Journal of Forecasting
J. Forecast. **21**, 301–315 (2002)
Published online 29 May 2002 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/for.834

Conditional Predictability of Daily Exchange Rates

DEMOSTHENES N. TAMBAKIS[1]* AND ANNE-SOPHIE VAN ROYEN[2]
[1] *Pembroke College, Cambridge, UK*
[2] *State Street Associates, Boston, USA*

ABSTRACT
At what forecast horizon is one time series more predictable than another? This paper applies the Diebold–Kilian conditional predictability measure to assess the out-of-sample performance of three alternative models of daily GBP/USD and DEM/USD exchange rate returns. Predictability is defined as a non-linear statistic of a model's relative expected losses at short and long forecast horizons, allowing flexible choice of both the estimation procedure and loss function. The long horizon is set to 2 weeks and one month ahead and forecasts evaluated according to MSE loss. Bootstrap methodology is used to estimate the data's conditional predictability using GARCH models. This is then compared to predictability under a random walk and a model using the prediction bias in uncovered interest parity (UIP). We find that both exchange rates are less predictable using GARCH than using a random walk, but they are more predictable using UIP than a random walk. Predictability using GARCH is relatively higher for the 2-weeks-than for the 1-month long forecast horizon. Comparing the results using a random walk to that using UIP reveals 'pockets' of predictability, that is, particular short horizons for which predictability using the random walk exceeds that using UIP, or vice versa. Overall, GBP/USD returns appear more predictable than DEM/USD returns at short horizons.

KEY WORDS conditional predictability; exchange rate forecasting; bootstrap

INTRODUCTION

At what forecast horizon is one time series more predictable than another? Recently, researchers have recognized that predictability is a *conditional* notion, concerning variances conditional on varying information sets. Given a particular information set, predictability is also a function of the model used to generate forecasts, and of the loss function employed for forecast evaluation.[1] Conditional upon the data set and loss function, predictability can then be used to assess the

* Correspondence to: Demosthenes N. Tambakis, Pembroke College, Cambridge, CB2 1RF, UK. E-mail: dnt22@cam.ac.uk
[1] See Diebold (1998), Diebold and Kilian (2001), Granger and Pesaran (1999), Hong and Billings (1999) and Rothemberg and Woodford (1996).

performance of competing models at different forecast horizons through the properties of their predictions.

Granger and Newbold (1986) and Beran (1994) define the predictability of covariance stationary series as the proportion of the unobservable unconditional variance explained by the mean squared error (MSE) of a conditional forecast at a given horizon. Diebold and Kilian (2001)—henceforth DK—generalize this notion and define predictability as inversely related to the ratio of the expected losses of conditionally optimal linear forecasts at 'short' and 'long' forecast horizons. Thus predictability is a relative statistic comparing the evolution of forecasting models' out-of-sample expected losses at two different horizons specified by the modeller. The DK measure has the advantage that the degree of 'long' forecast accuracy is directly observable. 'Short' horizon predictability can then be measured against this benchmark under a particular loss function. DK estimate confidence intervals for the predictability of US quarterly macro series over the post-war period.

This paper applies the DK predictability measure to assess multi-step-ahead dynamic forecasts generated by three different models of daily dollar exchange rate returns of the deutschemark and pound sterling over the period 1988–1998. All models are evaluated under the MSE criterion. The long forecast horizon is set to 2 weeks or one month ahead and the short horizon takes all intermediate values. First, GARCH models are used to compute bias-corrected predictability confidence intervals.[2] As the predictability statistic is non-linear and possibly non-normally distributed, we use the parametric bootstrap to construct confidence intervals for the forecasts generated by the GARCH models. These are then compared to predictability using a random walk and an uncovered interest parity model.

Our motivation is the apparent failure of most parametric and non-parametric exchange rate models to outperform a random walk process at various frequencies.[3] The random walk is a consequence of efficient markets arbitraging away profitable trades; as such it constitutes the appropriate benchmark to use. Conveniently for our purposes, the predictability of a random walk according to MSE is linearly decreasing in the forecast horizon, reflecting the linearly increasing forecast variance. Thus random walk processes are a natural benchmark against which to evaluate predictability using other models.

The main motivation for using uncovered interest parity (UIP) to predict exchange rate returns is our interest in exploiting the stylized empirical violation of UIP, reflecting the presence of risk premia and resulting 'peso problems'. Under the assumption of risk neutrality, or certainty equivalence, UIP reduces to a no-arbitrage relationship. It is often interpreted jointly with the random walk, in the form of rational expectations, as an expression of the weak-form efficient market hypothesis.[4]

The main findings are as follows. First, both DEM/USD and GBP/USD exchange rates returns are less predictable using GARCH than a random walk, but they are more predictable using the violation of UIP than a random walk. This conclusion applies to both the 2-weeks-and 1-month long forecast horizons. Second, the distribution of the bootstrap predictability statistic using GARCH is

[2] Using GARCH to forecast the conditional mean is appropriate if the conditional variance is time-varying. See Bera and Higgins (1995).
[3] On structural models' out-of-sample performance against a random walk see Meese and Rogoff (1983) and Meese (1990). For time series models see Brooks (1997) and Diebold and Nason (1990).
[4] Indeed, Fama (1984) showed that predictable excess returns should then be zero, as the foreign exchange forward premium for a given maturity is an unbiased predictor of the future realized exchange rate return over the same maturity. See also Begg (1984) and Lewis (1995).

negatively skewed, and its median is higher for the 2-weeks-than for the 1-month long horizon for both exchange rates. This suggests that GARCH models for modelling the conditional mean are better suited to shorter forecast horizons. Finally, comparing the evolution of predictability using a random walk to that using the UIP model reveals 'pockets' of predictability, that is, specific short horizons for which predictability using the random walk outperforms predictability using UIP, or vice versa. These appear more prominent for DEM/USD than GBP/USD returns. The latter are relatively more predictable using UIP than a random walk, confirming the findings by Fisher *et al.* (1990) for earlier data. Overall, our results are in line with past research, surveyed in Brooks (1997), indicating that sterling–dollar exchange rate returns are relatively more predictable than deutschemark–dollar returns for short forecast horizons.

