Forecasting the volatility of the Australian Dollar using high frequency data: Does estimator accuracy improve forecast evaluation?

George Bailey & James M. Steeley

Abstract

We compare forecasts of the volatility of the Australian Dollar / US Dollar exchange rate to alternative measures of ex-post volatility. We develop and apply a simple test for the improvement in the ability of loss functions to distinguish between forecasts when the quality of a volatility estimator is increased. We find that both realized variance and the daily high-low range provide a significant improvement in loss function convergence relative to squared returns. We find that a model of stochastic volatility provides the best forecasts, relative to a set of GARCH models, which includes a GARCH(1,1) that is second best.

Keywords: Volatility forecasting, exchange rate, Australian Dollar, stochastic volatility, realized variance, high-low range.

July 2017

Both authors: School of Management, Keele University, Keele, ST5 5BG, UK. Email: j.steeley@keele.ac.uk.

Forecasting the volatility of the Australian Dollar using high frequency data: Does estimator accuracy improve forecast evaluation?

Abstract

We compare forecasts of the volatility of the Australian Dollar / US Dollar exchange rate to alternative measures of ex-post volatility. We develop and apply a simple test for the improvement in the ability of loss functions to distinguish between forecasts when the quality of a volatility estimator is increased. We find that both realized variance and the daily high-low range provide a significant improvement in loss function convergence relative to squared returns. We find that a model of stochastic volatility provides the best forecasts, relative to a set of GARCH models, which includes a GARCH(1,1) that is second best.

Keywords: Volatility forecasting, exchange rate, Australian Dollar, stochastic volatility, realized variance, high-low range.

July 2017

1. Introduction

A key purpose of volatility models is to produce forecasts of volatility, which are useful for the pricing of financial assets that depend heavily on the evolution of volatility, such as derivatives, and also for input into economic policy analysis and forecasting.[1] Because 'true' volatility is unobservable, volatility estimators are required when one wishes to undertake a comparison of forecasted values. However, the early studies of volatility forecast evaluation,[2] found that the forecasts performed poorly when compared to the standard volatility estimator, squared daily returns. Andersen and Bollerslev (1998) demonstrated that more sophisticated estimators would increase the explanatory power offered by the volatility models, and recommended the use of volatility measured from high-frequency (intra-day) returns, which has become known as realized variance.

A key consideration in the evaluation of forecasts, more generally, is the selection of the loss function. Hansen and Lunde (2005) examined 330 ARCH related models to an exchange rate and a stock return series and found that different loss functions selected different models as providing the best volatility forecasts.[3] Combining the two issues, Hansen and Lunde (2006) and Patton (2011) explore the theoretical and practical interactions between the selected loss function and the quality of the volatility estimator. Both papers identify those loss functions that are less sensitive to the noise in volatility proxies and show that more sophisticated measures of volatility are better able to distinguish between forecasts.

In this paper we also focus on the interaction between the choice of volatility estimator and the selection of loss functions in the evaluation of volatility forecasts, but seek to answer the following specific question. Does the use of more sophisticated volatility proxies generate greater convergence amongst the forecast rankings across different loss functions?

We contribute and innovate in number of important ways. Both Hansen and Lunde (2006) and Patton (2011) examine the volatility of IBM stock returns, over similar time frames, ending in 2003. We examine an exchange rate, specifically the Australian Dollar / US Dollar (AUD/USD) rate, and over recent years. Although, Hansen and Lunde (2005) did

1. McKenzie (1998), Anderton and Skudelny (2001) and Choudhry and Hassan (2015), for example, demonstrate the significant impact of exchange rate volatility on international trade flows.
2. Examples, using exchange rate data, include Cumby et al (1993), Jorion (1995) and Figlewski (1997).
3. A further example of the sensitivity of forecast rankings to loss functions is Brailsford and Faff (1996) who show that asymmetric loss functions are more likely to favour forecasts from asymmetric volatility models.

examine an exchange rate in their comparison of forecasts, they use only high-frequency measures of volatility and, like many volatility forecasting studies, used the German Mark (or Euro) / US Dollar rate. The AUD/USD is the fourth highest traded currency pair (BIS, 2013) and has approximately double the trading activity of the next two highest currency pairs.[4] The AUD itself is the fifth highest traded currency, after the US Dollar, Euro, British Pound and Japanese Yen. Using this currency pair, which is still highly traded, avoids the pitfalls from over-working a dataset and has greater potential to provide new empirical findings.

While Patton examined forecasts from past squared returns, using either a rolling window average or a partial adjustment mechanism, and Hansen and Lunde (2006) examine forecasts from a small number of GARCH models, we consider a model of stochastic volatility (Harvey et al, 1994) as well as a selection of GARCH models, also relaxing the distributional assumptions in Hansen and Lunde (2006). Harvey et al (1994) showed that the stochastic volatility model fitted well to four US Dollar exchange rates, and comparisons to GARCH models for exchange rate data have been conducted by Heynen and Kat (1994), Dunis et al (2003) and Lopez (2001). However all of these studies focus on the cross rates of the US Dollar with the German Mark (or Euro), the Japanese Yen, the British Pound, the Canadian Dollar and the Swiss Franc, and none use high frequency data. By contrast, Chortareas et al (2011) do use high frequency data, but only high frequency data, in their comparison of stochastic volatility to (symmetric) GARCH models for exchange rate data, and also for a subset of the aforementioned currencies. Moreover, the balance of evidence from these studies is unable to clearly distinguish between the forecasts from these two classes of model. So, our study will compare the stochastic volatility model to a selection of GARCH models using a fresh data set, of comparable trading activity levels, that includes both low and high frequency data.

We use the set of loss functions examined in Hansen and Lunde (2005), which is broader than, but encompasses, those used by Patton (2011) and Hansen and Lunde (2006). Our selection of volatility estimators is similar to Patton (2011). As well as examining squared returns and realized variance from high frequency returns, he also considers the range based variance estimator of Parkinson (1980). While he calibrates the efficiency losses of this volatility measure relative to those of squared returns and realized variance measures, like Hansen and Lunde (2006), he only uses realized variance and squared returns in his

4. The US Dollar against the Canadian Dollar and the Swiss Franc.

empirical application. We believe our study to be the first to include all three measures in the empirical application. Moreover, we consider a different asset class and a much more recent period of time.

The question of whether there is increased convergence across loss functions of relative forecasting performance when higher quality volatility estimators are used is examined only indirectly by Hansen and Lunde (2006) and Patton (2011). Hansen and Lunde (2006) provide comparative performance tables for the forecasts from different models across different loss functions using either squared returns or realized variance, while Patton (2011) formally tests differences between such forecasts using the tests proposed by Diebold and Mariano (1995) and West (1996). In this study, we address the question directly, by conducting paired t-tests of the mean (across forecasts/models) of the standard deviation of the loss function rankings (across loss functions). A smaller value of this mean indicates a greater convergence of the rankings across loss functions, since converged rankings would generate a zero standard deviation across the loss functions for each model. We supplement this, with a related procedure, similarly constructed, that tests whether the ability of loss functions to distinguish between the first and second best models is enhanced by using more sophisticated variance estimators.

