, the conclusions drawn above are based on straight comparisons of the estimated utilities obtained in real time using the optimal portfolios suggested by the various models compared, in each case, with the utility achieved using the random walk model. While figures in excess of unity indicate that a model outperforms the M_{RW}, it is difficult to interpret the statistical significance of these figures. Table 5 provides an indication of the significance of the figures by reporting the outcome of a simulation exercise in which at each of our 51 replications, we simulated 1000 artificial 'futures' based on the estimated random walk model and conducted precisely the same evaluation exercise as described above for each of the artificial datasets; i.e. calculating the (average) end-of-period utilities obtained following the investment strategies suggested by the different economic models and expressing these as a ratio to the utility obtained using the M_{RW} model. The simulated distributions of utility ratios shows the range of outcomes that would be obtained if the random walk model were the true data generation and comparison of the utilities in Table 5 with the 95th, 90th, 80th and 70th percentiles of these distributions gives an indication of the statistical significance of the figures. The simulated distributions demonstrate that there is reasonable variability in the utility ratios that would be obtained, but the superscripts attached to the figures in the table show that there are a reasonable number which appear to be relatively small numbers but are 'significantly' greater than unity by this standard.

[INSERT TABLE 5 HERE]

6 Concluding Comments

The results described above illustrate strikingly that judgements on the forecasting performance of the various models can be quite different depending on whether the evaluation is based on a statistical approach or a decision-based approach. According to the statistical view based on RMSEs, the simple random walk model M_{RW} systematically outperforms the structural models at all horizons, but is defeated by a model average. In contrast, according to the decision-based criteria, models incorporating economically-meaningful relations outperform the the M_{RW} model and the artificial model averages (with the model accommodating the efficient markets hypothesis performing relatively well for both the US-UK and US-Japan case).

While it will remain unusual for the decision-making environment to be fully articulated, it is clear from this empirical exercise that, when it is possible, models and their forecasts should be evaluated according to the purpose to which they will be used. The exercise also show that the technical issues involved in decision-based forecast evaluation can be readily addressed using the methods outlined in the paper, based on relatively straightforward simulation exercises, even where complex objective functions or many variables or model uncertainty are involved.

Acknowledgements

We are grateful to Stephen Hall, Lucio Sarno and participants at the conference of the Society for Computational Economics for helpful comments on earlier versions of the papers.

Data Appendix

The sources and transformations for the data are as follows:

e_t : the natural logarithm of the UK Sterling and Japanese Yen per US Dollar nominal spot exchange rate. Source: International Financial Statistics (IFS), codes 112AGZF and 158AEZF respectively.

f_t : the natural logarithm of the Sterling/Dollar and Yen/Dollar one month forward exchange rate. For the Sterling/Dolllar rate, source: Bank of England, code XUMLDS1. For Yen/Dollar rate, we used three sources: (i) for 1979m1-1992m8, the data is from Hai et al. (1997) available from the *JAE Data Archive* (ii) for 1992m9-1996m12, we constructed the data assuming covered interest parity i.e. $f_t = r_t - r_t^* + e_t$.and (iii) for 1997m1-2003m6, the data was collected from Datastream, code USJPF.

r_t : the US (domestic) three month treasury bill rates, expressed as a monthly rate: $r_t = 1/12 \times \ln[1 + (R_t/100)]$, where R_t is the annualised rate. Source: IFS, code 11160C.

r_t^* : for foreign short term interest rates we defined monthly rates as: $r_t^* = 1/12 \times \ln[1 + (R_t^*/100)]$, where R_t^* is the annualised rate. For the UK we used the three month treasury bill rates, source: IFS, code 11260C and for Japan the three month discount rate, source: IFS, code 15860ZF. .

y_t : the natural logarithm of US industrial production, constant 1995 prices, 1995=100. Source: IFS, code 11166 CZF.

