Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models

Forthcoming, Journal of Banking and Finance

Shiuyan Pong[a], Mark B. Shackleton[a],

Stephen J. Taylor[a,*] and Xinzhong Xu[b]

15 October 2003

Abstract

We compare forecasts of the realized volatility of the pound, mark and yen exchange rates against the dollar, calculated from intraday rates, over horizons ranging from one day to three months. Our forecasts are obtained from a short memory ARMA model, a long memory ARFIMA model, a GARCH model and option implied volatilities. We find intraday rates provide the most accurate forecasts for the one-day and one-week forecast horizons while implied volatilities are at least as accurate as the historical forecasts for the one-month and three-month horizons. The superior accuracy of the historical forecasts, relative to implied volatilities, comes from the use of high frequency returns, and not from a long memory specification. We find significant incremental information in historical forecasts, beyond the implied volatility information, for forecast horizons up to one week.

JEL Codes: C22; C53; G13; G14
Keywords: Realized volatility, Fractional integration, Forecasting, Implied volatilities, Exchange rates

a Department of Accounting and Finance, Lancaster University, England LA1 4YX

b Guanghua School of Management, Peking University, China

* Corresponding author, tel. +44 1524 593624, fax +44 1524 847321, e-mail S.Taylor@lancaster.ac.uk

1 Introduction

The future volatility of an asset's prices can be forecast using historical information about those prices and/or the information provided by option prices. In addition to all historical information, option traders have other information about future events that may be relevant. As a result, option implied volatilities are potentially the more accurate forecasts.

We compare *four* methods for forecasting volatility in this paper for the first time. The methods produce short memory (ARMA) and long memory (ARFIMA) forecasts from intraday returns, GARCH forecasts from daily returns and implied volatilities from option prices. We compare forecasts of volatility for three exchange rates (pound/\$, mark/\$ and yen/\$) over the forecast horizons of one day, one week, one month and three months. We find intraday exchange rates provide more accurate forecasts than implied volatilities for the one-day and one-week forecast horizons while implied volatilities are at least as accurate as the forecasts from intraday data for the one-month and three-month horizons. We show that the enhanced performance of the historical forecasts, relative to implied volatilities, comes from the use of high frequency returns, and not from the use of a long memory model.

Many studies that use daily prices favor the conclusion that option prices provide more accurate forecasts than historical information. Furthermore, these studies involving low-frequency data often find that all the relevant information for volatility prediction is in option prices. See, for example, Jorion (1995) and Xu and Taylor (1995) for foreign exchange, Christensen and Prabhala (1998) and Fleming (1998) for U.S. equity indices and the survey papers by Figlewski (1997) and Poon and Granger (2003). Recent research has emphasized that the additional historical information in intraday prices can be used to produce volatility forecasts of higher accuracy. *Realized volatility*, defined as the sum of intraday squared returns, provides a more accurate estimate of the latent process that defines volatility than is given by daily squared returns (Andersen and Bollerslev, 1998). The theoretical and empirical properties of realized volatility are derived in Andersen, Bollerslev, Diebold and Labys (2001) for foreign exchange. Further empirical evidence is provided in Andersen, Bollerslev, Diebold and Ebens (2001) for U.S. equities. We apply some of the methodology of Andersen *et al* in this paper and refer, as appropriate, to their work as ABDL and ABDE. Related research into the econometric properties of realized volatility includes Barndorff-Nielsen and Shephard (2001, 2002a, 2002b) and Areal and Taylor (2002).

There are, at present, few comparisons of the forecast accuracy of option prices and the high-frequency challenger called realized volatility. Taylor and Xu (1997) find volatility information in five-minute Mark/$ returns incremental to option implied volatilities, when forecasting volatility one-hour ahead. On the other hand, Blair, Poon and Taylor (2001) claim almost all the useful predictive information is in option prices when forecasting S & P 100 volatility one or more days into the future. There is, however, a further way to attempt to improve historical information forecasts. Realized volatility behaves like a *long memory* process (ABDL, 2001, ABDE, 2001) and this feature can be used to construct potentially more accurate forecasts. These forecasts are evaluated in ABDL (2003) for mark/$ and yen/$ rates, but these authors do not make the critical comparison with forecasts from option prices. We make these comparisons here. They are also made in independent and concurrent research by Li (2002) for exchange rates and by Martens and Zein (2002) for the S&P 500 index, the yen/$ rate and crude oil prices. These other authors find incremental information in long memory forecasts.

Despite the empirical appeal of long memory models for realized volatility, analysis of long memory forecasts is insufficient when evaluating historical forecasts. It is possible that the forecasting improvements attributed to long memory forecasts are merely a consequence of using high frequency data. It is therefore essential to also consider forecasts from *short memory* models, applied to high frequency data, to provide a complete forecasting competition. We include forecasts from the ARMA(2,1) model in our comparisons. Our investigation of the ARMA(2,1) model is motivated by the research of Gallant, Hsu and Tauchen (1999), who show that the sum of two AR(1) processes is capable of capturing the persistent nature of asset price volatility, and of Alizadeh, Brandt and Diebold (2002), who show that the sum of two AR(1) processes describes FX volatility better than one AR(1) process. The sum of two AR(1) processes can always be represented by an ARMA(2,1) model (Granger and Newbold, 1976). Interestingly and importantly, our empirical results do show that the short memory ARMA(2,1) model is as good as long memory ARFIMA models when forecasting future volatilities. Thus we show that the accuracy of historical forecasts can be enhanced by using high-frequency returns rather than by the selection of a long memory model.

Section 2 describes the high-frequency data that we use and the properties of daily realized volatility. Using five-minute price changes for the exchange rates from 1987 to 1998, we confirm that the logarithm of realized volatility has the major properties identified by ABDL : it is approximately Gaussian and highly persistent. Short and long memory AR(FI)MA models are estimated for realized volatility, by maximizing the likelihood function in the frequency domain.

Section 3 describes the methods for obtaining forecasts of realized volatility. It also includes a description of the implied volatilities used in this study. These are one and three month at-the-money values supplied by a bank. Section 4 compares the forecasts using several criteria. We discuss their accuracy (defined as the mean squared forecast error), followed by analysis of regressions of realized volatility on one or two forecasts. Conclusions are presented in Section 5.

2 Modeling realized volatility

One of the important issues in volatility forecasting is identifying a suitable proxy for realized volatility. Absolute daily returns and squared daily returns have been chosen as indicators of daily volatility in many previous studies. However, Andersen and Bollerslev (1998) emphasize that both measures are noisy estimators for daily volatility. They show that the sum of intraday squared returns is a closer proxy for daily volatility than either absolute or squared daily returns, providing that sampling is sufficiently frequent. By making use of the theory of quadratic variation and arbitrage-free processes, ABDL (2001, 2003) provide theoretical justification for the construction of realized volatility from high-frequency intraday returns.

2.1 High-frequency data

The exchange rates used in our study are provided by Olsen and Associates. The raw data is in the form of five-minute exchange rate differences computed using mid-quote exchange rates.[1] The sample period is from July 1, 1987 to December 31, 1998. All returns are calculated as differences in the logarithm of the exchange rate.

---

[1] The mid-quote price at a specific time point is estimated by linear interpolation between the average of the bid and ask quote just before and after that particular point in time.

Trading volume and the volatility of the market diminish significantly during weekends and holidays as shown, for example, by Bollerslev and Domowitz (1993). To eliminate the weekend effect, we exclude the period from 21:00 GMT Friday to 21:00 GMT Sunday from the calculations. We also exclude several holidays: Memorial Day, Independence Day, Labour Day and Thanksgiving. In addition, the returns during Christmas (24/12-26/12), New Year (31/12-01/01), Good Friday, Easter Monday and the day after Thanksgiving are also removed. [2]

We partition the sample period into a six-and-a-half-year period for in-sample parameter estimation, from July 1987 to December 1993, and a five-year period for out-of-sample forecasting, from January 1994 to December 1998.