Our key findings are as follows. Using our test for ranking convergence, we find that using either a range based estimator (Parkinson, 1980) or a realized variance estimator rather than squared returns as the measure of ex-post volatility results in a significant increase in the convergence of loss function rankings. With better quality measures of volatility, loss functions converge more strongly on the preferred forecast. This is the case both for loss functions classified as robust by Patton (2011) and those not regarded as robust. Moreover, we also find a significant increase in the ability of loss functions to distinguish between the best and second best models when either the range measure or the realized variance are used. Thus, the margin by which the best model is chosen is significantly increased by using the higher quality volatility measures.

While there are no significant differential gains in terms of loss function ranking convergence between the range measure and the realized variance measure, we find that the latter is more often able to facilitate the rejection of one or more competing forecasts against a benchmark forecast. If squared returns are used as the measure of ex-post volatility, the forecasts comparison tests are unable to distinguish between any of the competing forecasts.

Taken together, these results also suggest that while the range measure may provide a significant improvement upon the use of squared returns, and so be a valuable substitute for realized variance when high frequency data is unavailable or too costly to collect, it is still not going to be as reliable as using realized variance.

We find that the regression based tests are not enhanced by the use of realized variance, which is in contrast to early studies, such as Andersen and Bollerslev (1998) but is in line with more recent studies, such as McMillan and Speight (2012) and Chortareas et al (2011). This is the case both in regards to the regression R^2 and tests of forecast unbiasedness, although the only model that generates unbiased forecasts across all measures of ex-post volatility turns out to be the best model according to the loss function based tests.

Regarding the specific characteristics of the Australian Dollar / US Dollar exchange rate, we find that the stochastic volatility model is ranked highest by the loss functions, and is significantly better than the next best model when realized variance is used as the ex-post volatility measure. The best performing GARCH model out of sample is the base-line GARCH(1,1) model, despite there being significant asymmetries in the volatility in sample. These asymmetries imply that an unanticipated depreciation of the Australian Dollar will have a greater impact of the volatility of the exchange rate than an unanticipated appreciation, relative to the US Dollar. In contrast to earlier studies, we find no evidence that power ARCH models add value to the modelling or forecasting of the exchange rate. This perhaps indicates a shift in the properties of exchange rate volatility in recent years. We also find no differences between forecasts generated assuming normally distributed errors as opposed to those from *t*-distributed errors, although the latter provides a significantly better fit in sample.

The remainder of the paper proceeds as follows. Section 2 reviews related literature to provide some context for our work and enable comparisons to our findings to be drawn. Section 3 describes the stochastic volatility model and the GARCH models that will be used to forecast the exchange rate, describes the computation of the ex-post volatility measures, and explains the forecast evaluation testing procedures. In Section 4, we describe the data and report the results of the model estimation in sample and the out of sample forecast evaluations. Section 5 summarizes and concludes.

2. Related literature

The literature on volatility forecasting and the evaluation of these forecasts is extensive, and has grown alongside the development of models of volatility.[5] However, the literature on evaluating forecasts of exchange rate volatility is relatively small, and so presents further opportunities for discovery.[6] Our work contributes to a number of literatures including, but not limited to, the literature on forecasting exchange rates, both with low and high frequency data, forecast comparisons using high frequency data, and to studies of specific financial markets, in this case the AUD/USD exchange rate.

Evaluating forecasts of the volatility of exchange rates was first undertaken by Taylor (1986) who compared forecasts from his AMARCH model with the exponentially-weighted moving average (EWMA) model in Taylor and Kingsman (1979), using daily data for the £/$ between 1974 and 1982, and marginally favouring the EWMA model. Forecast comparison has thereafter broadened out to evaluate forecasts from a variety of models in the GARCH class, stochastic volatility models, autoregressive models of squared and absolute returns and models including implied volatility from options. Bera and Higgins (1997), who examined the weekly $/£ rate between 1985 and 1991, favoured a GARCH model over a bilinear model, while Cumby et al (1993), who examine the weekly Yen/$ between 1977 and 1990 favoured an EGARCH model over historical measures of volatility. By contrast, Figlewski (1997), who examine the daily DM/$ rate over the period 1971 to 1995, favoured an historical measure over a GARCH model. Jorion (1995) who examined both the DM/$ and the Yen/$ using daily data between 1985 and 1992, also found that GARCH models added little incremental explanation of volatility compared to historical and implied measures. Comparing models within the GARCH class and a nonparametric estimator, and for 5 weekly observed cross rates with the US $ over the period 1973 to 1989, Lee (1991) found little difference between the forecasting performance of the alternative GARCH specifications, but better performance for the nonparametric specification. By contrast, West and Cho (1995) using an almost identical data set, found that the GARCH models outperformed a

5. Early studies of the volatility of financial asset returns by, for example, Mandelbrot (1963), Fama (1965), Praetz (1969), Clark (1973), and Taylor and Kingsman (1979), found that the variances of financial asset prices vary across time. The development of the ARCH models of conditional variances of time series, Engle (1982), whose properties also matched closely the empirical distributions of many asset return classes, led to a huge growth in the development of variance modelling. Early examples the application of these models to exchange rate data include Taylor (1980), Diebold (1988), Hsieh (1989), Baillie and Bollerslev (1989) and Bera and Higgins (1992). Surveys include Bollerslev et al (1992) and Bollerslev et al (1994).

6. In the survey paper by Poon and Granger (2003) of 93 papers available up to that time, fewer than 20 papers focussed on the volatility of spot exchange rates, with all focussing on some subset of the cross rates between the US Dollar, Canadian Dollar, Japanese Yen, German Mark, French Franc, Swiss Franc and Italian Lira.

nonparametric estimator, but was inferior to a constant variance assumption, but that the differences were not statistically significant. Studies that have introduced forecasts from stochastic volatility models into the comparisons include Heynan and Kat (1994) and Dunis et al (2003). They use daily data for the most active cross rates against the USD (1980-1992) and the USD and DM (1990-1998), respectively, and find that GARCH based models forecast volatility better than stochastic volatility models. By contrast, Lopez (2001) using daily data for four cross rates against the US Dollar between 1980 and 1995 could not distinguish between the forecasts of GARCH based models and a stochastic volatility model. Not only are these studies inconclusive in aggregate as to a preferred volatility forecasting model for exchange rates, the performance of any of the models in these studies is low. The typical value of the coefficient of variation (R^2) in their regression tests of forecast unbiasedness was less than 5 percent. [7]

Andersen and Bollerslev (1998) proved that regression methods would anyway give low values of R^2 because they are noisy estimates of volatility. This also explained why models that seemed to fit well in sample were being found to fail out of sample. Specifically, they show that the value of R^2 in a volatility unbiasedness regression is bounded above by the inverse of the kurtosis of the underlying returns series. They show that intra-day returns can be used to construct a realized variance (RV) measure (from the daily sum of squared intra-day returns) that eliminates the noise in measurements of daily volatility. Using 5 minute and 1 hour returns on the German Mark / US Dollar, they find that the R^2 for forecasts from a GARCH(1,1) model increased to 0.48 and 0.33, respectively. In a series of papers, Anderson et al (2001, 2003) and Barndorf-Nielsen and Shephard (2002a,b, 2004a,b) develop further the theoretical basis for using realized variance as a measure of the daily conditional variance and provide further empirical validation. Specifically, these papers show that the ex-post value of realized variance is an unbiased estimate of the conditional return variance of returns for log price processes that are special semi-martingales – in which class fit most financial models.