y_t^* : the natural logarithm of UK and Japanese industrial production, constant 1995 prices, 1995=100. Source: IFS, codes 11266 CZF and 15866 CZF respectively.

p_t : the natural logarithm of US (domestic) consumer prices, index 1995=100. Source: IFS, code 11164ZF.

p_t^* : the natural logarithm of UK and Japanese (foreign) consumer prices, index 1995=100. Source: IFS, codes 11264ZF and 15864ZF respectively.

m_t : the natural logarithm of US (domestic) narrow money (M1 seasonally adjusted). Source: IFS, code 11159MA.

m_t^* : the natural logarithm of UK and Japanese (foreign) narrow money (M0 seasonally adjusted). Source: IFS, codes 11259MC ZF and 15834BZF respectively.

Appendix A: Calculating $\Pr(\mathbf{Z}_{T+1,H} \mid \mathbf{Z}_T, M_i)$ in the Linear Case

To illustrate more practically how we evaluate and make use of predictive density functions using simulation methods, assume that each of the models M_i can be written in the VAR form

$$\mathbf{z}_t = \sum_{s=1}^{p} \Phi_s \mathbf{z}_{t-s} + \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{v}_t,\ t = 1, 2, ..., T, T+1, .., T+\ H, \tag{9}$$

where Φ_s is an $n \times n$ matrix of parameters, $\mathbf{a}_0$, and $\mathbf{a}_1$are $n \times 1$ parameter vectors and $\mathbf{v}_t$ is assumed to be a serially uncorrelated *iid* vector of shocks with zero means and a positive definite covariance matrix, Σ. Using this model, an estimate of the probability distribution function of the forecasts can be obtained using stochastic simulation techniques.

Specifically, suppose that the ML estimators of the parameters in (9) Φ_s, $i = 1, \ldots, p$, $\mathbf{a}_0$, $\mathbf{a}_1$ and Σ are denoted by $\hat{\Phi}_s$, $i = 1, \ldots, p$, $\hat{\mathbf{a}}_0$, $\hat{\mathbf{a}}_1$ and $\hat{\Sigma}$, respectively. Then the point estimates of the h-step ahead forecasts of $\mathbf{z}_{T+h}$ conditional on Ω_T, denoted by $\hat{\mathbf{z}}_{T+h}$, can be obtained recursively by

$$\hat{\mathbf{z}}_{T+h} = \sum_{s=1}^{p} \hat{\Phi}_i \hat{\mathbf{z}}_{T+h-s} + \hat{\mathbf{a}}_0 + \hat{\mathbf{a}}_1 (t+h),\ h = 1, 2, \ldots, \tag{10}$$

where the initial values, $\mathbf{z}_T, \mathbf{z}_{T-1}, \ldots, \mathbf{z}_{T-p+1}$, are given. Hence, abstracting from parameter uncertainty for the time being, we can obtain an estimate of $\Pr(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T}, M_i)$ using stochastic simulation, obtaining forecast values of $\mathbf{z}_{T+\ H}$ using

$$\mathbf{z}_{T+h}^{(r)} = \sum_{s=1}^{p} \hat{\Phi}_i \mathbf{z}_{T+h-s}^{(r)} + \hat{\mathbf{a}}_0 + \hat{\mathbf{a}}_1 (t+h) + \mathbf{v}_{T+h}^{(r)}, \quad h = 1, 2, ..., H \quad \text{and} \quad r = 1, 2, ..., R, \tag{11}$$

where superscript '(r)' refers to the r^{th} replication of the simulation algorithm, and $\mathbf{z}_T^{(r)} = \mathbf{z}_T$, $\mathbf{z}_{T-1}^{(r)} = \mathbf{z}_{T-1}, \ldots$, $\mathbf{z}_{T-p+1}^{(r)} = \mathbf{z}_{T-p+1}$ for all r. The $\mathbf{v}_{T+h}^{(r)}$'s can be drawn either by parametric methods based on $\hat{\Sigma}$ or by non-parametric methods based on the estimated residuals on which $\hat{\Sigma}$ is calculated (see GLPS for more details).