2.2 Construction of realized volatility

Realized volatility is constructed by summing the squared intraday returns sampled at a particular frequency. ABDL (2001, 2003) have shown that as sampling becomes more frequent the realized volatility is an increasingly accurate measure of the integrated return volatility, when certain regularity conditions apply. This could suggest that the sampling frequency should be chosen as high as possible. However, bid-ask bounce is a market microstructure feature that hinders the use of the highest frequencies. The negative serial correlation induced by bid-ask bounce is greater when a higher sampling frequency is employed. A higher sampling frequency also increases

[2] After the returns for all the weekends and specified holidays are removed, there remain several long sequences of constant returns in the five-minute return series. Detailed analysis of these sequences showed that the only plausible explanation is missing quotations. We removed the days in which constant returns are found for eight or more hours. Similar features are observed by ABDL (2001) who use the same dataset.

the impact of interpolation upon volatility calculations. These factors can distort the quality of the constructed realized volatility as a proxy for true volatility.

The optimal frequency for constructing realized volatility is unknown and a variety of frequencies have been tried by ABDL/E. To obtain robust conclusions, we construct volatility series using two sampling frequencies, five minutes and thirty minutes, that represent the most popular choices in previous studies. Daily volatility series constructed from five-minute returns are here named five-minute series and the same convention applies to the thirty-minute returns. To conserve space, our discussion mainly focuses on the empirical results for the five-minute series unless the thirty-minute series has different properties.

The realized variance for day t is defined by

$$s_t^2 = \sum_{j=1}^{n} r_{t,j}^2 \tag{1}$$

where $r_{t,j}$ is the return in interval j on day t and n is the number of intervals in a day. For the five-minute sampling frequency, n equals 288 while n equals 48 for the thirty-minute sampling frequency. The realized standard deviation, s_t, is hereafter called realized volatility, and is used to define the forecast target in our study. We choose to commence the first interval at 12:00 local time in London because our implied volatility quotes, described in Section 3.2, are the last available quotes before 12:00 London time.

2.3 Distribution of daily realized volatility

The distributional characteristics of the realized volatility series for the in-sample period are shown in Table 1 for the five-minute series. The values of skewness and kurtosis indicate a highly skewed and leptokurtic distribution for all three exchange rates; these statistics are even larger for the thirty-minute series. The Ljung-Box test statistics show that there is substantial serial correlation among realized volatility, with the magnitudes of the autocorrelations from one to five lags all highly significant.

Taking the logarithms of these realized volatilities produces distributions that are approximately Gaussian. Table 1 shows that the skewness and kurtosis of realized logarithmic volatility, $\ln(s_t)$, are close to the Normal values of zero and three respectively.

The autocorrelations of log realized volatility are characterized by a slow decay.[3] While the patterns are similar for both frequencies, the autocorrelations of the five-minute series are always higher than those of the thirty-minute series at the same lag. This reflects a more accurate measurement of the latent volatility process when the higher frequency is used.

2.4 AR(FI)MA models

The autocovariance features of log realized volatility suggest that its dynamics may be represented by a *long memory* process. Granger and Joyeux (1980) and Hosking (1981)

[3] Similar decay patterns are also found for the logarithm of realized volatility by ABDL (2001) for exchange rate markets and by ABDE (2001) and Areal and Taylor (2002) for stock markets.

introduced a flexible class of long memory processes, called autoregressive, fractionally integrated, moving average models. For an excellent survey of long memory processes, including applications to financial data, see Baillie (1996). The ARFIMA(*p*, *d*, *q*) model for a process y_t is defined by

$$f(L)(1-L)^d (y_t - m) = q(L)e_t \qquad (2)$$

where d is the order of fractional integration, L is the lag operator, $f(L) = 1 + f_1 L + ... + f_p L^p$ is the polynomial that defines the autoregressive component, $q(L) = 1 + q_1 L + ... + q_q L^q$ is the moving average polynomial and m is the expectation of y_t. The roots of $f(L)$ and $q(L)$ lie outside the unit circle and e_t is a zero-mean white noise process whose variance is here denoted by x^2. If $0 < d < 1$, the process has a long memory property; it is covariance stationary when $d < 0.5$. The spectral density function of an ARFIMA(*p*, *d*, *q*) model is given by Baillie (1996), and is

$$f_{arfima}(w) = (\frac{x^2}{2p}) |q(e^{-iw})|^2 |f(e^{-iw})|^{-2} |1 - e^{-iw}|^{-2d}, \qquad 0 < w < 2p. \qquad (3)$$

Gallant, Hsu and Tauchen (1999) show that the sum of a particular pair of AR(1) models is an alternative to long memory that is also able to explain the highly persistent characteristic of volatility. Alizadeh, Brandt and Diebold (2002) find that the sum of two AR(1) models describes FX volatility better than the AR(1) model, with an appropriate sum composed of a persistent component and a transient component. Any process formed by adding two AR(1) models is an ARMA(2, 1) model, which is categorized as having a *short memory*. For this reason, the ARMA(2, 1) model is also included as one of the candidates in the forecast comparisons. The spectral density function of the general ARMA(*p*, *q*) model is simply

$$f_{arma}(w) = (\frac{x^2}{2p})\,|q(e^{-iw})|^2\,|f(e^{-iw})|^{-2}\,. \tag{4}$$

2.5 Maximum likelihood estimation

The parameters of an ARFIMA(*p*, *d*, *q*) model[4] can be estimated efficiently by maximizing the likelihood function.[5] Whittle (1951) derived the log-likelihood function in the frequency domain for a Gaussian process, here calculated from T observations as

$$K = -\frac{1}{2}\sum_{j=1}^{T-1} \ln(2pf(w_j)) + \hat{f}(w_j)/f(w_j) \tag{5}$$

where f is the theoretical spectral density of the process. The spectral density f at frequency $w_j = 2pj/T$ is given by equations (3) and (4), while $\hat{f}(w_j)$ is the value of the periodogram at frequency w_j. As the distribution of the realized logarithmic volatilities is approximately Gaussian, it is appropriate to estimate the model parameters by maximizing the above log-likelihood.[6]

We estimate four models for $y_t = \ln(s_t)$ from the in-sample data, that ends in 1993. Model 1 includes only the fractional differencing operator, model 2 adds an autoregressive component, i.e. ARFIMA(1, *d*, 0), while model 3 is the simplest that

---

[4] There are $p+q+3$ parameters: p autoregressive terms, q moving average terms, the order of integration d, the mean m and the variance of the residuals x^2. There are only $p+q+2$ parameters for the ARMA model, that has $d=0$.

[5] The Geweke and Porter-Hudak (1983) method and the scaling law method of ABDL (2001) provide alternative ways to estimate the order of integration d for the ARFIMA model. However, Smith, Taylor and Yadav (1996) show that the MLE of d outperforms the semi-parametric methods in terms of bias and mean squared error.

[6] The likelihood estimate is robust for non-Gaussian distributions, according to the detailed simulation study of Taqqu and Teverovsky (1997).

includes both autoregressive and moving average components, i.e. ARFIMA(1, *d*, 1). Model 4 is the ARMA(2, 1) model. The maximum likelihood estimates and their standard errors are shown in Table 2 for the five-minute series. The estimates of *d* equal 0.43, 0.44 and 0.48 for the pound, mark and yen respectively when model 1 is estimated. When autoregressive and moving-average components are included, the estimates of *d* decrease to 0.33 and 0.34 for the pound and the mark while that for the yen only changes a little, to 0.46. All the estimates of *d* define covariance stationary processes, although many of the 95% confidence intervals for *d* extend beyond 0.5.