The remainder of the paper is arranged as follows. The next section defines the predictability measure and outlining its theoretical properties. The third section summarizes the data and the estimated GARCH models and develops the bootstrap methodology for constructing predictability confidence intervals. The fourth section applies this procedure to the 2-week and one-month-ahead predictability of daily exchange rate returns using GARCH models. Predictability is also computed using a random walk and a UIP model, and implications for exchange rate forecasting are drawn by comparisons across the three models. The final section presents a conclusion.

A REVIEW OF PREDICTABILITY MEASURES

Granger and Newbold (1986) define the predictability of univariate covariance-stationary processes according to MSE loss by analogy to the familiar formula for R^2:

$$G(s, l) = \frac{var(\widehat{y}_{t+s,t})}{var(y_{t+l})} = 1 - \frac{var(e_{t+s,t})}{var(y_{t+l})} \tag{1}$$

where $\widehat{y}_{t+s,t}$ is the optimal linear s-step-ahead forecast of random variable y_t and $e_{t+s,t} = y_{t+s} - \widehat{y}_{t+s,t}$ is the associated forecast error. To the extent that the unconditional variance of y_t appears in the denominator, $G(s, l)$ may be thought of as an absolute predictability measure.[5]

DK extend this definition to non-stationary processes evaluated under general loss functions. Following the intuition of the $G(s, l)$ measure, but taking into account that the unconditional variance is unobservable, the predictability of y_t is defined as being the ratio of the conditionally expected loss of an optimal 'short-run' forecast, $E(L(e_{t+s,t}))$, to that of an optimal 'long-run' forecast, $E(L(e_{t+l,t}))$, where $s < l$. Intuitively, if $E(L(e_{t+s,t})) < E(L(e_{t+l,t}))$ then the process is more predictable at horizon s relative to horizon l. In contrast, if $E(L(e_{t+s,t})) \simeq E(L(e_{t+l,t}))$ the process is almost equally predictable at horizon s relative to l. DK thus define the predictability $P_t(s, l, L, \Omega)$ $(s < l)$ of y_t to be:

$$P_t(L, \Omega_t, s, l) = 1 - \frac{E(L(e_{t+s,t}) \mid \Omega_t)}{E(L(e_{t+l,t}) \mid \Omega_t)} \tag{2}$$

where the information set Ω_t can be either univariate or multivariate and y_t can be either stationary or non-stationary. The choice of short and long horizons s and l is flexible provided l is finite.

[5] The Granger–Newbold measure has been employed in macroeconomics to relate one-step-ahead inflation forecast errors to long-run inflation variability, as in Barsky (1987).

Equivalently, we can assess expected forecast losses for period t made in periods $t-s$ and $t-l$ by comparing the relative accuracy of s and l−step−ahead forecasts of y_t, where $l > s$:

$$P_t(L, \Omega_t, s, l) = 1 - \frac{E_{t-s}(L(e_t))}{E_{t-l}(L(e_t))} \tag{3}$$

For univariate AR(1) processes, $y_t = \rho y_{t-1} + u_t$, the s-step-ahead predictability of white noise ($\rho = 0$) is zero for all s, as short- and long-run forecasts are equal. In contrast, DK show that a dataset's predictability using a univariate random walk according to MSE is just $P_t(MSE, \Omega, s, l) = 1 - s/l$.[6] In general, comparing the predictability of covariance-stationary data at two short term horizons s_1 and s_2 with $s_1 < s_2$, the information set used for s_2-steps−ahead forecasts is likely to be poorer than that for s_1 steps ahead, so we expect predictability to decline as we compare forecasts further into the future.

For covariance-stationary data one can show that $0 < P_t(L, \Omega, s, l) < 1$. However, the predictability of non-stationary data can also be negative, reflecting the fact that short-term expected losses may exceed long-term ones. In general, increasing the short horizon need not imply a lower $P(s, l)$. As our focus is on univariate information sets and MSE loss, we shall simplify notation and write just $P_t(L, \Omega_t, s, l) = P_t(s, l)$. In the remainder of this paper, 'predictability' always refers to predictability of a certain dataset using a particular forecasting model evaluated according to MSE loss.[7]