Forecasts of exchange rate volatility measured as intra-day realized variance have also been examined by, for example, Taylor and Xu (1997), Martens (2001), Li (2002), Pong et al (2004), Chortareas et al (2011) and McMillan and Speight (2012). Taylor and Xu (1997) find that realized variance from 5 minute returns on the DM/$ during 1992-93 contains

7. This result is not specific to exchange rates, see for example, Akgiray (1989), Pagan and Schwert (1990), and Day and Lewis (1992) for US stock indexes and Brailsford and Faff (1996) for individual Australian stocks. Somewhat higher values are reported by Blair et al (2002) for a US stock index but, for non-combined forecasts, are under 23 percent for GARCH and historical models.

incremental information for forecasting beyond that contained in implied volatility. Martens (2001) compares forecasts constructed using different intra-day intervals and finds that the higher the frequency the better the out of sample performance. Li (2002) finds that implied volatility has no incremental information relative to high frequency data, for forecasts of the volatility of the US Dollar against each of the British Pound, Japanese Yen and the German Mark, between 1994 and 1999. Pong et al (2004) find that high frequency data applied to relatively simple models can generate more accurate forecasts than those from more complex models that use lower frequency data. Chortareas et al (2011) examine four Euro exchange rates (US Dollar, Japanese Yen, British Pound and the Swiss Franc) between 2000 and 2004 and find better forecasting performance from those models that use intra-day (15 minute) returns than those that use daily returns. McMillan and Speight (2012) also examine three Euro exchange rates (against the US Dollar, British Pound and Japanese Yen) for the period 2002 to 2006. They find that forecasts from GARCH models that use intra-day data outperform those that use daily data. The contribution and focus of all these studies is, however, in showing that intra-day returns contain valuable information for exchange rate volatility forecasting, specifically that intra-day returns are a valuable input (to improving the volatility forecast) rather than (also) being a valuable output (as an improved measure of ex-post volatility). All of these studies take the use of realized variance as the ex-post measure of volatility as a given. By contrast, our focus is on showing that high frequency data can help exchange rate forecast evaluation by generating convergence among the rankings from different loss functions.

By contrast to the US Dollar, the British Pound, the Euro (or the German Mark prior to 1999) and the Japanese Yen, the Australian Dollar has received much less attention. McKenzie (1997) documents the presence of ARCH effects in 21 bilateral Australian Dollar exchange rates, finding greater effects in daily data than lower frequency data. McKenzie and Mitchell (2002) feature the Australian Dollar / US Dollar rate among a study of 17 heavily traded exchange rates between 1986 and 1997, using GARCH models. They find an asymmetry in the impact of shocks to the volatility process such that an unanticipated depreciation of the Australian Dollar has a bigger impact on future volatility than an unanticipated depreciation of the US Dollar. A similar asymmetry is reported in Villar (2010) who examines this same bilateral exchange rate, including more recent data, for the period 1994 and 2007. Comparative volatility forecasting performance, but for the Australian Dollar versus the Malaysian Ringit for the period 2010 to 2011, is considered by Ramasamy and

Munisgamy (2012). They find no asymmetry and report extremely large forecasting errors. McMillan and Speight (2004) include the Australian Dollar / US Dollar exchange rate among their data for 1990 to 1996 for forecast comparisons between three GARCH and two historical volatility measures and find that the preferred forecast is highly sensitive to the loss function, even with the use of a high frequency measure of ex-post volatility. Our paper adds to this evidence by considering alternative measures of ex-post volatility, including stochastic volatility as well as GARCH based models, calculating a wide range of loss functions and focussing on the convergence in loss function rankings when using high-frequency ex-post measures of volatility.

3. Methods

3.1 GARCH models

Forecasts for the volatility of the USD/AUD exchange rate will be generated from a set of GARCH models and a model of stochastic volatility. The models employed are listed in Table 1. GARCH based models, introduced by Engle (1982) and Bollerslev (1986), are autoregressive models of the conditional variance of a time series that depend on the squared residuals from an underlying model of the conditional mean of the time series. It is typical to parameterize the conditional mean of a time series of log changes in an exchange rate by either a low order ARMA process or a constant mean white noise process, as these series display little if any autocorrelation, see, for example, Taylor (2005).

So, the general structure of the models that we consider is

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t \qquad u_t \sim IID(0, \sigma_t^2) \tag{1}$$

where $\boldsymbol{y_t}$ is the log change in the daily closing value of the exchange rate, $\boldsymbol{\mu}$ is the constant mean, and the residual $\boldsymbol{u_t}$has mean zero and conditional variance $\boldsymbol{\sigma_t^2}$, given by one of the GARCH or stochastic volatility models in the Table 1. We assume that the residuals are independently and identically normally distributed for all models, and additionally consider the possibility that the residuals may be *t*-distributed for three of the models.[8]

Our base line volatility model, is the GARCH(1,1) specification, denoted “GARCH”, in Table 1. In this model the conditional variance depends on the size of its past lagged squared residuals, through the parameter $\boldsymbol{\alpha}$, and on its own past values, through the parameter $\boldsymbol{\beta}$. While most applications of the GARCH model consider the first lagged squared residuals and lagged own values, Hansen and Lunde (2005) did consider second lags in their forecast comparisons. However, they found that the higher order ARCH-type models rarely performed better out-of-sample than lower order lag alternatives, so we confine our study to the GARCH(1,1) specification.

In an early application of GARCH models to exchange rates, Bera and Higgins (1992), explored a non-linear version of the GARCH model, now known as the Power GARCH (or PGARCH) model. This permits the lagged residuals and the conditional standard deviation to take any positive power, including the special case of 2, which is the GARCH model, and where the power is a free parameter to be estimated. In their study of 6 US Dollar exchange rates (against the Canadian Dollar, French Franc, Swiss Franc, German Mark, British Pound and Japanese Yen) over the period 1973 to 1985, they found that the non-linear model fitted the data better than the linear (GARCH) model for some of the exchange rates.

A key consideration in the estimation of GARCH models is ensuring that the parameter estimates are constrained to prevent negative estimates of the conditional variance. Nelson (1991) proposed an exponential form of the GARCH model, now known as the EGARCH model, which models the natural log of the conditional variance rather than the conditional variance directly. The model also introduced the possibility of asymmetry in the volatility process, whereby negative shocks could have a different effect on future volatility than positive shocks. The EGARCH model captures effect of the sign of the past residual through the parameter γ and the effect of the size of the shock through the parameter α. If

8. Baillie and Bollerslev (1989) first recommended the use of the t-distribution for modelling exchange rates with GARCH processes and this recommendation is repeated in the study by Hansen and Lunde (2005).

$\gamma < 0$, then negative shocks will cause the conditional variance to rise more than it does in response to positive shocks.

The so-called "leverage effect", whereby an increase in corporate leverage can increase both the required return (which induces a decrease in the stock price) and the risk of equity, provides a theoretical basis for a negative relationship between stock returns and subsequent stock return variances, see Black (1976) and Christie (1982).[9] By contrast, for exchange rates, the possibility of an asymmetric relationship between the sign of unanticipated returns and variance is largely an empirical matter.[10] However, evidence of such an asymmetry can be interpreted as a differential volatility impact of domestic versus foreign unanticipated currency depreciations. What is striking among those studies that identify an asymmetry is that they all find that the impact of the local currency depreciating against a more major currency has a bigger impact on the exchange rate volatility than a depreciation of the major currency. Hu et al (1997) attribute this to the greater likelihood of there being a policy intervention in this situation.