For all three currencies the estimates for the ARMA(2, 1) model can be interpreted as the sum of persistent and transient AR(1) components. The component autoregressive parameters solve the quadratic equation $z^2 f(z^{-1}) = 0$. For example, the pound series has $\hat{f}_1 = -1.18$, $\hat{f}_2 = 0.20$ giving component AR(1) parameters equal to 0.964 and 0.205 that are similar to the estimates in Alizadeh, Brandt and Diebold (2002). The variance of the persistent AR(1) component equals 59% of the variance of the ARMA(2, 1) process for the pound series.

Likelihood ratio tests help to identify the most suitable model for log realized volatility during the in-sample period. From the test statistics given in Table 2 we conclude that the most appropriate fractionally integrated models are ARFIMA(1, *d*, 1) for the pound and the mark while the ARFIMA(0, *d*, 0) specification is best for the yen. The ARMA(2, 1) model has log-likelihood values that almost equal those of the ARFIMA(1, *d*, 1) model and consequently the short memory ARMA model merits evaluation in our forecasting competition.

The maximum likelihood estimates for the thirty-minute series are generally similar to those of the five-minute series except that the estimates of *d* are lower, with the reduction being approximately 0.07 for the ARFIMA(0, d, 0) model. The likelihood ratio tests yield results consistent with those for the five-minute series.

3 Forecasting realized volatility

Four methods for forecasting realized volatility are described in this section. Three of the methods use time series information. The first uses a GARCH model and the second and third utilize the ARFIMA(p, d, q) model with either short ($d = 0$) or long ($d > 0$) memory properties. The fourth method obtains forecasts from implied volatilities.

The GARCH(1,1) model is included as one of the candidates in the forecast comparisons because of its popularity among practitioners and academics. We use the simplest credible model from the GARCH family to obtain volatility forecasts from the information in daily returns. The parameters are estimated by the standard quasi-likelihood method. We employ a rolling estimation methodology to obtain time series of volatility forecasts. No intraday price information is used in the construction of these forecasts.

3.1 Forecasting using AR(FI)MA models

To forecast the logarithm of volatility one day ahead from an ARFIMA model, we represent equation 2 by the infinite-order autoregressive formulation:

$$\ln(s_t) - m = \sum_{k=1}^{\infty} p_k [\ln(s_{t-k}) - m] + e_t \quad (6)$$

where p_k is minus the coefficient of L^k in the power series expansion of $B(L) = (1-L)^d f(L) q(L)^{-1}$. The infinite-order polynomial $B(L)$ must be truncated and we do so at $k = 1200$.

The one-step-ahead forecast made at time t is the expectation of the next observation, conditional on the information I_t in the series up to that time, :

$$E(\ln(s_{t+1}) \mid I_t) = m + \sum_{k=1}^{1200} p_k [\ln(s_{t+1-k}) - m]. \tag{7}$$

The j-step ahead forecast of the volatility logarithm made at time t is as follows for $j \geq 2$,

$$E[\ln(s_{t+j}) \mid I_t] = m + \sum_{k=1}^{j-1} p_k E[\ln(s_{t+j-k}) - m \mid I_t] + \sum_{k=j}^{1200} p_k [\ln(s_{t+j-k}) - m]. \tag{8}$$

An estimate of the mean m is required to implement these equations. The spectral density of a process is not a function of m and hence the likelihood method of Section 2.5 does not provide an estimate. Providing $B(L)$ is truncated,

$$m = E[B(L)\ln(s_t)] / B(1), \tag{9}$$

with $B(1) = 1 - (p_1 + ... + p_k)$, and again we truncate after 1200 lags. The estimate of m follows from the average value of the filtered series.

To obtain volatility forecasts, we can not simply apply the exponential function to the logarithmic forecasts given by equations (7) and (8) as biased forecasts are then obtained. Instead we employ the method of Granger and Newbold (1976) to obtain the volatility forecast $g_{t,j}$:

$$g_{t,j} = \exp[f_{t,j} + \frac{1}{2} S^2(j)], \tag{10}$$

where

$$f_{t,j} = E[\ln(s_{t+j}) \mid I_t] \quad \text{and} \quad S^2(j) = \text{var}[\ln(s_{t+j}) \mid I_t].$$

When the forecast horizon is N days, the forecast of total volatility within the period is expressed as:

$$F_{t,N} = \left(\sum_{j=1}^{N} g_{t,j}^{2}\right)^{\frac{1}{2}}. \tag{11}$$

3.2 Forecasting using implied volatility

Option implied volatilities provide market information about the expected exchange rate return volatility for the period until the expiry date of the option. Unlike the realized volatilities, the implied volatilities are forward looking. However, implied volatility may be a biased representation of market expectations, for example if volatility risk is priced or transaction prices do not represent equilibrium market prices or the option pricing model is mis-specified. Despite these concerns, implied volatilities have often been found to be a better volatility forecast than those given by historical price models which use low-frequency returns as noted in Section 1.

The options data in this study are daily volatility quotes from the over-the-counter (OTC) market from July 1, 1988 to December 31, 1998.[7] The quotes are the averages of the bid and ask implied volatility at 12:00 in London, obtained from European at-the-money options.

Two series of implied volatilities are used in our study for each of the three exchange rates. The time remaining until the options expire is constant for each series and equals either one month or three months. Table 3 shows the summary statistics for the six series of volatility quotes. The standard deviation, range, skewness and kurtosis coefficients of the one-month implied volatilities are higher than those of the three-month implied volatilities, consistent with the studies of Xu and Taylor (1994) and Campa and Chang (1995).

Many previous studies find that equity implied volatilities are biased forecasts for future volatility. There are several possible causes. Day and Lewis (1992) and Christensen and Prabhala (1998) point out that their implied volatilities contain measurement errors whose magnitudes are unknown. Fleming, Ostdiek and Whaley (1995) and Blair, Poon and Taylor (2001) use a weighted index of American implied volatilities from eight near-the-money, near-to-expiry options called VIX to eliminate measurement errors and the smile effect. Even then, Blair *et al* find that VIX is a biased predictor of subsequent volatility. More recently, Bakshi and Kapadia (2003) show that the volatility risk premium is negative for index options and consequently at-the-money implied volatilities are upward-biased forecasts for future volatilities.

We apply an empirical design intended to minimize any bias incurred for the above reasons. We assume there is a linear relationship between the implied volatilities ($IV_{t,N}$) and the integrated realized volatilities ($s_{t,N}$) so that

$$s_{t,N} = k_{1,t,N} + k_{2,t,N} IV_{t,N} + e_{t,N}. \tag{12}$$

Here $s_{t,N}$ is defined as the annualized realized volatility within a *N*-day period starting from day t and $IV_{t,N}$ is the implied volatility for an option with N trading days to maturity traded on day *t*. When the forecast horizon does not match an option expiry date, it is necessary to redefine $IV_{t,N}$ as the implied volatility for an option whose

[7] The fact that implied volatility data starts half a year later than the in-sample period for modeling realized volatility does not affect our forecasting comparison which is only carried out for the out-of-the sample period starting from January 1994.

time until maturity is nearest to the forecast horizon. The parameters $k_{1,t,N}$ and $k_{2,t,N}$ are estimated through rolling regressions, always using five years of prior monthly observations. The forecast made on day t is $FIV_{t,N} = \hat{k}_{1,t,N} + \hat{k}_{2,t,N} IV_{t,N}$.