These simulation exercises provide estimates of $\Pr(\mathbf{Z}_{T+1,T+H} \mid \mathbf{Z}_{1,T}, M_i)$ which can be used as predictive densities assuming a particular model is appropriate, or which can be used in a model averaging exercise. For any particular density, the simulations also allow us to evaluate $\mathrm{E}[\nu(W_{T+\ H}) \mid \Omega_T]$ in (3) for a range of values of ω (in practice calculating $\nu(W_{T+\ H}(\omega_0))$ in each replication for various values of ω_0 and calculating the mean value across replications). The investor's decision then simply involves choosing the ω associated with the maximum value of the simulated expected wealth. Specifically:

1. For a given model, M_i, with a fixed set of parameters, $\boldsymbol{\theta}_i$, we generate a sequence of forecasts for $\Delta_H e_{T+H}^{(r)}$, for $h = 1, .., H$ and $r = 1, ..., R$ (where $R = 10,000$ and $i = 1, ..., 4$) based on draws from a distribution of errors. These are non -parametric draws in our case.

2. For each replication r, we compute the value of $W_{T+H}^{(r,\omega)}$ using equation (1) and assuming that r_{T+h} and r_{T+h}^*, $h = 1, ..., H$, are known, where ω has 101 values $\omega = 0, ..., 1$ in step lengths of 0.01. Hence, we have a total of $R \times 101$ values of $W_{T+H}^{(r,\omega)}$ for each forecast horizon H. The forecast horizons considered in the paper are $H = 1, 3, 6, 12, 24, 36$, and 48.

3. We translate $W_{T+H}^{(r,\omega)}$ into the utility $v(W_{t+H}^{(r,\omega,A)})$, using CRRA utility defined in equation (2), for each level of risk aversion $A = 2, 5$ and 10. Then we compute, for the given A, H and ω,:
$$\frac{1}{R}\sum_{r=1}^{R} v(W_{T+H}^{(r,\omega,A)}).$$

4. The optimal portfolio weight, for each forecast horizon H and level of risk aversion A, is the value of ω which maximizes the above expression; i.e. the maximum utility over the 101 different values of the portfolio weight ω.

5. Repeat for all models $i = 1, ..., 4$ and the equal weight and AIC average models.

References

Abhyankar, A., Sarno, L., and G. Valente (2005), "Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability", *Journal of International Economics*, 66(2), 325-348.

Avramov, D. (2002), "Stock Return Predictability and Model Uncertainty", *Journal of Financial Economics*, 64, 423-458.

Baberis,N. (2000), "Investing for the Long Run when Returns Are Predictable", *Journal of Finance*, Vol. LV, N0. 1, 225-264.

Berkowitz, J. and L. Giorgianni (2001), "Long-Horizon Exchange Rate Predictability?", *Review of Economics and Statistics*, 83(1), 81-91.

Campbell, J.Y. and R.J. Shiller (1987), "Cointegration Tests and Present Value Models", *Journal of Political Economy*, 95, 1002-1088.

Campbell, J.Y. and R.J. Shiller (1991), "Yield Spreads and Interest Rate Movements: A Birds Eye View", *Review of Economic Studies*, 58, 495-514.

Campbell, J.Y. and L.M., Viceira (2002), *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Claredon Lectures in Economics, Oxford University Press: Oxford.

Cheung, Y., M.D. Chinn and A.G. Pascual (2005), "Empirical Exchange Rate Models of the Nineties: Are They Fit to Survive?", *Journal of International Money and Finance,* 24 (7), 1150-1175.

Clarida, R.H. and M.P. Taylor (1997), "The Term Structure of Forward Premiums and the Forecastability of Spot Exchange Rates", *Review of Economics and Statistics*, 89, 353-361.