We test for the presence of asymmetric effects of shocks to the conditional variances using the sign and size bias tests of Engle and Ng (1993), together with a preliminary examination of the correlation between returns and squared returns. The sign bias test indicates whether asymmetric effects may be present in the variance, while the size bias tests indicate whether the size of the shock influences the asymmetric effect of the shock. The coefficients in the following equations are estimated

Sign Bias Test: $$\hat{u}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \psi_t \tag{2a}$$

Negative Size Bias Test: $$\hat{u}_t^2 = \phi_0 + \phi_2 S_{t-1}^- u_{t-1} + \psi_t \tag{2b}$$

Positive Size Bias Test: $$\hat{u}_t^2 = \phi_0 + \phi_3 S_{t-1}^+ u_{t-1} + \psi_t \tag{2c}$$

where $\hat{u}_t^2$ is the residual from the Gaussian symmetric GARCH(1,1), S_{t-1}^- is an indicator variable that takes the value 1 if $u_{t-1} < 0$ and is zero otherwise, $S_{t-1}^+ = 1 - S_{t-1}^-$, and ψ_t is

9. There is an alternative explanation. If volatility is priced, it can feedback into returns requiring a decrease in price, see, for example French et al (1987) and Campbell and Hentschel (1992). Moreover, Duffee (1995) shows that the "leverage" effect is due to a contemporaneous positive relationship between returns and volatility that is strongest in firms with little actual leverage, and Figlewski and Wang (2001) suggests that it is a "down-market" effect that has little connection to leverage. Bollerslev et al (2006) show that high frequency data facilitates the distinction between alternative explanations of asymmetry.
10. Asymmetric effects are modelled in exchange rate volatility by, for example, Tse and Tsui (1997), Hu, Jiang and Tsoukalas (1997), McKenzie and Mitchell (2002), McMillan and Speight (2004), Hansen and Lunde (2005), Huang et al (2009), Villar (2010), Abdalla (2012) and Ramasamy and Munisamy (2012) although the relative forecasting performance of models that accommodate asymmetry is mixed.

an i.i.d error term.[11] The presence of significant $\phi_i,\ i = 1,2,3$, coefficients would indicate that there are asymmetries present that are not accommodated by the symmetric GARCH model.

The asymmetric GJR-GARCH model, Glosten et al (1993), proposes adding an interaction variable to the standard GARCH(1,1) model. This variable comprises the lagged squared residual multiplied by an indicator variable that takes the value unity when the lagged residual is positive, and takes the value zero when the lagged residual is not positive. The asymmetry works through the parameter γ. Similarly to the EGARCH model, if $\gamma < 0$, then negative shocks will cause the conditional variance to rise more than it does in response to positive shocks.[12]

The threshold GARCH (TGARCH) model of Zakoïan (1994) also captures the sign of shocks with an interaction term, but adds this to the AMARCH model of Taylor (1986), which models the conditional standard deviation (as a function of the absolute value of the residuals) rather than the conditional variance. Again, if the coefficient on this interaction term is negative, then negative shocks will cause the conditional variance to respond more than it does in response to positive shocks.

Building upon the work of Taylor (1986), Schwert (1989) and the PARCH model of Bera and Higgins (1992), Ding, Engle and Granger (1993) proposed an asymmetric power ARCH (APARCH) model. In their study of the S&P 500 daily closing prices from 1928 to 1991, they estimated the value for the power function to be 1.43 which means that the process lies between two models, namely the GARCH(1,1) d=2 and the TGARCH model, d=1. Validation of this model for exchange rate volatility is provided in studies by Tse and Tsui (1997) and McKenzie and Mitchell (2002). Both this model and the PARCH model have been shown to be able to capture well the long lag autocorrelation of some volatility series.[13] This is the final GARCH based specification that we examine and, once again, a negative sign on the term capturing asymmetry indicates a greater response to negative shocks.

---

11. The repeated notation for the coefficients and the error term across the four different equations, (2a) to (2d), is for notational simplicity and does not imply that the coefficients nor error term values would be expected to be the same across the four equations.
12. In the original formulation of the GJR GARCH model, the indicator variable takes the value of one when the shock is negative, and is zero otherwise. The reversed formulation here (indicating positive shocks) enables the interpretation of the sign of γ to remain the same across all the models incorporating asymmetry.
13. Long memory in volatility, which gives rise to such autocorrelations, can also be modelled using the Component GARCH model of Engle and Lee (1999) and the Fractionally Integrated GARCH model of Baillie et al (1996). Applications to exchange rate volatility include Vilasuso (2002), Chortareas et al (2011) and McMillan and Speight (2012). Since Lamoureux and Lastrapes (1990) have shown that structural shifts can generate spurious persistence in volatility, and extensive simulations in Ding el al (1993) show that simple GARCH models are capable of capturing the autocorrelation structure in financial data, we examine only the PARCH and APARCH models.

3.2 Stochastic Volatility

This paper will consider the simple SV model from Ruiz (1994) and Harvey, Ruiz and Sheppard (1994), in which the log of the conditional variance, σ_t^2, is assumed to follow a first order autoregressive process,

$$\ln \sigma_t^2 = \omega + \phi \ln \sigma_{t-1}^2 + \eta_t \tag{3}$$

where the residual in equation (1) is now $u_t = \sigma_t z_t$, where $z_t \sim NID(0,1)$ and where $\eta_t \sim NID\left(0, \sigma_\eta^2\right)$. Taking the log of the squared residuals, a state space form of the model can be generated, where the measurement equation is

$$\ln(u_t^2) = E(\ln(z_t^2)) + \ln \sigma_t^2 + \xi_t \tag{4}$$

and the transition equation is given by equation (3). In the measurement equation, $\xi_t = \ln(z_t^2) - E(\ln(z_t^2))$ and is a non-Gaussian, zero-mean white noise. It is assumed that ξ_t and η_t are uncorrelated (see Harvey, Ruiz and Shephard 1994, p.261), and the autoregressive parameter in the volatility equation satisfies, $|\phi| < 1$.[14] Since we assume that $z_t \sim NID(0,1)$, we know that the mean and variance of $\ln(z_t^2)$ are -1.27 and $\pi^2/2$, respectively (Abramowitz and Stegun 1970), and so ξ_t also has variance of $\pi^2/2$. This model is therefore time invariant for the variances of ξ_t and η_t. Estimation of the state space form of the SV model is undertaken using the Kalman Filter, which is quasi-maximum likelihood.

3.3 Ex-post Volatility

We will use and compare three measures of ex-post volatility, squared daily returns, realized volatility calculated from intra-day returns, and the range-based estimator of Parkinson (1980).