The median values of our estimates $k_{1,t,N}$ and $k_{2,t,N}$ are close to zero and one, respectively, when N is either one or five. However, for the one and three month periods the median intercept is positive and the median slope is below one; the ranges for the medians are from 2.6 to 5.8 for the intercept and from 0.50 to 0.73 for the slope, when the volatility measures are annualized percentage quantities.

3.3 Methodological choices

An important choice when comparing forecasts is the selection of forecast horizons. Despite the increasing popularity of long memory models in volatility forecasting, research has mainly focused on the short term with forecast horizons that are one month or less. For example, ABDL (2003) evaluated the forecasting abilities of the ARFIMA model for horizons of one trading day and ten trading days. An exception is Li (2002) who considers horizons up to six months. We evaluate forecasts for a broad range of horizons : one day, one week, one month and three months.

Another choice we need to make is the interval between the forecasts. Choosing a short interval increases the power of statistical inferences but may simultaneously create an overlapping data problem. To strike a balance, we set the intervals equal to the forecast horizons except for the three-month forecasts. The overlapping data problem then occurs only for the three-month horizon. Forecasts that overlap induce serial correlation among the forecast errors and extra care is required when tests are performed. We choose to make one-week forecasts on each Wednesday and one-month and three-month forecasts on the third Wednesday of each month.

The implied volatility quotes are annualized percentages while the realized volatilities and the forecasts from historical price models are all quantities for specific horizons less than one year. To make their scales compatible, we annualize the realized volatilities, ARFIMA forecasts, ARMA forecasts and GARCH forecasts. The formula used is as follows:

$$v^{*}_{t,N_1} = 100\sqrt{\left(\frac{N_2}{N_1} v_{t,N_1}{}^{2}\right)} \tag{13}$$

where v^{*}_{t,N_1} represents the percentage volatilities after annualization, N_1 is the number of trading days in a given forecast horizon starting on day t and N_2 is the number of trading days in the year in which day t falls. All the subsequent empirical results are for volatilities expressed as annualized percentages.

4 Comparisons of forecast performance

We now compare the accuracy of the various sets of forecasts. There are several measures that can be used to do this, that are reviewed by Diebold and Lopez (1996) and Poon and Granger (2003). We choose three important measures for our forecast comparisons. First, we use the mean squared error (MSE) as a criterion for comparison. Second, we carry out regression based efficiency tests and compare the coefficient of determination (R^2) from different models. Finally, we make use of bivariate regressions to decide if time series forecasts contain incremental information that is not captured by the implied volatilities.

4.1 Mean squared error

A loss function is often used to measure forecast accuracy. The loss function depends on pairs of forecasts and realized volatilities. In many applications, it is expressed directly as a function of the forecast errors $e_{j,t}$, where $e_{j,t}$ represents the forecast error of method j made at time t. The mean squared error (MSE) for a sample of size n is the quadratic loss function defined as $MSE = \frac{1}{n}\sum_{j=1}^{n} e_{j,t}{}^2$.

Table 4 shows the magnitudes of the MSEs for the four forecasting methods. The four MSEs can be ranked for each of 24 cases, given by combining the four forecast horizons, the three currencies and the two frequencies used to calculate the realized volatilities. The 24 sets of ranks yield mixed results, with implied volatilities (FIV) the most accurate for nine cases, ARFIMA best for eight cases, ARMA(2, 1) best for six and GARCH best for only one. There are six sets of ranks for each forecast horizon. The AR(FI)MA forecasts have six top ranks one-day ahead and either two or three top ranks for the other horizons. In contrast, the FIV forecasts have 0, 2, 4 and 3 top ranks as the horizon increases.

Two consistent results are observed from Table 4. First, the AR(FI)MA forecasts generally produce smaller MSEs than the GARCH forecasts for all forecast horizons with only a couple of exceptions for the yen. Second, the ARFIMA and ARMA models produce forecasts of similar accuracy. This shows that high frequency data enhances volatility forecasts more than the selection of a long memory specification.

We now focus on the results for the five-minute series and start with the results for the pound. The numbers in Table 4 show that the MSE for FIV is higher than for the ARFIMA and ARMA forecasts for the one-day and one-week horizons. As the horizon increases to one month and then three months, FIV produces forecast accuracy very close to that of the AR(FI)MA forecasts. The MSEs for the GARCH forecasts are highest for all forecast horizons. The test of Diebold and Mariano (1995) is used to decide if two forecasting methods provide the same accuracy.[8] Table 5 shows the p-values of the test when FIV is compared with different historical forecasts. These values show that the ARFIMA and ARMA forecasts are significantly more accurate than FIV for the one-day and one-week horizons at the 2% significance level. FIV has accuracy superior to GARCH forecasts for all forecast horizons at the same significance level.

Next consider the results for the mark. When the forecast horizon is one day, again the AR(FI)MA forecasts have smaller MSE than FIV and the GARCH forecasts give the highest MSE. The Diebold-Mariano test shows that the ARFIMA and ARMA forecasts are significantly more accurate than FIV at the 6% significance level. The relative performance of FIV improves as the forecast horizon become longer. FIV gives the least MSE for the one-week horizon while it produces MSEs of similar magnitudes to the ARFIMA and ARMA forecasts for one-month and three-month horizons. The hypothesis test can not reject the hypothesis of equal accuracy between the AR(FI)MA forecasts and FIV at the 25% significance level when the forecast horizons are one week, one month and three months.

We now turn our attention to the results for the yen. The magnitudes of the MSE are much larger than those of the pound and the mark because the yen exchange

[8] The test statistic is calculated from the loss-differential time series defined as the difference between squared forecast errors from two different forecast methods. It equals the sample mean of the loss-differential series divided by an estimate of its standard error that takes account (when necessary) of serial correlation in the series. The null hypothesis is a zero population mean for the loss-differential series.

rate was much more volatile during the period for which the forecasts are evaluated. The MSE of FIV is the highest when the forecast horizon is one day and even exceeds that of the GARCH forecasts. The test results again show that the AR(FI)MA forecasts are significantly more accurate than FIV at the 10% level. All the methods have comparable accuracy when the forecast horizon is one week. The MSE of FIV is the least for the one-month and thee-month horizons. The tests show that FIV is superior to the historical forecasts at the 6% significance level for the one-month horizon and at the 13% level for the three-month horizon.

When the thirty-minute series are used instead, the magnitudes of the MSEs are generally higher than those of the five-minute series. This is expected because volatility is measured less accurately when the returns are less frequent. Despite this, similar rankings are obtained as can be seen from Table 4.

4.2 Regression analysis

In addition to using MSE as a criterion, we report R^2 for regressions of the form

$$s_{t,N} = y_{j,0} + y_{j,1} F_{t,N} + e_t \tag{14}$$

where $s_{t,N}$ is the *N*-day integrated realized volatility and $F_{t,N}$ is the forecast made at time t. The coefficient of determination R^2 from a regression represents the information content of a particular forecasting method. Table 6 shows the values of R^2 for all three exchange rates. These values are usually similar for the ARFIMA and ARMA forecasts, while the GARCH(1,1) forecasts generally produce the lowest values with a few exceptions for the yen at short horizons.

Ranking the four values of R^2 for the four forecasting methods again gives a top rank for each of 24 cases, by combining the four forecast horizons, the three currencies and the two frequencies used to calculate the realized volatilities. The implied volatilities (FIV) have the highest R^2 for eleven cases, one of the two AR(FI)MA methods is best for ten cases, there is a three-way tie for one case and GARCH is best for the remaining two cases. These counts are similar to those for the MSE comparisons, but the best method identified by the MSE and R^2 criteria is different for ten cases. The top ranks for the AR(FI)MA forecasts are fairly evenly spread across the horizons, while the FIV forecasts have 2, 2, 4 and 3 top ranks as the horizon increases from one day to three months.