Clarida, R.L., L. Sarno, M.P. Taylor and G. Valente (2003), "The Out of Sample Success of Term Structure Models as Exchange Rate Predictors: a step beyond", *Journal of International Economics*, 60, 61-83.

Clemen, R.T. (1989), "Combining Forecasts: A Review and Annotated Bibliography with Discussion, *International Journal of Forecasting*, 5, 559-608.

Clements, M. and D. Hendry (2005), "Information in Economic Forecasting", *Oxford Bulletin of Economics and Statistics*, 67 (Supplement), 713-753.

Draper, D. (1995), "Assessment and Propagation of Model Uncertainty," *Journal of Royal Statistical Society Series B*, 58, 45-97.

Faust, J., J.H. Rogers and J.H Wright (2003), "Exchange Rate Forecasting: the error we've really made", *Journal of International Economics,* 60, 35-59.

Fauvel, Y., A. Paquet and C. Zimmerman (1999), "A Survey on Interest Rate Forecasting", CREFE Working Paper 87, University of Quebec.

Fernandez, C., E. Ley and M. F. J. Steel, (2001), "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 100, 381-427.

Garratt, A., K. C. Lee, M.H. Pesaran and Y. Shin, (2003), Forecast Uncertainties in Macroeconometric Modelling: An Application to the UK Economy, *Journal of American Statistical Association, Applications and Case Studies*, 98, 464, 829-838.

Garratt, A., K. C. Lee, M.H. Pesaran and Y. Shin, (2006), *Global and National*

Macroeconometric Modelling: A Long-Run Structural Approach, Oxford University Press, Oxford.

Granger, C. W. J. and M. H. Pesaran (2000a), "A Decision Theoretic Approach to Forecast Evaluation," in Statistics and Finance: An Interface, eds. W. S. Chan, W. K. Li and H. Tong, London: Imperial College Press, pp. 261-278.

Granger, C. W. J. and M. H. Pesaran (2000b), "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, 19, 537-560.

Hai, Mark and Wu (1997), "Understanding Spot and Forward Exchange Rate Regressions", *Journal of Applied Econometrics*, 12, 715-734

Harvey, D.I. and P. Newbold (2005), "Forecast Encompassing and Parameter Estimation", *Oxford Bulletin of Economics and Statistics*, 67 (Supplement), 815-835.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). "Bayesian Model Averaging: A Tutorial", *Statistical Science*, 14, 382-417.

Killian, L. and M.P. Taylor (2003), "Why is it so Difficult to Beat a Random Walk Forecast of Exchange Rates?", *Journal of International Economics*, 60, 85-107.

Kandel, S. and R.F. Stanbaugh (1996), "One the Predictability of Stock Returns: An Asset-Allocation Perspective", *Journal of Finance*, 51, 385-424.

Leitch, G. and J.E. Tanner (1991), "Economic Forecasts Evaluation: Profits Versus the Conventional Measures", *American Economic Review*, 81, 580-590.

Mark, N.C. (1995), "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability", *American Economic Review*, 85, 201-218.

Mark, N.C. and D. Sul (2001),"Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post Bretton Woods Panel", *Journal of International Economics,* 53, 29-52.

Meese, R.A., and K. Rogoff (1983),"Empirical Exchange Rate Models of the Seventies: Do they Fit Out of Sample?", *Journal of International Economics*, 14, 3-24.

Pesaran, M. H. and Timmermann, A. (1995), "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50, 1201-1228.

Pesaran, M.H. and S. Skouris (2002),"Decision-based Methods for Forecast Evaluation", in Clements, M.P. and Hendry, D.F. (eds.), *A Companion to Economic Forecasting,* Oxford: Blackwell.

Sarno, L. and G. Valente (2004), "Comparing the Accuracy of Density Forecasts form Competing Models", *Journal of Forecasting*, 23, 541-557.

Timmerman, A. (2006), "Forecast Combinations". In: Elliot, G., Granger, C.W.J., Timmerman, A. (Eds.) *Handbook of Economics Forecasting,* Elsevier, Amsterdam, pp. 135-196.