Let N is the total number of observations, $N - l$ of which are reserved for estimation and l for out-of-sample forecast evaluation. The expected MSE of s-steps-ahead forecasts then is:

$$E_t(L(e_{t+s,t})) = MSE_s = \frac{1}{s} \sum_{i=N-l+1}^{N-l+s} (y_{t+i} - \widehat{y}_{t+i,t})^2 \tag{4}$$

where the realized value of the variable at time t is y_t, and the conditionally optimal s-steps-ahead forecast made in period t is $\widehat{y}_{t+s,t} = E(y_{t+s} \mid \Omega_t)$. Thus, if the predictability of a dataset at short horizon s relative to long horizon l is higher using forecasting model A than alternative model B we may write:

$$P^A_{(s,l)} > P^B_{(s,l)} \Leftrightarrow 1 - \frac{MSE^A_s}{MSE^A_l} > 1 - \frac{MSE^B_s}{MSE^B_l}$$

implying:

$$\frac{MSE^A_s}{MSE^A_l} < \frac{MSE^B_s}{MSE^B_l} \tag{5}$$

It is possible that a dataset's predictability at horizons (s, l) is greater using model A even though its respective MSEs exceed those of model B at both 'short' and 'long' horizons. Therefore, predictability is a *relative* concept: in order to conclude that the data is absolutely more predictable at horizon s, the MSEs at the long horizon also have to be of similar magnitude. We return to this

[6] A random walk without drift is defined as $y_t = y_{t-1} + u_t$, where u_t is serially uncorrelated with constant variance. It satisfies the weak martingale property that expected changes in y_t are zero.
[7] The loss function need not be restricted to the quadratic specification. For the theory and applications of non-quadratic and/or asymmetric loss functions see Diebold and Mariano (1995) and Christoffersen and Diebold (1996).

issue below when assessing the predictability of exchange rate returns using a random walk and uncovered interest parity.

THE DATA AND BOOTSTRAP METHODOLOGY

The data

The information set consists of mid-day spot exchange rates of pound sterling and the deutschemark against the dollar, respectively denoted GBP and DEM, over the period from 1/1/1988 to 7/4/1998. There are $N = 2,678$ log return observations for each exchange rate. Tables I and II (Panel (a)) summarize the data's distributional properties. Augmented Dickey–Fuller (ADF) unit root tests indicate that both return series are covariance-stationary. The in-sample returns variance is much larger than the mean, and skewness is close to zero. Returns are strongly leptokurtic and normality is rejected. Also, the Ljung–Box Q_x statistic is insignificant at 20 lags for DEM returns, but significant at the 1% level for GBP, indicating the presence of linear autocorrelation.

Panels (b) of Tables I and II present the selected GARCH models for the conditional mean of returns. The AR order is determined using the Schwartz Information Criterion (SIC). As this penalizes an overspecified structure more than Akaike's Information Criterion, it is less likely to

Table I. DEM/USD daily returns

(a) Summary statistics[a]	
Mean	0.0000577
Standard deviation	0.0066
Skewness	0.0328
Excess kurtosis	5.0023
Normality	447.83**
ADF(5)	−21.84**
$Q_x(20)$	22.21
(b) Model specification[b]	
AR	11
MA	0
GARCH(p,q)	(1,1)
ARMA errors	No
SIC	−7.2408
$\overline{R^2}$	0.004852
$Q_x(20)$	8.65
ARCH LM (20)	1.46*

Notes:
[a] $N = 2678$ observations: 1/1/1988-7/4/1998. ADF(5) is a unit root test with 5 lags. Normality is the Bera–Jarque test, asymptotically distributed $\chi^2(2)$. $Q_x(20)$ is the Ljung–Box statistic of order 20. ARCH(5) is Engle's LM test for ARCH. * and ** denote 10% and 1% significance levels.
[b] The AR order was obtained using SIC. $\overline{R^2}$ is adjusted R^2. Q_x and ARCH report the Ljung–Box and ARCH LM statistics for the nulls of no linear dependence in the error levels and squares, respectively. * and ** denote 10% and 1% significance levels.

Table II. GBP/USD daily returns

(a) Summary statistics[a]	
Mean	0.0000446
Standard deviation	0.0062
Skewness	0.02780
Excess kurtosis	5.6367
Normality	810.24**
ADF(20)	−20.76**
$Q_x(20)$	46.23**
(b) Model specification[b]	
AR	7
MA	0
GARCH(p,q)	(1,1)
ARMA errors	No
SIC	−7.4370
$\overline{R}^2$	0.005073
$Q_x(20)$	13.58
ARCH LM (20)	1.21*

Notes: See Table I.

overestimate the required number of lags. The lag orders are AR(7) for GBP and AR(8) for DEM. Because of the penalty associated with an overspecified lag order, a GARCH(1,1) specification is always preferred with no MA structure in the errors for both time series. Not surprisingly, the explanatory power is very low. The standard diagnostics indicate linear dependence in both the levels and squares of the errors from the GARCH models.

Bootstrap predictability for GARCH models

Estimation is carried out in two stages. First, *baseline predictability* is computed as follows. After fitting the selected GARCH model to the first $N - l$ return observations, the estimated coefficients are used to generate forecasts for the out-of-sample range: $1 \leq s \leq l$, $l = 10, 22$. These are compared with the respective two–week and one-month out-of-sample observations—10 and 22 days ahead, respectively—and the baseline value of the predictability statistic is computed for all out-of-sample short horizons. In the second stage, the baseline is used to construct bias-corrected confidence intervals for predictability following the five-step parametric bootstrap of Freedman and Peters (1985):[8]

(1) AR(p)-GARCH models are fitted to all N observations using constrained maximum likelihood. Each chosen model is then used to compute the corresponding fitted values of y_t for all $t = p + 1, \ldots, N$ given by $\widehat{y}_t = \mu + \sum_{i=1}^{p} \widehat{\alpha}_i y_{t-i} + e_t$. Subsequently, the set of all observations is used to construct $N - p$ pseudo-data vectors. Each model's true in-sample error vector $\widehat{\mathbf{e}}$ $((N - p) \times 1)$ is the difference between the true vector $\mathbf{y}$ and the conditional fitted values: $\widehat{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}}$.