We first comment on the results for the pound. The values of R^2 for the FIV forecasts are lower than those for the AR(FI)MA forecasts when the horizon is one day or one week, but they are higher for the one-month horizon. The relative performance of these methods for the three-month horizon depends on the frequency used to define realized volatility.

For the mark, the values of R^2 are always similar for the FIV, ARFIMA and ARMA forecasts regardless of the horizon or the frequency that defines realized volatility. The GARCH forecasts always have the lowest values.

The results for the yen can not be summarized as simply as those for the mark. When the five-minute series is used, AR(FI)MA is superior to FIV for the one-day and one-week horizons, is similar for the one-month horizon and is inferior for the three-month horizon. Unlike the results for the pound and the mark, the GARCH forecasts have the highest R^2 for one-week forecasts and the second highest for the one-day horizon, although GARCH again has the lowest values for the longer horizons.

Switching to the thirty-minute series provides values of R^2 that are similar for FIV, ARFIMA and ARMA forecasts for horizons up to one month while FIV clearly has the highest value for the three-month horizon. The GARCH results are again the best for the one-week horizon.

4.3 Test for incremental information

In addition to finding out which forecast method is best, we are also interested in assessing the incremental information in exchange rates and the prices of their options. The test for incremental information is performed by regressing the realized volatilities on FIV and one of the historical forecasts (HIST), as follows :

$$s_{t,N} = a + b_{FIV}(\mathrm{FIV}_{t,N}) + b_{HIST}(\mathrm{HIST}_{t,N}) + e_t . \tag{15}$$

If b_{HIST} is equal to zero, there is no incremental predictive information from that particular historical forecast. The null hypothesis $b_{HIST} = 0$ can be tested using a standard regression test.[9]

The results for the five-minute series and the thirty-minute series are very similar. Therefore, we focus our discussion on the encompassing regression results for the five-minute series. These results are shown in Tables 7, 8 and 9, respectively for the pound, the mark and the yen.

When the forecast horizon is one day, the coefficient of the AR(FI)MA forecast always exceeds that of the option forecast, FIV, for all three exchange rates i.e. $b_{HIST} > b_{FIV}$. The null hypothesis $b_{HIST} = 0$ is rejected at low significance levels,

[9] For the three-month forecasts, the standard errors of the coefficients are computed using the Newey-West method to take account of the autocorrelation in the residuals from the regression.

as the corresponding t-ratios all exceed three. Thus there is significant incremental information in the ARFIMA, ARMA and GARCH forecasts beyond that provided by FIV. The other null hypothesis of interest is $b_{FIV} = 0$, which is rejected at low levels for the pound and the mark but not for the yen.

At the next forecast horizon of one week, $b_{HIST} > b_{FIV}$ for the AR(FI)MA forecasts of pound and yen volatility but $b_{HIST} < b_{FIV}$ for the mark. The smallest t-ratio for b_{HIST} equals 2.18, so all the coefficients on the historical forecasts are then significantly different from zero at the 5% level for all exchange rates. As the smallest t-ratio for b_{FIV} is 2.64, there is also significant evidence for incremental information in the implied volatilities.

Different conclusions are obtained for the longer horizons of one month and three months, that reflect the small number of forecasts that can then be made. Now $b_{HIST} < b_{FIV}$ for five of the six regressions for the pound, for two cases for the mark and all six cases for the yen. There is only one rejection of $b_{HIST} = 0$ at the 5% significance level, that occurs for the one-month ARMA(2,1) forecasts of pound volatility. There are more rejections of $b_{FIV} = 0$ at the 5% level : four for the pound, one for the mark and three for the yen; three cases involve AR(FI)MA forecasts and five GARCH forecasts.

The information content results for implied volatilities and ARFIMA forecasts are consistent with those of Martens and Zein (2002) for the yen. They found information in long memory forecasts incremental to implied volatilities for the one-day horizon but not for a 20-day horizon, based on a non-overlapping methodology. Li (2002) has different conclusions and claims that overlapping ARFIMA forecasts have incremental information over implied volatilities for the one-month and three-month forecast horizons, for the same exchange rates that are investigated in this paper. However, as pointed out by Fleming (1998), Christensen and Prabhala (1998) and Martens and Zein (2002), evidence for incremental information in historical forecasts can be a statistical artifact caused by using overlapping forecasts for the encompassing test.

5 Conclusions

The main contribution of this paper is to compare the volatility forecasting ability of *four* methods for the first time: GARCH forecasts obtained from daily returns, short memory ARMA and long memory ARFIMA forecasts obtained from high-frequency returns, and implied volatilities obtained from option prices.

Only recently have forecasts obtained from long memory models and/or high-frequency data begun to be assessed. ABDL (2003) find that long memory models outperform traditional historical price models which use low frequency returns, including the IGARCH and GARCH models, for exchange rate volatility forecasting. Li (2002) finds long memory and implied predictors have comparable accuracy for one-month ahead forecasts of exchange rate volatility, while time series forecasts perform best for the six-month horizon. Martens and Zein (2002) find incremental information in long memory forecasts for yen volatility for the one-day horizon but none for the twenty-day horizon, relative to implieds. These studies do not show if the incremental information of the historical forecasts originates from the use of a long memory model or from high frequency returns. We have shown in this paper that the forecasting performances of short memory forecasts and long memory forecasts, relative to implied volatilities, are very similar. This shows that the forecasting accuracy of historical forecasts is enhanced by using high-frequency returns rather than by the selection of a long memory model.

We use high frequency returns to construct realized volatility estimates for the exchange rates of the dollar against the pound, the mark and the yen, which define the target in our forecast comparison. Two volatility series are constructed for each exchange rate, one using five-minute returns and the other using thirty-minute returns. We find that the unconditional distribution of the logarithm of realized volatilities is approximately Gaussian for both volatility series and their dynamics are best represented either by a long memory process or a particular short memory process.

Our forecast comparisons focus on the out-of-the-sample period from January 1994 to December 1998. We compare the forecasting ability of implied volatilities, the ARFIMA model, the ARMA(2, 1) model and the GARCH(1, 1) model for forecast horizons ranging from one day to three months. Using a mean-squared error (MSE) criterion, we find that the ARFIMA and ARMA forecasts generally perform better than implied volatilities for short forecast horizons while implied volatilities produce more accurate forecasts for longer forecast horizons. The GARCH forecasts are the least accurate for most of the evaluations, although incremental information is found to exist over implied volatilities for the shorter forecast horizons. Consistent results are obtained for both the five-minute and thirty-minute volatility series.

The R^2 statistic from a regression of realized volatility on its predicted value is an alternative criterion to the MSE, that is an indicator of the richness of the information content provided by the forecasts. The ARFIMA and ARMA forecasts produce values of R^2 higher than those for implied volatilities for two currencies (the pound and the yen) when the forecast horizon is either one day or one week. For longer horizons, the implied volatilities are more informative forecasts. For the third currency (the mark), the values of R^2 are similar for implieds and AR(FI)MA forecasts, for all forecast horizons.

In addition to comparing the forecasts individually, we perform tests to see if ARFIMA, ARMA and GARCH forecasts have incremental information that is not found in implied volatilities. We find that all of these forecasts have incremental information over implied volatilities for the short forecast horizons in this study, which are one day and one week. In contrast, implied volatilities are found to incorporate most of the relevant information for all exchange rates when the horizon is either one month or three months.