Wright, J.H. (2003), "Bayesian Model Averaging and Exchange Rate Forecasts", downl;aodable at http://www.econ.jhu.edu/People/Wright/index.html, forthcoming *Journal of Econometrics.*

Wallis, K. F. (1999), "Asymmetric Density Forecasts of Inflation and the Bank of England's Fan Chart," *National Institute Economic Review*, 106-112.

West, K.D., H.J. Edison and D. Cho (1993), "A Utility-Based Comparison of Some

Models of Exchange Rate Volatility", *Journal of International Economics*, 35, 23-45.

Figure Legends

(1) Figure 1: Logarithm Pound-Dollar Spot Exchnage Rate

(2) Figure 2: Logarithm Pound-Yen Spot Exchnage Rate

(3) Figure 3: Change in the Pound-Dollar Spot Exchange Rate (percent)

(4) Figure 4: Change in the Yen -Dollar Spot Exchange Rate (percent)

(5) Figure 5: Short Term Interest Rates (Annual percent)

(6) Figure 6: Short Term Interest Rate Differentials (Annual percent)

(7) Figure 7: Excess Returns Pound-Dollar (percent)

(8) Figure 8: Excess Returns Yen -Dollar (percent)

(9) Figure 9: US-UK AIC Model Weights

(10) Figure 10: US-Japan AIC Model Weights

Note there are only two separate graph (pdf) files provided as Figures 1-8 and Figures 9-10 are on the same page. The figures were generated in Excel and then pasted into Scientific Word where pdf files were generated. I can, if required, produce separate files for each figure - but I am unclear what is needed here.

Table 1: **Exchange Rate Equation Diagnostics**

(a) US-UK

Model	LL	R^2	$F-test$	$S.E.$	$\chi^2_{SC}[12]$	$\chi^2_{H}[12]$	$\chi^2_{ARCH}[12]$
M_E	534.8	0.068	0.68 [0.87]	0.0321	11.76 [0.47]	68.33 [0.04]	11.93 [0.45]
M_M	543.0	0.126	1.34 [0.14]	0.0311	30.63 [0.00]	43.94 [0.71]	19.78 [0.07]
M_P	534.2	0.064	0.63 [0.91]	0.0320	15.73 [0.20]	63.98 [0.09]	8.86 [0.72]
M_A	530.4	0.036	0.77 [0.68]	0.0317	18.13 [0.11]	19.19 [0.74]	10.22 [0.60]
RW	525.6	0.000	-	0.0316	8.08 [0.78]	-	14.50 [0.27]

(b) US-Japan

Model	LL	R^2	$F-test$	$S.E.$	$\chi^2_{SC}[12]$	$\chi^2_{H}[12]$	$\chi^2_{ARCH}[12]$
M_E	517.3	0.094	0.96 [0.52]	0.0344	13.96 [0.30]	48.27 [0.54]	11.90 [0.45]
M_M	518.1	0.100	1.03 [0.43]	0.0343	13.18 [0.36]	38.43 [0.88]	12.27 [0.42]
M_P	521.7	0.124	1.32 [0.15]	0.0341	14.69 [0.26]	37.07 [0.91]	10.69 [0.56]
M_A	513.5	0.067	1.47 [0.14]	0.0339	14.40 [0.28]	15.76 [0.90]	10.94 [0.53]
RW	504.6	0.000	-	0.0343	17.77 [0.12]	-	8.75 [0.72]

Notes: RW denotes a random walk ‘benchmark’ model; models $M_A - M_T$ are described in the text. For model comparison and diagnosis, LL is the Log Likelihood, S.E. is the standard error of the regression, SC tests for the presence of serial correlation in the residuals, H tests for heteroscedasticity and ARCH tests for autoregressive conditional heteroscedasticity. P-values are given in [ ] brackets and the period of estimation is 1981m1-2002m6.