[8] This procedure, which has also been applied by Masarotto (1990), is reviewed by Berkowitz and Kilian (2000) and Davison and Hinkley (1997), who refer to it as 'post–blackening' of the data.

(2) Resampling with replacement from the empirical distribution of $\widehat{\mathbf{e}}$ yields B bootstrap error vectors $\mathbf{e}^b$ $(b = 1, \ldots, B)$. We use $B = 500$ replications throughout.[9] As discussed above, the levels and squares of GARCH estimated errors are correlated. Moreover, there is a large body of evidence of non-linear dependence in daily exchange rate returns working through the conditional mean and higher moments.[10] Therefore, the i.i.d. requirement is rejected, suggesting that estimation procedures yield inconsistent parameter estimates. Therefore, the issue of bootstrap resampling from a block index becomes important.

We apply the non-overlapping block resampling method of Künsch (1989) so that the bootstrap replicate blocks are asymptotically i.i.d. This is done by partitioning each true error vector into blocks of fixed size and resampling with replacement from the block index. Intuitively, a very large number of blocks would destroy any error autocorrelation in the true data; a moderate number is better suited. We choose a block size of 25 days throughout.[11]

(3) Each bootstrap error vector is then used to construct a pseudo-data vector $\mathbf{y}^b$ of dimension $N - p$ using the true coefficients $(\widehat{\boldsymbol{\alpha}})$ for each model. In this way we construct B pseudo-data vectors are constructed, the components of which are given by $y_t^b = \mu + \sum_{i=1}^{p} \widehat{\alpha}_i y_{t-i}^b + e_t^b$, $t = p + 1, \ldots, N$, $b = 1, \ldots, B$. Each pseudo-data vector is evaluated recursively, with the first p observations corresponding to the true returns data. These initial values should not affect the pseudo-data distribution for large sample sizes.

(4) $N - l - p$ of the $N - p$ pseudo-sample returns observations are reserved for in-sample estimation and the last l for out-of-sample forecast evaluation. Step (1) is then repeated for each pseudo-data vector using the first $(N - l - p)$ observations. The AR coefficient vector $\boldsymbol{\alpha}^b$ is estimated from $y_t^b = \mu^b + \sum_{i=1}^{p} \alpha_i^b y_{t-i}^b + e_t^{*b}$, while the out-of-sample values are the 'true' pseudo-data for each bootstrap replication.[12]

(5) Finally, for each pseudo-data vector we generate a dynamic pseudo-forecast for each short forecast horizon $(s = 1, \ldots, l)$. These are given by $y_{t+s}^{*b} = \mu^b + \sum_{i=1}^{p} \alpha_i^b y_{t+s-i}^{*b} + e_{t+s}^{*b}$, with $e_{t+s}^{*b} = 0$ for all s, and all unknown future errors set to their conditional expectation of zero. The expected MSE loss at short horizon s then is:

$$MSE_s^b = \frac{1}{s} \sum_{t=N-p-l+s}^{N-p} (y_{t+s}^b - y_{t+s,t}^{*b})^2 \tag{6}$$

Conditional predictability $\widehat{P}(s, l)$ is obtained by computing 1 minus the ratio of expected losses at the short and long forecast horizons s and l for each pseudo-data vector and averaging the outcome over all B bootstrap replications:

$$\widehat{P}(s, l) = 1 - \frac{E(L(e_{t+s}^b))}{E(L(e_{t+s}^b))} = 1 - \frac{MSE_s^b}{MSE_l^b} \tag{7}$$

[9] Bootstrap statistics can be estimated consistently using least squares provided the model residuals used for resampling are i.i.d. For a detailed treatment see Efron and Tibshirani (1993) and Davison and Hinkley (1997).

[10] Brock, Hsieh and LeBaron (1993) and Mills (1996) document non-linear dependence in various daily exchange rates.

[11] The results can be sensitive to changes in the block size: see Aczel and Josephy (1992). However, the non-linearity of the predictability statistic precludes using the block size rule of Hall, Horowitz and Jing (1995). An extension would involve the stationary bootstrap of Politis and Romano (1994) which uses a random block size.

[12] Due to computational constraints, the AR order p for each pseudo-data estimation is assumed to be the same as that used for the true data.

Confidence intervals (CIs) for conditional predictability can be simply constructed from the empirical percentiles of the bootstrap distribution. However, they are not reported here as they are very wide for both exchange rates.[13] A generic problem with the percentile CIs is their failure to account for the possibility of bias in the bootstrap distribution, thus resulting in worse coverage accuracy. This shortcoming can be addressed by using Efron and Tibshirani's (1993) bias-corrected confidence intervals (BCCIs), which adjust the endpoints of the bootstrap percentile confidence intervals to correct for possible bias. In contrast to t-percentile confidence intervals, BCCIs correct for the fact that the empirical distribution of the predictability statistic is not symmetric around baseline predictability—in this way, they have improved effective coverage.[14]

This is done by comparing the predictability of each bootstrap replication to the baseline predictability statistic from the first stage. The bias of the standard bootstrap CIs is then defined as the proportion of bootstrap replications which are less than the value of baseline predictability. In this way, the latter serves as a benchmark for bias evaluation: if a large number of bootstrap statistics is less (greater) than the baseline, the CI bounds are shifted to the right (left) by an appropriate amount to account for the bias. The correct sign of adjustment can also be inferred by the sample skewness of the bootstrap distribution, as discussed below.