Acknowledgements

The authors thank the two referees for their detailed and constructive advice. A previous version of this paper, entitled "Forecasting Sterling/Dollar volatility …", was presented in 2000 at the SIRIF volatility conference in Glasgow, the European Finance Association meeting in Berlin, the Australasian Banking and Finance conference in Sydney and the Global Finance Association meeting in Beijing. We thank participants at these meetings for their comments. Particular thanks are due to Martin Martens and Theo Nijman.

References

Alizadeh, S., M.W. Brandt and F.X. Diebold, 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* 57, 1047-1091.

Andersen, T. G. and T. Bollerslev, 1998, Answering the skeptics: yes standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885-905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, The distribution of stock return volatility, *Journal of Financial Economics* 61, 43-76.

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys, 2001, The distribution of exchange rate volatility, *Journal of the American Statistical Association* 96, 42-55.

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579-625.

Areal, N. and S. J. Taylor, 2002, The realized volatility of FTSE-100 futures prices, *Journal of Futures Markets* 22, 627-648.

Baillie, R.T., 1996, Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5-59.

Bakshi, G. and N Kapadia, 2003, Delta-hedged gains and the negative market volatility risk premium, *Review of Financial Studies* 16, 527-566.

Barndorff-Nielsen, O.E. and N. Shephard, 2001, Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics, *Journal of the Royal Statistical Society B* 63, 167-241.

Barndorff-Nielsen, O.E. and N. Shephard, 2002a, Estimating quadratic variation using realised variance, *Journal of Applied Econometrics* 17, 457-477.

Barndorff-Nielsen, O.E. and N. Shephard, 2002b, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society B* 64, 253-280.

Blair, B., S-H. Poon and S. J. Taylor, 2001, Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns, *Journal of Econometrics* 105, 5-26.

Bollerslev, T. and I. Domowitz, 1993, Trading patterns and prices in the interbank foreign exchange, *Journal of Finance* 48, 1421-1443.

Campa, J. M. and P. H. Chang, 1995, Testing the expectations hypothesis on the term structure of volatilities in foreign exchange options, *Journal of Finance* 50, 529-547.

Christensen, B. J. and N. R. Prabhala, 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50, 125-150.

Day, T. E. and C. M. Lewis, 1992, Stock market volatility and the informational content of stock index options, *Journal of Econometrics* 52, 267-287.

Diebold, F. X. and J. Lopez, 1996, Forecast Evaluation and Combination, G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics*, Amsterdam: North-Holland, 241-268.

Diebold F. X. and R. S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

Figlewski, S., 1997, Forecasting volatility, *Financial Markets, Institutions and Instruments* 6, 1-88.

Fleming, J., 1998, The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance* 5, 317-345.

Fleming, J., B. Ostdiek and R. E. Whaley, 1995, Predicting stock market volatility: A new measure, *Journal of Futures Market* 15, 265-302.

Gallant, R., C-T Hsu and G. Tauchen, 1999, Using daily range data to calibrate volatility diffusions and extract the forward integrated variance, *The Review of Economics and Statistics* 81, 617-631.

Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221-238.

Granger, C. W. J. and P. Newbold, 1976, *Forecasting Economic Time Series*, Academic Press. New York.

Granger, C. W. J. and R. Joyeux, 1980, An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-39.

Hosking, J. R. M., 1981, Fractional differencing, *Biometrika* 68, 165-176.

Jorion, P., 1995, Predicting volatility in the foreign exchange market, *Journal of Finance* 50, 507-528.

Li, K., 2002, Long-memory versus option-implied volatility predictions, *Journal of Derivatives* 9, No. 3, 9-25.

Martens, M. and J. Zein, 2002, Predicting financial volatility : high-frequency time-series forecasts vis-à-vis implied volatility, working paper, Erasmus University, Rotterdam.

Poon, S-H. and C. W. J. Granger, 2003, Forecasting financial market volatility : a review, *Journal of Economic Literature* 41, 478-539.

Smith, J., N. Taylor and S. Yadav, 1997, Comparing the bias and misspecification in ARFIMA models, *Journal of Times Series Analysis* 18, 507-527.

Taqqu, M. S. and V. Teverovsky, 1997, Robustness of Whittle type estimators for time series with long range dependence, *Stochastic Models* 13, 723-757.

Taylor, S. J. and X. Xu, 1997, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance* 4, 317-340.

Whittle, P., 1951, *Hypothesis Testing in Time Series Analysis*, Almquist and Wiksells, Uppsala.

Xu, X. and S. J. Taylor, 1994, The term structure of volatility implied by foreign exchange options, *Journal of Financial and Quantitative Analysis* 29, 57-74.

Xu, X. and S. J. Taylor, 1995, Conditional volatility and the informational efficiency of the PHLX currency options markets, *Journal of Banking and Finance* 19, 803-821.

Table 1. The distribution of realized volatility and its logarithm for three exchange rates

	USD/GBP		USD/DEM		USD/JPY	
	Realized Volatility	*Logarithm of RV*	*Realized Volatility*	*Logarithm of RV*	*Realized Volatility*	*Logarithm of RV*
Mean	0.67	-0.44	0.70	-0.42	0.66	-0.48
Standard Deviation	0.23	0.31	0.25	0.34	0.24	0.34
Skewness	1.75	0.34	1.65	0.04	1.44	0.16
Kurtosis	9.82	3.49	8.85	3.67	7.06	3.17
Maximum	2.72	1.00	2.64	0.97	2.13	0.76
Minimum	0.24	-1.43	0.21	-1.57	0.20	-1.61
Autocorrelation						
lag(1)	0.63	0.65	0.62	0.67	0.64	0.67
lag(2)	0.54	0.57	0.52	0.59	0.53	0.56
lag(3)	0.50	0.53	0.48	0.54	0.47	0.50
lag(4)	0.47	0.49	0.45	0.51	0.44	0.47
lag(5)	0.48	0.49	0.46	0.52	0.43	0.46
Ljung Box Test						
Q(20)	5445.73	5864.11	5126.36	6580.41	3996.49	4794.65

The daily realized volatilities are for the in-sample period from July 1, 1987 to December 31, 1993. They are constructed using five-minute returns.

USD/GBP refers to the exchange rate of the dollar against the British pound in all the Tables. Likewise, USD/DEM and USD/JPY refer to rates for the Deutsche mark and the Japanese yen.

Table 2. Maximum likelihood estimation of ARFIMA models for log realized volatility

A.

ARFIMA	USD/GBP *Model 1 (0,d,0)*	*Model 2 (1,d,0)*	*Model 3 (1,d,1)*	*Model 4 (2,0,1)*	USD/DEM *Model 1 (0,d,0)*	*Model 2 (1,d,0)*	*Model 3 (1,d,1)*	*Model 4 (2,0,1)*	USD/JPY *Model 1 (0,d,0)*	*Model 2 (1,d,0)*	*Model 3 (1,d,1)*	*Model 4 (2,0,1)*
d	0.4339	0.4768	0.3285		0.4390	0.4752	0.3392		0.4789	0.4638	0.4639	
	(0.0275)	(0.0233)	(0.0711)		(0.0252)	(0.0420)	(0.0610)		(0.0326)	(0.0524)	(0.0647)	
AR(1)		0.0739	-0.9329	-1.1693		0.0629	-0.9413	-1.1975		-0.0237	0.0028	-1.2561
		(0.0273)	(0.0543)	(0.0639)		(0.0514)	(0.0441)	(0.0515)		(0.05849)	(0.0351)	(0.0756)
AR(2)				0.1976				0.2201				0.2900
				(0.0537)				(0.0469)				(0.0641)
MA(1)			-0.8703	-0.7685			-0.8849	-0.7879			0.0266	-0.7646
			(0.0936)	(0.0516)			(0.0703)	(0.0395)			(0.0514)	(0.0648)
log likelihood value	1635.65	1637.74	1640.18	1639.86	1523.81	1525.21	1528.52	1528.55	1511.83	1512.00	1512.00	1511.52

B.