Table 2: **Average Model Weights, w_{it}^{AIC}, 1989m12-2002m6**

Model	US-UK	US-Japan
M_E	0.0932	0.1125
M_M	0.1221	0.0000
M_P	0.0000	0.0002
M_A	0.7847	0.8873

Notes: The weights reported here are the average calculated from the recursive regressions ran over 1981m1-1989m12 through to 1981m12-2002m6.

Table 3: **Root Mean Squared Error Ratios for the Exchange Rate (Relative to a Random Walk)**

(a) US-UK

Model	$H=1$	$H=3$	$H=6$	$H=12$	$H=24$	$H=36$	$H=48$
M_E	1.0408	1.0812	1.1125	1.1283	1.1564	1.1681	1.1481
M_M	1.1733	1.1678	1.2423	1.3320	1.4020	1.4077	1.3903
M_P	1.1643	1.1811	1.2221	1.2399	1.2705	1.2835	1.3428
M_A	1.0085	1.0036	1.0144	1.0129	1.0012	0.9942	0.9914
Equal-weight Av.	1.0065	0.9903	0.9739	0.9460	0.8988	0.8818	0.8617
AIC Average	1.0706	1.0419	1.0434	1.0307	0.9860	0.9755	0.9758

(b) US-Japan

Model	$H=1$	$H=3$	$H=6$	$H=12$	$H=24$	$H=36$	$H=48$
M_E	1.0984	1.0450	1.0757	1.0694	1.0468	1.0238	1.0221
M_M	1.0732	10686	1.0871	1.2206	1.3964	1.5666	1.7253
M_P	1.1005	1.0006	0.9887	0.9243	0.8692	0.8887	0.9215
M_A	1.0072	0.9906	1.0106	1.0197	1.0170	1.0079	1.0064
Equal-weight Av.	0.9838	0.9758	0.9673	0.9748	0.9832	0.9751	0.9473
AIC Average	1.0333	0.9948	1.0149	1.0155	0.9907	0.9858	0.9907

Notes: Reported statistics are average RMSEs ratios for the exchange rate, where a statistic less than one implies a superior RMSE performance to that of a random walk. We compute these, for each model and horizon, by calculating the RMSE of $c_T(H)$ for 51 quarterly recursions 1981m1-T, T=1989m12,...,2002m6 and then taking the average of the 51 RMSE ratios

Table 4: **Optimal Portfolio Shares Allocated to Foreign Assets**
(percentage averages for the recursions)
(a) US-UK

	Model M_E			Model: M_P		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	51.1	42.8	29.4	29.4	23.6	18.7
$H=3$	42.3	35.3	29.4	21.5	14.9	8.5
$H=6$	38.6	33.0	26.8	10.2	6.4	3.9
$H=12$	37.1	31.5	22.4	4.0	1.6	0.8
$H=24$	37.2	27.8	15.0	0.5	0.2	0.1
$H=36$	36.1	22.1	10.9	0.3	0.1	0.0
$H=48$	33.3	16.8	8.3	0.0	0.0	0.0

	Model M_M			Model: M_A		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	50.2	45.8	38.8	40.7	28.7	18.1
$H=3$	39.8	35.1	28.3	29.2	17.6	9.7
$H=6$	35.3	31.1	25.5	21.6	10.6	5.2
$H=12$	40.2	34.2	26.1	20.2	8.1	4.1
$H=24$	46.5	38.9	29.4	16.0	6.3	3.1
$H=36$	48.9	41.2	30.5	12.7	5.0	2.5
$H=48$	48.7	41.8	30.4	8.8	3.5	1.7

	Equal Weights Av. Model			AIC Average Model		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	32.8	26.8	18.6	42.4	29.8	19.7
$H=3$	25.4	18.3	11.3	25.5	18.0	11.9
$H=6$	25.8	14.8	7.4	24.8	15.3	7.7
$H=12$	26.6	12.2	6.1	27.4	13.3	6.6
$H=24$	21.2	8.5	4.2	25.4	11.9	6.0
$H=36$	15.8	6.2	3.1	20.9	9.8	4.9
$H=48$	11.9	4.7	2.3	17.0	7.6	3.8