For a trading day with daily returns, y_t, that are the sum of N_t intraday returns, the realized volatility for day t, RV_t, is defined as,

14. As we find that empirically we cannot reject the null hypothesis that all of the ARMA coefficients and the constant term in equation (1) are zero, the residual from equation (1) is $u_t = y_t$, and so the state space form uses log squared returns in the measurement equation (3).

$$RV_t = \sum_{j=1}^{N_t} y_{t,j}^2 \tag{5}$$

where $y_{t,j}^2$ are the squared intraday returns, $j = 1,2, \ldots, N_t$. In selecting the frequency for the intra-day observations, there is a trade-off between the increase in accuracy from increased frequency and the increase in microstructure frictions as the frequency increases. While Andersen et al (2003) in their study of two major exchange rates recommend 30 minute observations, more recent studies, such as Liu et al (2015) have suggested that 5 minute intervals are the best choice. Bandi and Russell (2006) suggest that the optimal sampling frequency for the UK is between 0.4 and 13.8 minutes. We select 15 minutes, as being close to the upper bound, and note that this interval has precedent for exchange rate forecast comparisons in Chortareas et al (2011), who use it in their forecasting models.[15]

The range-based estimator proposed by Parkinson (1980), which we denote by HL_t, uses the daily high and low prices and is defined as

$$HL_t = \frac{(\log(h_t) - \log(l_t))^2}{4\log(2)} \tag{6}$$

where h_t and l_t are the daily high and low prices respectively.[16]

3.4 Forecasting volatility and Forecast Evaluation

The dynamics of the volatility processes in Table 1 provide for the estimation of one-step ahead forecasts of volatility in the (out-of-sample) evaluation period, using the estimated parameter values for the estimation period. We use four procedures for forecast evaluation: Mincer-Zarnowitz (MZ) regressions, loss function scores, forecast comparison tests and new tests of loss function ranking convergence.

In the regression test method of Mincer and Zarnowitz (1969), a measure of (ex-post) variance, σ_t^2, is regressed on a constant and the predicted variance from a postulated model, $\hat{\sigma}_t^2$, equation (7), where the residual is assumed to be mean zero and serially uncorrelated.

15. Although Liu, Patton and Sheppard (2015) do not examine spot exchange rate volatility, they also find that 5 minute and 15 minute return intervals produce the best forecasts among over 400 models applied to data between 2000 and 2010 for individual stocks, stock indexes and futures contracts on indexes, interest rates and currencies.
16. The scaling of the squared log range depends on the underlying assumed data generating process. The value here, 4ln(2), is consistent with a diffusion having zero mean and constant conditional daily volatility, see Parkinson (1980) and Patton (2011).

$$\sigma_t^2 = \alpha + \beta\hat{\sigma}_t^2 + \varepsilon_t \tag{7}$$

An unbiased forecast will have $\alpha = 0$ and $\beta = 1$, while the R^2 value from the regression provides a simple measure of the level of predictability in the volatility process. We estimate this model for each of the GARCH and SV models in turn with each of the three different ex-post estimators, squared returns, the range measure and realized variance. Since returns are heteroscedastic, squared returns will be even more heteroscedastic and so robust standard errors are calculated, although this may not be adequate since both sides of the regression are affected.

Since estimation error in $\hat{\sigma}_t^2$ can cause downward bias in the estimates of β, and the R^2 values do not penalize biased forecasts, these regressions are usually accompanied by loss function calculations. Loss functions measure the distance between the observation and forecast, and have been further adapted to penalize more extreme distances more heavily, by using quadratic functions, and to include asymmetry whereby downside and upside losses have different weightings. Patton (2011) has shown that the loss function rankings of forecasts produced by competing models are dependent on the quality of the estimator of ex-post volatility. Poor proxies for ex-post volatility can lead to the selection of not the best forecasting model. Our use of three different measures of ex-post volatility is designed to see whether the results (the rankings of models) from loss function evaluations show more convergence with more sophisticated volatility measures. We use the following set of loss functions, from Hansen and Lunde (2005):

$$MAE_1 \equiv n^{-1}\sum_{t=1}^{n}|\sigma_t - h_t| \tag{8}$$

$$MAE_2 \equiv n^{-1}\sum_{t=1}^{n}|\sigma_t^2 - h_t^2| \tag{9}$$

$$MSE_1 \equiv n^{-1}\sum_{t=1}^{n}(\sigma_t - h_t)^2 \tag{10}$$

$$MSE_2 \equiv n^{-1}\sum_{t=1}^{n}(\sigma_t^2 - h_t^2)^2 \tag{11}$$

$$QLIKE \equiv n^{-1} \sum_{t=1}^{n} (\log(h_t^2) + \sigma_t^2 h_t^{-2}) \tag{12}$$

$$R^2LOG \equiv n^{-1} \sum_{t=1}^{n} [\log(\sigma_t^2 h_t^{-2})]^2 \tag{13}$$

where n is the number of forecasts in the evaluation period. The MAE criteria (8) and (9) are likely to be more robust to outliers than the MSE criteria (10) and (11). The MSE criterion (11) is equivalent (only if the constant term, $\alpha = 0$) to using the R^2 from the MZ regression, equation (7). An MZ regression using logs (rather than levels) of variances is similarly equivalent to using R^2LOG. $QLIKE$, criterion (12), corresponds to the loss function implied by a Gaussian likelihood. We are motivated to use a wide range of loss functions from the observation of Poon and Granger (2003) that when loss-functions use variances, it becomes increasingly difficult to find a significant difference between competing forecasts, and that this is compounded when they are squared (as for MSE_2).

While loss functions provide a ranking and these may show convergence across loss functions and variance estimators, they do not determine whether, say, the first best and second best models produce forecasts that are statistically significantly different. To determine whether this is the case, we apply the modified Diebold-Mariano (DM) test of Harvey, Leybourne and Newbold (1998) (HLN-DM) to the three highest ranked forecasts from the MZ regressions and the loss function tests. The test can itself use a variety of functional forms to measure the distance between two sets of forecasts and we report results of both a quadratic and an absolute distance function. The objective is to see whether this test is more often able to distinguish between competing forecasts when the accuracy of the volatility estimator increases.

To further address the question of improved convergence of forecast rankings by loss functions when better quality volatility proxies are used, we examine the mean (across forecasts/models) of the standard deviation of the loss function rankings (across loss functions). A smaller value of this mean indicates a greater convergence of the rankings across loss functions, since converged rankings would generate a zero standard deviation across the loss functions for each model. Our test is constructed as follows. Let the rank of loss function k for model m using volatility estimator p be denoted by $r_{k,m,p}$, $k = 1,2, \ldots K$, and $m = 1,2, \ldots, M$ and similarly for volatility estimator q be denoted by $r_{k,m,q}$. The null hypothesis is that volatility estimator q provides the same amount of convergence among the loss functions as volatility estimator p, and the alternative hypothesis is the volatility estimator q provides for greater convergence than volatility estimator p. Our null hypothesis is formally that $\overline{s_p} = \overline{s_q}$, where

$$\overline{s_p} = M^{-1} \sum_{m=1}^{M} \left((K-1)^{-1} \sum_{k=1}^{K} \left(r_{k,m,p} - \overline{r_{m,p}} \right)^2 \right) \tag{14}$$

where $\overline{r_{m,p}} = K^{-1} \sum_{k=1}^{K} r_{k,m,p}$, and $\overline{s_q}$ is similarly defined. The null hypothesis can be tested using a one-side paired sample t-test. By applying this test across the different volatility proxies, we provide direct evidence on the improvements in convergence from using higher quality data.