USD/GBP *Model Comparison*	*Likelihood Ratio Test Statistic*		USD/DEM *Model Comparison*	*Likelihood Ratio Test Statistic*		USD/JPY *Model Comparison*	*Likelihood Ratio Test Statistic*	
ARFIMA			ARFIMA			ARFIMA		
(0,d,0) vs (1,d,0)	4.1867	(0.0407)	(0,d,0) vs (1,d,0)	2.7996	(0.0942)	(0,d,0) vs (1,d,0)	0.3347	(0.5629)
(1,d,0) vs (1,d,1)	4.8849	(0.0271)	(1,d,0) vs (1,d,1)	6.6199	(0.0101)	(0,d,0) vs (1,d,1)	0.3429	(0.8424)

The estimates are for the five-minute realized volatility series, from July 1, 1988 to December 31, 1993. Panel A shows the estimates of the parameters of four different models obtained by MLE. The values in parentheses are the standard errors of the estimates. Panel B shows the likelihood ratio test statistics when comparing two models. The values in parentheses are the levels of significance of the test statistics.

Table 3. Summary statistics for the implied volatility quotes

	USD/GBP		USD/DEM		USD/JPY	
	One-month Implied Volatilities	*Three-month Implied Volatilities*	*One-month Implied Volatilities*	*Three-month Implied Volatilities*	*One-month Implied Volatilities*	*Three-month Implied Volatilities*
Mean	10.40	10.61	11.21	11.31	11.08	11.16
Standard Deviation	2.81	2.25	2.30	1.75	2.98	2.44
Skewness	0.48	0.22	0.71	0.15	1.34	1.27
Kurtosis	3.66	2.86	4.53	3.39	6.76	5.30
Maximum	24.00	19.00	24.00	19.00	36.50	28.00
Minimum	4.15	4.60	6.25	6.60	6.25	6.75
Range	19.85	14.40	17.75	12.40	30.25	21.25

The summary statistics are for the average of daily bid and ask volatility quotes, for options with one and three months to expiration, for the USD/GBP, USD/DEM and USD/JPY exchange rates for the period from July 1, 1988 to December 31, 1998.

Table 4. Values of the mean squared error for the volatility forecasts

A. USD/GBP

	5-min series				30-min series			
Forecast/Horizon	*1 day*	*1 week*	*1 month*	*3 month*	*1 day*	*1 week*	*1 month*	*3 month*
FIV	4.16	3.14	1.95	2.14	7.03	4.34	2.53	2.84
ARFIMA(1,d,1)	3.81	2.48	1.96	2.15	6.32	3.90	2.89	2.79
ARMA(2,1)	3.80	2.48	1.91	2.21	6.30	3.88	2.84	2.81
GARCH(1,1)	4.75	4.19	3.57	3.92	9.29	7.77	4.17	9.29

B. USD/DEM

	5-min series				30-min series			
Forecast/Horizon	*1 day*	*1 week*	*1 month*	*3 month*	*1 day*	*1 week*	*1 month*	*3 month*
FIV	7.09	4.89	4.34	4.45	10.79	7.05	5.02	5.26
ARFIMA(1,d,1)	6.55	5.17	4.42	4.23	10.32	7.24	4.95	4.61
ARMA(2,1)	6.55	5.14	4.45	4.21	10.38	7.31	5.14	4.87
GARCH(1,1)	7.47	6.25	6.23	5.10	12.78	9.81	9.38	7.26

C. USD/JPY

	5-min series				30-min series			
Forecast/Horizon	*1 day*	*1 week*	*1 month*	*3 month*	*1 day*	*1 week*	*1 month*	*3 month*
FIV	19.37	12.72	9.01	8.39	24.00	15.14	10.48	9.19
ARFIMA(0,d,0)	15.11	12.96	11.09	11.39	20.00	16.34	12.64	13.67
ARMA(2,1)	15.24	13.14	11.43	11.50	20.44	16.55	12.81	13.91
GARCH(1,1)	17.51	12.91	16.47	12.36	21.51	13.57	18.42	15.04

The table shows the magnitude of the Mean Square Error (MSE) for different forecast horizons. The MSE is defined as the mean square of the difference of the annualized realized volatility and the forecast volatility. The results cover the out-of-sample period from January 1, 1994 to December 31, 1998.

Table 5. The p-values of the sample means of the loss differentials

A. USD/GBP

Forecast Comparison	5-min series 1 day	1 week	1 mth	3 mths	30-min series 1 day	1 week	1 mth	3 mths
FIV vs ARFIMA(1,d,1)	0.02*	0.01*	0.49	0.50	0.00*	0.06*	0.04	0.45*
FIV vs ARMA(2,1)	0.01*	0.01*	0.42*	0.38	0.00*	0.07*	0.10	0.48*
FIV vs GARCH (1,1)	0.00	0.01	0.00	0.02	0.00	0.00	0.01	0.00

B. USD/DEM

Forecast Comparison	5-min series 1 day	1 week	1 mth	3 mths	30-min series 1 day	1 week	1 mth	3 mths
FIV vs ARFIMA(1,d,1)	0.06*	0.26	0.35	0.40*	0.14*	0.36	0.46*	0.35
FIV vs ARMA(2,1)	0.06*	0.29	0.34	0.41*	0.19*	0.31	0.44	0.42
FIV vs GARCH (1,1)	0.13	0.01	0.03	0.03	0.00	0.00	0.00	0.02

C. USD/JPY

Forecast Comparison	5-min series 1 day	1 week	1 mth	3 mths	30-min series 1 day	1 week	1 mth	3 mths
FIV vs ARFIMA(0,d,0)	0.08*	0.37	0.05	0.09	0.10*	0.08	0.01	0.04
FIV vs ARMA(2,1)	0.09*	0.29	0.06	0.13	0.13*	0.07	0.02	0.05
FIV vs GARCH (1,1)	0.27*	0.47	0.02	0.01	0.11*	0.26*	0.01	0.02

The table shows the p-values of the Diebold-Mariano test, which tests for equal forecast accuracy between two forecasting methods. The loss differentials are defined as the differences of squared forecast errors from two different forecasting methods. The mean of the loss differentials is equal to the difference between the mean squared errors of the two forecast methods, which are shown in Table 4. Asterisks indicate a positive mean difference for the loss differentials, i.e. FIVs produces a higher MSE than the historical forecasts.