	Model M_{RW}		
	$A=2$	$A=5$	$A=10$
$H=1$	46.7	26.8	13.4
$H=3$	47.4	26.8	13.4
$H=6$	47.8	26.9	13.4
$H=12$	49.8	26.7	13.3
$H=24$	52.1	25.8	12.8
$H=36$	54.3	24.7	12.3
$H=48$	53.4	22.7	11.2

(b) US-Japan

	Model M_E			Model: M_P		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	54.1	50.4	42.1	43.1	37.9	33.7
$H=3$	55.7	47.8	37.2	55.9	46.2	33.7
$H=6$	55.3	46.6	34.7	46.7	40.6	28.2
$H=12$	53.1	42.8	28.9	43.1	35.9	21.7
$H=24$	55.0	37.6	21.2	38.4	21.9	11.4
$H=36$	57.5	33.8	17.3	24.2	12.1	6.0
$H=48$	57.7	29.9	14.7	12.9	5.7	2.8

	Model M_M			Model: M_A		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	27.5	23.7	18.6	51.9	42.5	31.1
$H=3$	18.9	15.0	9.2	48.4	35.8	20.7
$H=6$	9.3	7.2	5.5	54.3	29.6	14.7
$H=12$	3.5	1.8	0.9	53.6	25.5	12.7
$H=24$	0.0	0.0	0.0	55.0	24.5	12.1
$H=36$	0.0	0.0	0.0	54.8	22.9	11.4
$H=48$	6.0	3.0	1.5	52.5	21.0	10.4

	Equal Weights Av. Model			AIC Average Model		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	35.7	30.5	25.3	54.7	47.6	33.7
$H=3$	41.0	32.2	18.6	55.8	42.3	25.7
$H=6$	38.6	24.5	13.1	62.5	39.2	20.3
$H=12$	32.7	18.3	9.2	56.4	33.2	16.6
$H=24$	22.5	10.1	5.0	56.7	29.1	14.4
$H=36$	17.8	7.2	3.6	56.8	26.2	13.0
$H=48$	15.7	6.1	3.0	55.5	23.6	11.7

	Model M_{RW}		
	$A=2$	$A=5$	$A=10$
$H=1$	4.9	1.9	1.0
$H=3$	4.9	1.9	1.0
$H=6$	5.1	2.1	1.0
$H=12$	4.7	1.9	0.9
$H=24$	3.5	1.4	0.7
$H=36$	2.1	0.8	0.4
$H=48$	0.5	0.2	0.1

Notes: The statistics relate to the optimal share held on average across 51 quarterly recursions over the period 1989m12-2002m6.

Table 5: **End-Period Utility Ratios from Home-Overseas Investments**
(relative to random walk model, percentage averages for the recursions)

(a) US-UK

Model M_E				Model: M_P		
$A=2$	$A=5$	$A=10$		$A=2$	$A=5$	$A=10$
1.0009	1.0077	1.0225††	$H=1$	0.9974	0.9925	0.9883
1.0010	1.0186†	1.0715†	$H=3$	0.9971	0.9980	0.9933
1.0016	1.0304	1.1285†	$H=6$	0.9934	0.9837	0.9769
1.0041	1.0424	1.1655†	$H=12$	0.9989	0.9993	0.9990
1.0089	1.0510	1.0666	$H=24$	0.9956	0.9945	0.9934
1.0099	1.0185	1.0233	$H=36$	0.9956	0.9926	0.9930
1.0167	1.0352	1.0414	$H=48$	0.9925	0.9879	0.9869