RESULTS AND DISCUSSION

Predictability using GARCH versus a random walk

The pure random walk process (without drift) is arguably the most popular benchmark for assessing the out-of-sample performance of exchange rate return forecasting models. Figures 1 and 2 present the 2.5% and 97.5% bias-corrected confidence bounds of DEM and GBP predictability using the GARCH models, along with the sample mean and median of the bootstrap predictability statistic's distribution. These are shown together with exchange rate returns predictability using a random walk process, which is declining linearly in the short-run horizon from equation (3). Predictability is evaluated according to MSE, and Panels A and B set the long forecast horizon to 2 weeks ahead ($l = 10$ days) and one month ahead ($l = 22$ days), respectively. Note that $\widehat{P}(s, l)$ is bounded in [0,1], so changes in magnitude can be interpreted in percentage terms.

Figure 1 considers DEM returns. On the one hand, the sample mean of the bootstrap predictability distribution is well below predictability under a random walk. On the other hand, the sample median lies well above the mean, implying that the bootstrap predictability distribution is negatively skewed. This suggests that the predictability of a proportion of bootstrap replications lies below the baseline predictability statistic. As discussed earlier, such negative bias in the bootstrap predictability distribution implies that the BC confidence intervals have to shift to the right compared to the standard bootstrap percentiles. However, the BCCIs of GARCH predictability are very wide despite the adjustment: the 2.5% lower bound is negative throughout and the 97.5% upper bound exceeds random walk predictability.

At the 2-week-long horizon (Panel A), median GARCH predictability also exceeds random walk predictability, by an amount of nearly 0.25 for 4-days-ahead and 8-days-ahead forecasts. At the one-month long horizon (Panel B), the median of GARCH predictability is declining almost linearly

[13] Normal and Student-t confidence intervals are also not reported pending results on the asymptotic normality of $\widehat{P}$.
[14] Endogenous AR order selection for each bootstrap, as described in Kilian (1998) and/or the jacknife-accelerated confidence intervals of Efron and Tibshirani (1993), would further improve CI coverage.

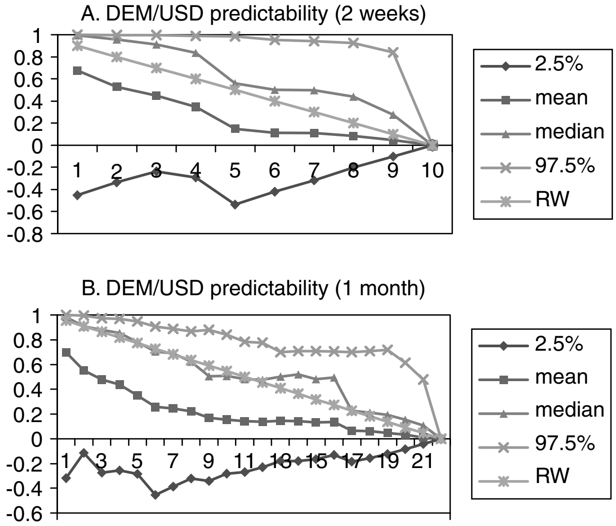

Figure 1. Conditional predictability: GARCH versus random walk. MSE predictability of daily DEM/USD spot returns using GARCH ($N = 2{,}678$) is based on 500 bootstrap replications. Mean and median refer to the bootstrap predictability distribution, and 2.5% and 97.5% are the bounds of the 95% BC confidence intervals. RW is MSE predictability using a random walk. The long forecast horizon is $l = 10$ (Panel A) and $l = 22$ (Panel B) days-ahead. The short forecast horizon varies from 1 to l

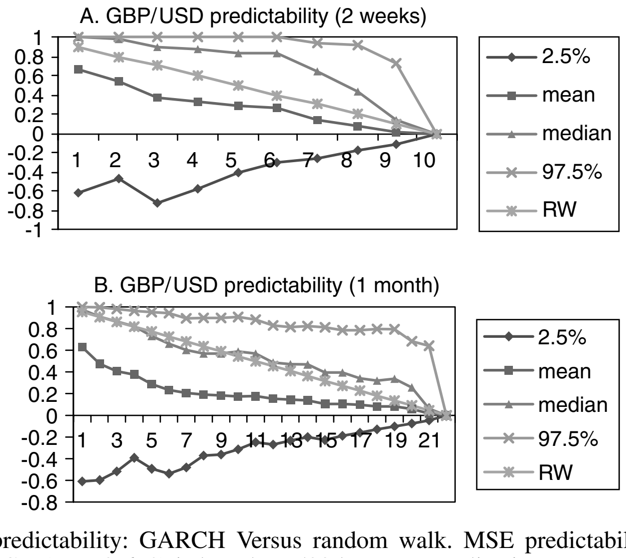

Figure 2. Conditional predictability: GARCH Versus random walk. MSE predictability of daily GBP/USD spot returns using GARCH ($N = 2{,}678$) is based on 500 bootstrap replications. Mean and median refer to the bootstrap predictability distribution, and 2.5% and 97.5% are the bounds of the 95% BC confidence intervals. RW is MSE predictability using a random walk. The long forecast horizon is $l = 10$ (Panel A) and $l = 22$ (Panel B) days-ahead. The short forecast horizon varies from 1 to l

alongside predictability using a random walk. It is only for 13-up to 16-days-ahead forecasts that median GARCH predictability exceeds random walk predictability by an average 0.16. Note also that mean GARCH predictability is less than random walk predictability for both long horizons.