Table 6. Values of the coefficient of determination for the regressions of realized volatility on volatility forecasts

A. USD/GBP

	5-min series				*30-min series*			
Forecast/Horizon	*1 day*	*1 week*	*1 month*	*3 month*	*1 day*	*1 week*	*1 month*	*3 month*
FIV	0.39	0.38	0.40	0.19	0.26	0.30	0.30	0.11
ARFIMA(1,d,1)	0.41	0.43	0.37	0.17	0.30	0.31	0.20	0.06
ARMA(2,1)	0.41	0.43	0.37	0.21	0.30	0.31	0.21	0.09
GARCH(1,1)	0.34	0.24	0.16	0.01	0.26	0.21	0.10	0.01

B. USD/DEM

	5-min series				*30-min series*			
Forecast/Horizon	*1 day*	*1 week*	*1 month*	*3 month*	*1 day*	*1 week*	*1 month*	*3 month*
FIV	0.45	0.44	0.34	0.17	0.35	0.35	0.28	0.13
ARFIMA(1,d,1)	0.45	0.40	0.35	0.17	0.34	0.34	0.31	0.15
ARMA(2,1)	0.45	0.41	0.36	0.19	0.34	0.34	0.31	0.16
GARCH(1,1)	0.40	0.33	0.21	0.12	0.31	0.27	0.18	0.09

C. USD/JPY

	5-min series				*30-min series*			
Forecast/Horizon	*1 day*	*1 week*	*1 month*	*3 month*	*1 day*	*1 week*	*1 month*	*3 month*
FIV	0.43	0.42	0.46	0.40	0.45	0.44	0.43	0.44
ARFIMA(0,d,0)	0.51	0.46	0.45	0.29	0.44	0.43	0.42	0.32
ARMA(2,1)	0.51	0.46	0.44	0.28	0.43	0.43	0.42	0.30
GARCH(1,1)	0.47	0.52	0.37	0.11	0.39	0.50	0.35	0.10

The table shows the value of the coefficient of determination, R^2 , from the regression of the annualized realized volatility on the volatility forecasts.

Table 7. The results of the test for incremental information for USD/GBP

Forecast Horizon=1 day

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	0.39	0.33	0.62	0.43
	(1.12)	(4.15)	(7.72)	
FIV vs ARMA(2,1)	0.32	0.33	0.63	0.43
	(0.92)	(4.03)	(7.18)	
FIV vs GARCH (1,1)	-1.06	0.55	0.53	0.41
	(-1.44)	(11.18)	(5.24)	

Forecast Horizon=1 week

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	0.65	0.37	0.58	0.45
	(1.12)	(3.41)	(4.81)	
FIV vs ARMA(2,1)	0.56	0.36	0.59	0.45
	(0.95)	(3.21)	(4.69)	
FIV vs GARCH (1,1)	-0.25	0.67	0.37	0.41
	(-0.22)	(8.14)	(2.65)	

Forecast Horizon=1 month

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	1.04	0.58	0.32	0.41
	(0.82)	(3.06)	(1.55)	
FIV vs ARMA(2,1)	0.62	0.53	0.41	0.42
	(0.47)	(3.20)	(2.08)	
FIV vs GARCH (1,1)	1.66	0.85	-0.01	0.40
	(1.37)	(4.73)	(-0.03)	

Forecast Horizon=3 month

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	3.54	0.43	0.17	0.20
	(1.41)	(1.26)	(0.35)	
FIV vs ARMA(2,1)	0.16	0.28	0.65	0.23
	(0.05)	(1.11)	(1.40)	
FIV vs GARCH (1,1)	6.29	0.65	-0.29	0.22
	(2.33)	(3.89)	(-1.04)	

The parameters are estimated from the regression $\sigma_{t,N} = \alpha + \beta_{FIV}(\text{FIV}) + \beta_{hist}(\text{historical forecast}) + \varepsilon_t$. The realized volatility $\sigma_{t,N}$ is constructed from the five-minute returns. The numbers in parentheses are t-statistics. When the forecast horizon is three months, the standard errors are estimated using the Newey-West method.

Table 8. The results of the test for incremental information for USD/DEM

Forecast Horizon=1 day

	a	*ß_{FIV}*	*ß_{Hist}*	R^2
FIV vs ARFIMA(1,d,1)	0.61	0.38	0.56	0.47
	(1.46)	(4.49)	(5.52)	
FIV vs ARMA(2,1)	0.55	0.38	0.57	0.47
	(1.29)	(4.32)	(5.46)	
FIV vs GARCH (1,1)	0.46	0.55	0.37	0.45
	(0.69)	(7.97)	(3.30)	

Forecast Horizon=1 week

	a	*ß_{FIV}*	*ß_{Hist}*	R^2
FIV vs ARFIMA(1,d,1)	0.81	0.57	0.37	0.46
	(1.01)	(3.65)	(2.18)	
FIV vs ARMA(2,1)	0.68	0.55	0.40	0.46
	(0.85)	(3.79)	(2.48)	
FIV vs GARCH (1,1)	0.18	0.67	0.30	0.46
	(0.18)	(6.19)	(2.31)	

Forecast Horizon=1 month

	a	*ß_{FIV}*	*ß_{Hist}*	R^2
FIV vs ARFIMA(1,d,1)	0.63	0.37	0.60	0.36
	(0.34)	(1.20)	(1.40)	
FIV vs ARMA(2,1)	0.00	0.38	0.66	0.38
	(0.85)	(1.59)	(1.70)	
FIV vs GARCH (1,1)	1.13	0.83	0.05	0.34
	(0.18)	(2.72)	(0.15)	

Forecast Horizon=3 month

	a	*ß_{FIV}*	*ß_{Hist}*	R^2
FIV vs ARFIMA(1,d,1)	2.27	0.15	0.65	0.17
	(0.34)	(0.28)	(1.40)	
FIV vs ARMA(2,1)	-0.51	0.23	0.86	0.20
	(-0.15)	(0.70)	(1.52)	
FIV vs GARCH (1,1)	3.28	0.47	0.17	0.14
	(1.05)	(1.10)	(0.29)	

The parameters are estimated from the regression $\sigma_{t,N} = \alpha + \beta_{FIV}(\text{FIV}) + \beta_{hist}(\text{historical forecast}) + \varepsilon_t$. The realized volatility $\sigma_{t,N}$ is constructed from the five-minute returns. The numbers in parentheses are the t-statistics. When the forecast horizon is three months, the standard errors are estimated using the Newey-West method.

Table 9. The results of the test for incremental information for USD/JPY

Forecast Horizon=1 day

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	-0.30	0.15	0.91	0.51
	(-0.41)	(1.19)	(6.31)	
FIV vs ARMA(2,1)	-0.68	0.16	0.92	0.51
	(-0.91)	(1.40)	(6.45)	
FIV vs GARCH (1,1)	1.18	0.31	0.65	0.49
	(1.23)	(2.11)	(3.45)	

Forecast Horizon=1 week

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	-0.45	0.48	0.59	0.49
	(-0.29)	(3.46)	(2.77)	
FIV vs ARMA(2,1)	-1.13	0.50	0.63	0.49
	(-0.71)	(4.31)	(3.50)	
FIV vs GARCH (1,1)	-0.16	0.40	0.68	0.55
	(-0.12)	(2.64)	(2.71)	

Forecast Horizon=1 month

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	0.49	0.67	0.35	0.47
	(0.23)	(1.71)	(0.83)	
FIV vs ARMA(2,1)	-0.98	0.61	0.54	0.48
	(-0.39)	(1.98)	(1.36)	
FIV vs GARCH (1,1)	1.47	0.77	0.15	0.48
	(0.79)	(3.49)	(0.70)	

Forecast Horizon=3 month

	a	$ß_{FIV}$	$ß_{Hist}$	R^2
FIV vs ARFIMA(1,d,1)	-3.73	1.59	-0.26	0.40
	(-0.79)	(1.88)	(-0.39)	
FIV vs ARMA(2,1)	-6.19	1.19	0.37	0.40
	(-1.24)	(1.91)	(0.50)	
FIV vs GARCH (1,1)	-3.91	1.35	0.01	0.40
	(-0.69)	(3.36)	(0.04)	

The parameters are estimated from the regression $\sigma_{t,N} = \alpha + \beta_{FIV}(\text{FIV}) + \beta_{hist}(\text{historical forecast}) + \varepsilon_t$. The realized volatility $\sigma_{t,N}$ is constructed from the five-minute returns. The numbers in parentheses are the t-statistics. When the forecast horizon is three months, the standard errors are estimated using the Newey-West method.