We supplement this evidence, with further tests that determine whether the ability of loss functions to distinguish between the first and second best models is also enhanced by using more sophisticated variance estimators. We regard a volatility estimator as more able to facilitate loss functions to distinguish between two forecasts if the percentage difference between the loss function values is greater. Specifically, let $\Delta r_{k,p}$ be the percentage difference between the loss function rank for the first and second best models for loss function k using volatility estimator p, and $\Delta r_{k,q}$ be similarly defined for volatility estimator q. Then, a test of the null hypothesis that two volatility estimators provide equal facility for loss functions to distinguish between two forecasts, is a test of $\overline{\Delta r_p} = \overline{\Delta r_q}$, against the alternative that $\overline{\Delta r_p} < \overline{\Delta r_q}$, where $\overline{\Delta r_p} = K^{-1} \sum_{k=1}^{K} \Delta r_{k,p}$, and $\Delta r_{k,p} = \left(r_{k,1,p} - r_{k,2,p} \right) / r_{k,2,p}$. The mean for volatility estimator q is defined similarly. The null hypothesis is tested using a paired sample t-test.

4. Data and Results

4.1 Data

Daily data for the AUD/USD exchange rate (Australian Dollars per one US Dollar) was collected from the Bloomberg Professional service on 19/07/2016 and covers the period 06/04/2010 to 15/07/2016. To enable out-of-sample forecast evaluation, the sample is split into an estimation period and evaluation period of 1540 observations (until 29/02/16) and 99 observations (from 01/03/16) respectively. These data are used to estimate the models and generate their forecasts for the evaluation period, as well as construct two of the variance estimators for the evaluation period; the squared returns and the high-low range measure.

For the realized variance estimator of volatility, we collected 9,493 15-minute intra-day observations for the 99 days of the evaluation period, again from Bloomberg Professional. The intra-day period runs from 17:00 hours the prior day to 16.45 hours on the day, UK time (GMT+1 during this period), the latter time point coinciding with the daily closing observation. We do not include observations between 17:00 hours on Friday to 16.45 hours on Sunday, to reduce any noise from weekend trading, and adjustments for the start and end of daylight savings times.[17]

4.2 Summary Statistics

Summary statistics for daily returns, log differences in the exchange rate, are shown in Table 2. Positive returns reflect a strengthening of the US Dollar (weakening of the Australian Dollar), while negative returns reflect a strengthening of the Australian Dollar (weakening of the US Dollar). The daily mean return is not significantly different from zero ($p=0.369$), while the daily variance corresponds to an annualized standard deviation of 11.20 percent. The Jarque-Bera test (JB Test) indicates that the returns series are not normally distributed ($p<0.001$), displaying excess kurtosis and a small positive skew. An augmented Dickey-Fuller test (ADF Test) strongly rejects the null hypothesis of non-stationarity in the returns series ($p<0.001$). A test for ARCH effects in the returns series also rejects the null hypothesis of no ARCH effects ($p=0.003$).

There is no significant autocorrelation in the daily returns at lags 1 to 40, which can be seen in Table 3. By contrast, the autocorrelations of squared returns, also reported in Table 3, are statistically significant. The non-normality in returns, the significant autocorrelation in squared returns and the results of the ARCH test all point to the returns series being characterised by changing volatility. The correlation between returns and squared returns is 0.064 and is significantly different from zero ($p=0.011$).

The results of the sign bias and size bias tests are reported in Table 4 and confirm the characteristic that positive shocks are associated with a bigger impact on volatility, than negative shocks. This means that an unanticipated depreciation in the Australian Dollar has a

17. A similar approach to removing slower weekend trading periods was used by McMillan and Speight (2012) following the direction in Bollerslev and Domowitz (1993).

bigger effect on volatility than an unanticipated depreciation in the US Dollar. Specifically, we find that the coefficient, ϕ_1, in the sign bias test is negative and significant (p<0.001), indicating a negative relationship between the sign of lagged shocks and current squared residuals. This result is supported by the positive size bias test, coefficient ϕ_3, that indicates that the size of the positive shocks (unanticipated depreciation of the Australian Dollar) also influences current exchange rate volatility. By contrast, there is no evidence that the size of any unanticipated depreciation in the US Dollar influences the exchange rate volatility. This result is consistent with earlier findings for the AUD/USD rate by McKenzie and Mitchell (2002) for the period1986 to 1997 and Villar (2010) for the period 1994 to 2007.[18]

4.3 Volatility Model Estimation

The coefficients of the competing forecasting models are given in Table 4, estimated using data for the "estimation" period of the sample. Although the simplest model reported is the GARCH(1,1), we also examined ARCH models with up to 15 lags. Both the AIC and BIC selected the GARCH(1,1) as the preferred model against any of the ARCH alternatives. The GARCH likelihoods are maximized using the BHHH and BFGS algorithms in turn. For the Stochastic Volatility model, the estimation is done using a Kalman filter, with the quasi-likelihood maximized using the BHHH and Newton Raphson algorithms.

The estimated coefficients from all the models meet their respective conditions for stationarity of the variance processes, and are indicative of a high level of persistence in volatility across all specifications. The coefficients are stable across the alternative distributional assumptions, showing very little numerical differences and similar statistical significance. However, likelihood ratio tests indicate that the t-distributed versions provide a significantly better fit to the data (p<0.01).

The power models estimate the power term to be around 2.2 which is close to the GARCH baseline of 2. Likelihood ratio tests comparing the power models to their closest GARCH counterpart are unable to reject the null hypothesis that the models provide an equal fit to the data (p>0.65). This finding is similar to that of Tse and Tsui (1997), who were examining the Malasian Ringit and Singapore Dollar against the US Dollar with data from 1978 to 1994. However, this finding contrasts with those of Bera and Higgins (1992) who found that, with the exception of the GBP/USD rate, the values of the power function were all

18. Evidence of a greater impact on volatility of local currency depreciations (against a major currency) can also be found in Hu et al (1997) and Tse and Tsui (1997).

less than 1, significantly so in three out of the five US Dollar exchange rates considered. The GBP/USD rate was not significantly different from either 1 or 2, while the Swiss France / US Dollar rate was not significantly different from either 1 or 0. For data between 1986 and 1997, McKenzie and Mitchell (2002) report a power function value of 1.295 for the AUD/USD, while Villar (2010) for data between 1994 and 2007 estimates the power to be 1.279. Our results suggest that in recent years the applicability of power ARCH models for the AUD/USD exchange rate may be much reduced.

The asymmetric models all have positive coefficient values on the term that captures asymmetry, which is consistent with the findings from the sign and size bias tests and suggests that an unanticipated depreciation of the Australian Dollar has a bigger impact on exchange rate volatility than an unanticipated appreciation, against the US Dollar. The coefficients are all statistically significant, except for the APGARCH model, which again suggests that this non-linear specification is currently less well suited to the AUD/USD exchange rate.

We explore the differences between the asymmetric specifications by calculating the news impact curves for each of the specifications. Pagan and Schwert (1990) first suggested news impact curves which graphically demonstrate how different models react to positive and negative shocks. The process involves substituting positive and negative values of the standardized residuals into an estimated model and then plotting the predicted variance against the lagged standardized residuals, where the lagged conditional variance is set equal to the unconditional variance. The news impact curves are shown in Figure 1, where it can be seen that the APGARCH and GJR-GARCH models generate much more asymmetry than the EGARCH and TGARCH models. Moreover, log likelihood tests cannot distinguish between the in sample fit of the APGARCH and GJR-GARCH models ($p=0.53$), or between the EGARCH and TGARCH models ($p=0.32$). However, each of the EGARCH and TGARCH models provide a significantly better in sample fits against either of the APGARCH or GJR-GARCH models ($p<0.01$, for all four possible pairwise comparisons).