Turning to Figure 2, the median GARCH predictability of GBP returns is similar to that of DEM returns in that it lies above the mean, reflecting the negative skewness of the bootstrap predictability distribution. Mean GARCH predictability is again lower than random walk predictability, especially at the 1-month long horizon (Panel B). In that case, median GARCH predictability is closely tracking predictability using a random walk. At the 2-week long horizon (Panel A), median GARCH predictability is above random walk predictability by almost 0.4 for 6-days-ahead forecasts.

Overall, the results on predictability according to MSE using the GARCH models indicate that GBP returns are more predictable than DEM returns. This is especially clear looking at the evolution of the median rather than the mean of the bootstrap predictability distribution. Figures 1 and 2 also suggest that the width of the 95% BCCIs for DEM returns is somewhat narrower than those for the GBP returns. We now turn to assess the data's predictability using GARCH to that using a model based on the stylized violation of the uncovered interest parity condition.

Predictability using UIP versus a random walk

A parametric alternative for forecasting exchange rate returns is based on the uncovered interest parity relationship (UIP):[15]

$$E_t s_{t+1} - s_t = \alpha_0 + \alpha_1 (r_t - r_t^*) + \varepsilon_t \tag{8}$$

r_t and r_t^* are the respective home (UK, Germany) and foreign (USA) overnight eurodeposit rates. Because the interest differential time series may be non-stationary for time periods as long as 10 years, we extend the sample period for r_t and r_t^* to 18 years (1/1/80–7/4/98, or 4763 return observations). Over the extended historical period the overnight interest differentials are stationary, so OLS estimation is appropriate.[16] Substituting $s_{t+1} = E_t s_{t+1}$ for all in-sample DEM/USD observations and regressing the 1-day-ahead spot returns upon the current overnight interest rate differentials yields:

$$s_{t+1}^{DEM} - s_t^{DEM} = \underset{(-0.42)}{-0.0000458} - \underset{(-1.32)}{0.0000384} \cdot (r_t^{DEM} - r_t^{USD}) \tag{9}$$

$$\overline{R}^2 = 0.0002, \ \text{DW} = 1.97$$

while for GBP/USD returns we have:

$$s_{t+1}^{GBP} - s_t^{GBP} = \underset{(2.25)}{0.000272} - \underset{(-2.79)}{0.0000874} \cdot (r_t^{GBP} - r_t^{USD}) \tag{10}$$

$$\overline{R}^2 = 0.014, \ \text{DW} = 1.86$$

[15] This definition uses a logarithmic approximation. Without approximation in terms of the levels of the actual and expected future spot rates, UIP says that

$$S_t / E_t S_{t+1} = 1 + \frac{r_t - r_t^*}{100}$$

The approximation is adequate for small rates of change such as those observed at daily frequency.

[16] The results of the ADF tests are available from the authors upon request.

The t-statistics are in parentheses. If UIP held, we should expect to find $\alpha_0 = 0$ and $\alpha_1 = 1$. The actual negative sign of the slope coefficients in both regressions reflects the stylized violation of uncovered interest parity: a positive interest differential in favour of the home currency induces it to appreciate against the dollar. Various explanations have been proposed for this finding, including time-varying risk premia, habit persistence, 'peso problems' and irrational expectations.[17] Importantly, the sign of the violation in equations (9) and (10) is robust to the sample size and the location of the cut–off point. However, whereas the magnitude of the GBP/USD coefficients is significant regardless of sample size, the significance of the DEM/USD coefficients increases gradually from about 10 years of data onwards until it peaks at the reported t-statistic of -1.32 for the chosen sample size of 18 years. This can be viewed as another example of the greater difficulty involved in forecasting the DEM/USD spot exchange rate.

The empirical violation of UIP can be exploited for predicting the future spot exchange rate. The estimated regression coefficients are used to obtain forecasts of exchange rate returns by repeatedly computing 1-day-ahead forecasts. In order to compare UIP predictability to that using the other two models the forecasts have to be dynamic: we therefore specify that the overnight interest rate differential follows a random walk out-of-sample. This is a reasonable assumption based on the underlying weak-form efficiency of interest rate markets at short horizons. By the law of iterated expectations, the interest differential's s-days-ahead conditional expectation is fixed at its last in-sample value:

$$E_t(i_{t+s} - i^*_{t+s} \mid \Omega_t) = i_t - i^*_t \tag{11}$$

Substituting the first expected future spot rate into (8) and applying (11) to obtain the 1-day-ahead interest differential yields the 2-days-ahead expected spot rate. Substituting the last 1-day-ahead spot rate forecast for the actual future spot rate yields the next forecast. Iteration continues until all out-of-sample spot rates ($1 \le s \le l$) have been forecast. These are evaluated against the true out-of-sample observations and baseline UIP predictability calculated for each short horizon.

Figures 3 and 4 compare the predictability of the exchange rate returns using a random walk to that using UIP. Note that, in contrast to the random walk, the evolution of UIP predictability is declining non-monotonically from 1 to 0. Its evolution compared to random walk predictability is similar for the two long forecast horizons. The predictability of DEM returns using UIP declines at first and then stabilizes at about 0.4 for the 2-weeks and 0.2 for the one-month forecast horizon. These minimum predictability levels occur at 4- and 9-day-ahead forecasts, respectively. In contrast, the predictability of GBP returns using UIP is almost always higher than predictability using a random walk.