Model M_M				Model: M_A		
$A=2$	$A=5$	$A=10$		$A=2$	$A=5$	$A=10$
0.9995	1.0009	1.0099	$H=1$	1.0009	1.0047	1.0141††
1.0037†	1.0201†	1.0573†	$H=3$	1.0010	1.0084†	1.0081
1.0057	1.0345†	1.0886†	$H=6$	0.9995	0.9979	0.9979
1.0035	1.0385	1.1071	$H=12$	0.9980	0.9980	0.9974
1.0098	1.0735	1.1601	$H=24$	1.0009	1.0025	1.0024
1.0094	1.0713	1.1993	$H=36$	1.0028	1.0042	1.0062
1.0194	1.1248	1.3783	$H=48$	1.0028	1.0044	1.0058

Equal Weights Av. Model				AIC Average Model		
$A=2$	$A=5$	$A=10$		$A=2$	$A=5$	$A=10$
1.0013	1.0059†	1.0145††	$H=1$	0.9979	1.0008	1.0090
1.0010	1.0106†	1.0208	$H=3$	1.0005	1.0107†	1.0195†
1.0009	1.0020	1.0030	$H=6$	0.9988	1.0003	1.0008
1.0029	1.0073	1.0087	$H=12$	0.9959	0.9970	0.9969
1.0004	1.0025	1.0025	$H=24$	1.0011	1.0187	1.0221
0.9959	0.9935	0.9938†	$H=36$	0.9949	0.9967	0.9984
0.9986	0.9978	0.9982	$H=48$	0.9976	1.0003	1.0003

(b) US-Japan

	Model M_E			Model: M_P		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	1.0022*	1.0087††	1.0133	1.0011†	1.0029	1.0085
$H=3$	0.9985	1.0118†	1.0514†	1.0067**	1.0196*	1.0521††
$H=6$	0.9974	1.0116	1.0716†	0.9988	1.0037	1.0317
$H=12$	0.9967	1.0349*	1.1223††	1.0060††	1.0366*	1.0684††
$H=24$	0.9926	1.0387*	1.1380*	1.0185**	1.0312*	1.0476*
$H=36$	0.9924	1.0149	1.0274	1.0131	1.0186†	1.0221†
$H=48$	0.9890	0.9952	1.0069	1.0207	1.0221	1.0242

	Model M_M			Model: M_A		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	0.9969	0.9930	0.9937	1.0032*	1.0137**	1.0274*
$H=3$	0.9976	0.9908	0.9840	1.0044†	1.0245**	1.0424**
$H=6$	0.9882	0.9697	0.9618	0.9937	0.9836	0.9821
$H=12$	0.9822	0.9524	0.9467	0.9940	0.9806	0.9778
$H=24$	0.9782	0.9412	0.9350	0.9939	0.9758	0.9739
$H=36$	0.9974	0.9768	0.9738	1.0009	0.9846	0.9811
$H=48$	1.0397	1.0603	1.0693	1.0014	0.9848	0.9824

	Equal Weights Av. Model			AIC Average Model		
	$A=2$	$A=5$	$A=10$	$A=2$	$A=5$	$A=10$
$H=1$	0.9989	0.9997	1.0011	1.0031*	1.0149**	1.0241*
$H=3$	1.0038*	1.0164**	1.0233*	1.0022†	1.0255**	1.0443**
$H=6$	0.9998	0.9938	0.9928	0.9954	0.9920	0.9982
$H=12$	1.0023	0.9969	0.9955	0.9962	1.0045	1.0069
$H=24$	1.0099	0.9901	0.9878	0.9950	1.0070†	1.0106
$H=36$	1.0317	1.0230	1.0230	1.0011	0.9975	0.9947
$H=48$	1.0539	1.0642	1.0697	1.0002	0.9920	0.9909

Notes: The statistics are average end-period utility ratios, calculated over 51 recursions, expressed relative to that obtained when modelling is undertaken using a random walk to forecast an exchange rate. Superscripts "**" and "*" indicate significance at the 5% and 10% levels respectively and "† †" and "†" indicate significance at the 20% and 30% levels.