4.4 Volatility Forecast Evaluation

The evolution of the three ex-post volatility estimators during the evaluation period is shown in Figure 2. Summary statistics are shown in Table 6, and show that the three measures of ex-post volatility display the skewness and excess kurtosis characteristics that are well-established features of high frequency exchange rate data.

4.4.1 Mincer-Zarnowitz Regression Tests

The results of the MZ regressions are shown in Table 7. Although there is some variation in the rankings of the models (by R^2 value, in parentheses below the R^2 value), all three variance estimators indicate that the two best models are the EGARCH and TGARCH specifications. This means that based on this test, the more sophisticated variance estimators do not appear to generate an advantage over the use of squared returns for measuring ex-post variance. However, all of the coefficients on the predictors take the wrong sign, with many significantly negative, and the goodness of fit is poor across all models and estimators. While the R^2 values for the GARCH models do improve when the ex-post volatility measure is changed from squared returns to the range measure, they do not improve with the use of realized volatility. This is in sharp contrast to earlier studies, such as Andersen and Bollerslev (1998), who use the observed increase in R^2 for their data to, in part, empirically validate the use of realized volatility in place of squared returns. However, studies using more recent exchange rate data, such as McMillan and Speight (2012) – who specifically note this result – and Chortareas et al (2011), document similarly small values of R^2 for GARCH models using realized volatility as the ex-post measure.[19]

The unbiasedness of the forecasts can be tested with a Wald test of the joint null hypothesis of a zero intercept and a unit slope. This hypothesis is strongly rejected ($p<0.05$) for all of the models, except the stochastic volatility model, for both the squared returns and high-low variance estimator. The stochastic volatility model also provides unbiased forecasts for the realized variance estimator, where too the GARCH and PGARCH models also do not reject the joint null hypothesis ($p>0.10$). However, these results are mostly driven by the large standard errors on the slope coefficients, and the cases of unbiased forecasts are those with the least good fit.

4.4.2 Loss Function Scores

The values of the loss functions for each of the estimators and for each of the predictor models are shown in Table 8 and should provide more robust results than the MZ

19. McMillan and Speight (2012) question the robustness of earlier results, but we conjecture that it is more likely that this weak testing framework is highly susceptible to the specifics of a given sample. By contrast to McMillan and Speight (2012), Chortareas et al (2011) document this finding for some but not all of the exchange rate series analysed.

regressions.[20] The loss functions rank the stochastic volatility model as the best model, for all loss functions for squared returns and the high-low estimator, and for four of the six loss functions for the realized variance estimator. The stochastic volatility model places second and third for the MSE2 and QLIKE loss functions, respectively, when realized variance is the estimator. While these results are in sharp contrast to the rankings implied by the R-squared values obtained from MZ regressions, it was only the stochastic volatility model that provided unbiased forecasts for all three ex-post volatility estimators. Among the GARCH based estimators, the base-line GARCH(1,1) model appears to perform the best, with the PGARCH model, which anyway was parametrically close to the GARCH model, appearing to the be the next best model. However, there also appears to be some clear interaction between the choice of variance estimator and the loss function rankings for the GARCH models. For the realized variance estimator, the symmetric GARCH and PGARCH specifications produce smaller losses than the asymmetric EGARCH and TGARCH specifications across all loss functions. However, for the squared returns and the high-low estimators, the MAE loss functions produce smaller losses for the asymmetric models. However, since the realized variance estimator has the highest kurtosis (Table 6), consistent with a greater presence of outliers, the MAE loss function, which is more robust to outliers than the other loss functions, should be the most reliable in that instance. This suggests that the symmetric models are preferred for out-of-sample forecasting, even though the asymmetric models seemed to provide the better fit in-sample. This finding is echoed even within the asymmetric models themselves, when those models that permit a more pronounced asymmetry (see Figure 1) perform relatively less well out-of-sample.

4.4.3 Modified Diebold-Mariano Tests

We supplement the above tests with the forecast comparison test of Harvey, Leybourne and Newbold (1998), hereafter HLN, which is a small sample version of the test developed by Diebold and Mariano (1995). This test permits direct comparisons between pairs of forecasts, using a given loss function. The model rankings reported above suggest that the stochastic volatility is the most preferred model and so this is the model against which all other models are tested. To provide for the most conservative test, and following the observation of Poon and Granger (2003) that forecasts comparisons are most difficult for metrics that are quadratic in the variance, we focus on the MSE2 loss function. But, to provide for a

20. As the loss function scores for the predictor models with *t*-distributed errors are mostly indistinguishable from the same model applied assuming normally distributed errors, we report in Table 8 only the latter, to make the comparison across different model specifications more apparent. The former results are available on request to the authors.

comparison and because the rankings were most different for this function, we also examine the MAE2 loss function.

By increasing the accuracy of the volatility estimator, the HLN tests, reported in Table 9, are able to reject more models in favour of SV. This is in contrast to Patton (2011) who did not find significant differences between these models, using more accurate volatility estimators for these loss functions. Using the squared returns, the HLN tests are unable to distinguish statistically between forecast accuracy of the stochastic volatility model and the competing GARCH specifications. The high-low estimator allows the test to indicate that the stochastic volatility model is preferred to the asymmetric GARCH specifications, which is consistent with earlier rankings based upon the loss function statistics themselves. For at least one of the two loss functions, the use of realized variance as an estimator of volatility enables the HLN test to reject each of the competing GARCH specifications against the stochastic volatility model. While the study by Hansen and Lunde (2005) found that nothing (within a wide class of GARCH models) could beat the base-line GARCH(1,1) model for exchange rate forecasting (of the DM/$), our study complements more recent evidence, such as Chortareas (2011) who find that the stochastic volatility model may outperform the GARCH(1,1) model for some exchange rates series.

4.4.4 Loss Function Convergence Tests

The loss function rankings, also reported in Table 8, appear to show that the more sophisticated variance estimators generate more convergence among the loss functions rankings. We test this formally by conducting paired t-tests of the mean (across models) of the standard deviation of the loss function rankings (across loss functions), equation (14) above. A smaller value of this mean indicates a greater convergence of the rankings across loss functions, since (in the limit) converged rankings would generate a zero standard deviation across the loss functions for each model. We find that the high-low estimator and realized variance estimators do not display different average levels of ranking convergence ($p=0.30$), but that the rankings using squared returns are significantly less converged than those for either the high-low estimator ($p<0.01$) or the realized variance estimator ($p=0.08$). This result, which is consistent with the study by Patton (2011), indicates that employing superior variance estimators can reduce the impact a loss function's characteristics have on forecast evaluation, thus reducing the importance of loss function selection.