Comparing the results with those in Figures 1 and 2 suggests that, in most cases, UIP predictability fluctuates between the mean and the median of GARCH predictability. UIP predictability of GBP returns is closer to the higher GARCH median than GARCH mean predictability level, and vice versa for DEM. In fact, UIP predictability of GBP for the one-month long horizon exceeds the GARCH median and follows the upper (97.5) confidence bound of the GARCH predictability distribution through to 10-days-ahead. Thus, it appears that forecasting using the UIP model yields higher overall predictability for GBP than for DEM returns.

We also note that DEM returns appear to display specific short forecast horizons at which each model yields relatively higher predictability. In particular, returns for the 2-weeks long horizon are

[17] Lewis (1995) presents a comprehensive survey of the empirical evidence on this and other puzzles in international finance.

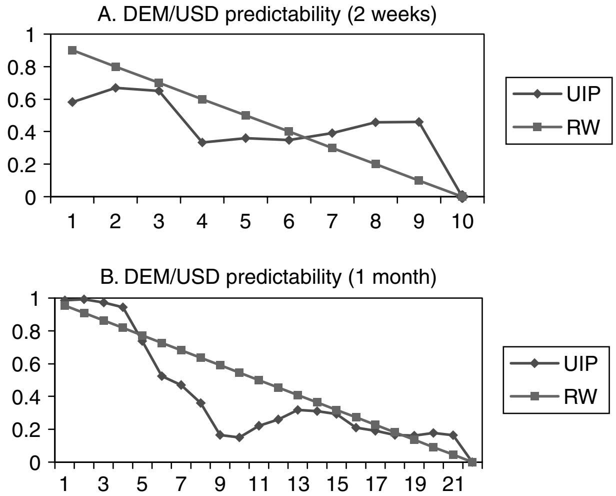

Figure 3. Conditional predictability: UIP versus random walk. MSE predictability of daily DEM/USD spot returns using uncovered interest rate parity (UIP, $N = 4{,}763$) and a random walk (RW). The long forecast horizon is $l = 10$ (Panel A) and $l = 22$ (Panel B) days-ahead and the short forecast horizon varies from 1 to l

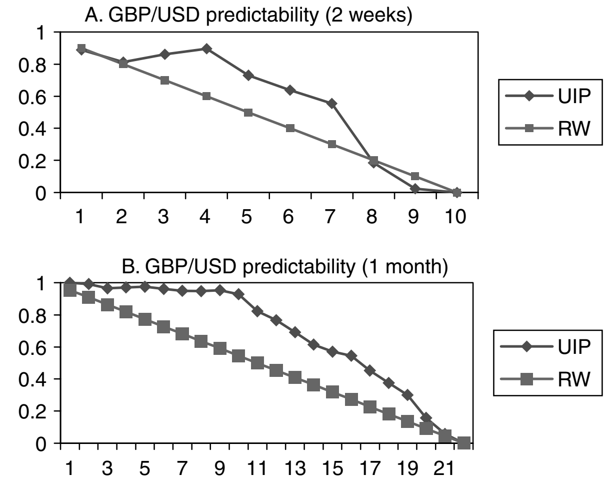

Figure 4. Conditional predictability: UIP versus random walk. MSE predictability of daily GBD/USD spot returns using uncovered interest rate parity (UIP, $N = 4{,}763$) and a random walk (RW). The long forecast horizon is $l = 10$ (Panel A) and $l = 22$ (Panel B) days-ahead and the short forecast horizon varies from 1 to l

more predictable using a random walk for up to 6 days ahead, and then using UIP from 7 to 9 days ahead. For returns at the one-month long horizon, the random walk yields higher predictability for all short horizons except the first 4 and the last 2 days ahead. In contrast, GBP returns for the 2-weeks long horizon are more predictable using UIP up to 8 days ahead, while for the one-month long horizon UIP always yields higher predictability than the random walk.

Identifying these 'pockets' of predictability for a given forecasting model could be a salient feature of its relative predictive performance against a random walk. Clearly, the significance of such pockets is sensitive to the forecaster's choices of loss function and length of the short and long forecast horizons. It should, however, be emphasized that the predictability statistic is only a relative indication of a model's out-of-sample properties. A necessary condition for an absolute comparison of predictability using two different models is that their expected losses at the long horizon are of similar magnitudes.

CONCLUSION

This paper applied the Diebold and Kilian (2001) conditional measure of time series predictability to DEM and GBP daily dollar exchange rate returns over a 10-year period. The predictability statistic enables a concrete evaluation of competing forecasting models' out-of-sample performance at different horizons. The forecasts generated by simple random walk, GARCH(1,1) and interest parity models were evaluated according to MSE loss. Because of the statistic's non-linearity bootstrap confidence intervals were used. It was found, first, that both exchange rate returns are less predictable using GARCH than using a random walk, but they are more predictable using the UIP model. Second, predictability using GARCH is higher for the 2-weeks-than for the one–month–ahead long horizon. This finding could be interpreted as *prima facie* evidence that GARCH models for modeling the conditional mean are better suited to a 2-week horizon. Third, comparing the evolution of the data's predictability using a random walk to that using UIP revealed 'pockets' of predictability, that is, specific short horizons for which the random walk outperforms the UIP model and vice versa. These were particularly prominent for DEM returns. Our results also confirmed earlier findings that GBP exchange rate returns are more predictable than DEM returns at short horizons.

The present framework can be extended to encompass a wider range of forecasting models and estimation procedures—our choice of simple GARCH, UIP and benchmark random walk was driven by the need for parsimony and to illustrate the underlying methodology. In particular, sensitivity analyses can be carried out by varying the in- and out-of-sample periods and/or by changing the length of the dataset. Confidence intervals for predictability under the UIP model could also be constructed along the same principles as those used for the GARCH model.

ACKNOWLEDGEMENTS

We wish to thank an anonymous referee, Alex Michaelides, Mark Salmon and participants at the Computational Finance 1999 (New York) and Forecasting Financial Markets 1999 (London) conferences for helpful comments. All errors remain our own.

REFERENCES

Aczel AD, Josephy NH. 1992. Using the bootstrap for improved ARIMA model identification. *Journal of Forecasting* **11**: 71–80.

Barsky RB. 1987. The Fisher hypothesis and the forecastability and persistence of inflation. *Journal of Monetary Economics* **19**: 3–24.

Begg DK. 1984. Rational expectations and bond pricing: modelling the term structure with and without certainty equivalence. *The Economic Journal* **94**: 45–58.

Bera A, Higgins M. 1995. On ARCH models: properties, estimation and testing. In *Surveys in Econometrics*, Oxley L, George D, Roberts C, Sayer S (eds). Blackwell: Oxford.

Beran J. 1994. *Statistics for Long-Memory Processes*. Chapman and Hall: New York.

Berkowitz J, Kilian L. 2000. Recent developments in bootstrapping time series. *Econometric Reviews* **19**: 1–54.

Brock A, Hsieh D, LeBaron B. 1993. *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*. The MIT Press: Cambridge, MA.

Brooks C. 1997. Linear and non-linear (non-)forecastability of high-frequency exchange rates. *Journal of Forecasting* **16**: 125–145.

Christoffersen P, Diebold FX. 1996. Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics* **11**: 561–572.

Davison AC, Hinkley DV. 1997. *Bootstrap Methods and their Application*. Cambridge University Press: Cambridge.

Diebold FX. 1998. The past, present and future of macroeconomic forecasting. *Journal of Economic Perspectives* **12**: 175–192.

Diebold FX, Kilian L. 2001. Measuring predictability: theory and macroeconomic applications. *Journal of Applied Econometrics* **16**: 657–669.

Diebold FX, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Diebold FX, Nason J. 1990. Nonparametric exchange rate prediction? *Journal of International Economics* **28**: 315–332.

Efron B, Tibshirani R. 1993. *An Introduction to the Bootstrap*. Chapman and Hall: New York.

Fama D. 1984. Forward and spot exchange rates. *Journal of Monetary Economics* **14**: 319–338.

Fisher PG, Tanna SK, Turner DS, Wallis KF, Whitley JD. 1990. Econometric evaluation of the exchange rate in models of the UK economy. *The Economic Journal* **100**: 1230–1244.

Freedman D, Peters S. 1985. Using the bootstrap to evaluate forecasting equations. *Journal of Forecasting* **4**: 251–262.

Granger CWJ, Newbold P. 1986. *Forecasting Economic Time Series*. Academic Press: Orlando, FL.

Granger CWJ, Pesaran MH. 1999. Economic and statistical measures of forecast accuracy. University of Cambridge, Department of Applied Economics working paper 9910.

Hall P, Horowitz JL, Jing B-Y. 1995. On blocking rules for the bootstrap with dependent data. *Biometrica* **82**: 561–574.

Hong X, Billings SA. 1999. Time series multistep-ahead predictability estimation and ranking. *Journal of Forecasting* **18**: 139–149.

Kilian L. 1998. Accounting for lag order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis* **19**: 531–548.

Künsch H. 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**: 1217–1241.

Lewis KK. 1995. Puzzles in international financial markets. In *Handbook of International Economics*, Vol. 3, Grossman G, Rogoff K (eds). North-Holland: Amsterdam.

Masarotto G. 1990. Bootstrap prediction intervals for autoregressions. *International Journal of Forecasting* **6**: 229–239.

Meese R. 1990. Currency fluctuations in the post Bretton-Woods era. *Journal of Economic Perspectives* **4**: 117–134.

Meese R, Rogoff K. 1983. Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics* **14**: 3–24.

Mills T. 1996. Non-linear forecasting of financial time series: an overview and some new models. *Journal of Forecasting* **15**: 127–135.

Politis D, Romano J. 1994. The stationary bootstrap. *Journal of the American Statistical Association* **89**: 1303–1313.

Rothemberg J, Woodford M. 1996. Real business-cycle models and the forecastable movements in output, hours and consumption. *American Economic Review* **86**: 71–89.

Authors' biographies:

Demosthenes N. Tambakis is Fellow in economics and finance at Pembroke College, Cambridge and Research Associate at the Cambridge Endowment for Research in Finance (CERF). He was previously at City University Business School in London, the International Monetary Fund and Princeton University.

Anne-Sophie Van Royen is senior quantitative analyst at State Street Associates.

Authors' addresses:

Demosthenes N. Tambakis, Pembroke College, Cambridge, CB2 IRF, UK.

Anne-Sophie Van Royen, State Street Associates, Boston, MA 02138, USA.