We conduct a further test of the benefit from using a more sophisticated variance estimator, by examining the ability of the loss functions to distinguish between the best and second best models. Loss function values across models, for a given loss function, are often very similar in magnitude, a characteristic noted by, for example, Brailsford and Faff (1996) for the Australian stock market. We calculate the average (across loss functions) of the percentage difference between the loss function scores for the best and second best models, for each of the three variance estimators. Our results echo those relating to the convergence of the rankings. We find no significant difference between the average differences for the range (high-low) and realized variance estimators (p=0.28), but significant differences between the smaller average differences for the squared returns estimators, and the larger average differences for both the high-low estimator (p=0.01) and the realized variance estimator (p=0.03). These results indicate that loss functions are much better able to distinguish between the first and second best models when either the high-low or realized variance estimators are used, than when the squared returns estimators is used.

5. Summary and conclusions

By forecasting the volatility of the Australian Dollar / US Dollar exchange rate using different proxies for ex-post volatility, we show how to measure and test the significance of the improvement in the convergence of loss function scores towards selecting the best forecast. Using our test for loss function ranking convergence, we find that using either a range based estimator or a realized variance estimator rather than squared returns as the measure of ex-post volatility results in a significant increase in the convergence of loss function rankings. With better quality measures of volatility, loss functions converge more strongly on the preferred forecast. Moreover, we also find a significant increase in the ability of loss functions to distinguish between the best and second best models when either the range measure or the realized variance are used. Thus, the margin by which the best model is chosen is significantly increased by using the higher quality volatility measures.

While we find that there are no significant differential gains in terms of loss function ranking convergence between the range measure and the realized variance measure, we do find that the latter is more often able to facilitate the rejection of one or more competing forecasts against a benchmark forecast. However, the forecasts comparison tests are unable to distinguish between any of the competing forecasts when squared returns are used as the measure of ex-post volatility. So while the range measure does provide a significant improvement upon the use of squared returns, and so be a valuable substitute for realized variance when high frequency data is unavailable or too costly to collect, high frequency data should still be used when it is available.

By contrast to early studies that advocated the use of realized variance, we do not find that the regression based tests are enhanced by the use of realized variance. However, other recent studies, such as McMillan and Speight (2012) and Chortareas et al (2011), have also noted that these regression based tests still perform poorly even with high frequency data, both in regards to the regression R^2 and tests of forecast unbiasedness. This suggests that for evaluating volatility forecasts, comparisons based upon loss functions are more likely to give reliable results. Our work supports studies that show that the use of loss functions is enhanced by the use of high frequency data for measuring ex-post volatility, and provides a simple test (using the standard deviation of the loss function rankings) of how much better the loss functions are at distinguishing between forecasts as the quality of the volatility estimator improves.

We choose the Australian Dollar / US Dollar exchange rate as this has had little empirical study yet is the fifth highest traded exchange rate; the four higher all having had their volatility extensively modelled and forecast. We find that a model of stochastic volatility provides superior forecasts of daily volatility compared to a range of GARCH models, and that the more simple GARCH formulations generate superior forecasts to those generated by more complex models featuring asymmetries and power functions. As prior studies of this exchange rate have found that asymmetric components and power functions can add to the modelling and forecasting of volatility, our results indicate a shift in the characteristics of this exchange rate in recent years. In particular, our study is the only study to consider the period post the 2008/09 financial crisis period.

References

Abdalla, Suliman Z. S. 2012. "Modelling Exchange Rate Volatility using GARCH Models: Empirical Evidence from Arab Countries." *International Journal of Economics and Finance* 4: 216-229.

Abramowitz, Milton, and Irene A. Stegun. 1970. *Handbook of Mathematical Functions.* New York: Dover Publications Inc.

Akgiray, Vedat. 1989. "Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts." *The Journal of Business* 62: 55-80.

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

Andersen, T., T. Bollerslev, F.X. Diebold, P. Labys (2001). The Distribution of Realized Exchange Rate Volatility, Journal of the American Statistical Association 96, 42-55.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. "Modeling and forecasting realized volatility." *Econometrica* 71: 579-625.

Anderton, Robert, and Frauke Skudelny. 2001. *Exchange Rate Volatility and Euro Area Imports*. ECB Working Paper No. 64.

Baillie, Richard T., Tim Bollerslev. 1989. "The message in daily exchange rates: A conditional variance tale", *Journal of Business and Economic Statistics* 7, 297-305.

Baillie, Richard T., Tim Bollerslev, and Hans O. Mikkelsen. 1996. " Fractionally integrated generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 74: 3-30.

Bandi, F.M., and J.R. Russell. 2006. "Separating microstructure noise from volatility". *Journal of Financial Economics* 79, 655-692.

Bank for International Studies. 2013. *Triennial Central Bank Survey - Foreign exchange turnover in April 2013: preliminary global results*. Monetary and Economic Department.

Barndorff-Nielsen, O.E., and N. Shephard (2002a). "Estimating Quadratic Variation Using Realized 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, Series B*, 64, 253-280.

Barndorff-Nielsen, O.E., and N. Shephard (2004a). "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 73(3), 885-926.

Barndorff-Nielsen, O.E., and N. Shephard (2004b). "How Accurate Is the Asymptotic Approximation to the distribution of Realized Volatility?," in *Identification and Inference for Econometric Models. a Festschrift in Honour of T.J. Rothenberg*, ed. by D.W.K. Andrews, J. Powell, P.A. Ruud, and J.H. Stock. Cambridge: Cambridge University Press.

Bera, Anil. K., and Matthew L. Higgins. 1992. "A Class of Nonlinear ARCH Models." *International Economic Review* 33: 137-58.

Bera, Anil. K., and Matthew L. Higgins. 1997. "ARCH and bilinearity as competing models for nonlinear dependence". *Journal of Business and Economic Statistics*, 15, 43-50.

Black, Fischer. 1976. "Studies of Stock Price Volatility Changes." Paper presented at the proceedings of the 1976 meetings of the American Statistical Association, Business and Economic Statistics Section, 177-181, Washington, D.C.

Blair, Bevan. J., Ser-Huang Poon, and Stephen 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, Tim, Julia Litvinova, and George Tauchen. 2006. "Leverage and Volatility Feedback Effects in High-Frequency Data." *Journal of Financial Econometrics* 4: 353-384

Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31: 307-327.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner. 1992. "ARCH modeling in finance." *Journal of Econometrics* 52: 5-59

Bollerslev, Tim and Ian Domowitz. (1993). "Trading patterns and prices in the interbank foreign exchange market", *Journal of Finance* 48, 1421-1443.

Bollerslev, Tim., Robert F. Engle., and Daniel B. Nelson. 1994. "ARCH models." In *Handbook of Econometrics, Volume 4*, edited by Robert F. Engle, and Daniel L. McFadden, 2961-3038. Amsterdam: Elsevier Science B.V.

Brailsford, Timothy, and Robert W. Faff. 1996. "An evaluation of volatility forecasting techniques." *Journal of Banking & Finance* 20: 419-438.

Campbell, J., and L. Hentschel, 1992. "No news is good news: An asymmetric model of changing volatility in stock returns." *Journal of Financial Economics*, 31, 281-331.

Chortareas, Georgios, Ying Jiang, and John C. Nankervis. 2011. "Forecasting exchange rate volatility using high-frequency data: Is the euro different?" *International Journal of Forecasting* 27: 1089:1107.

Choudry, Taufiq, and Syed S. Hassan. 2015. "Exchange rate volatility and UK imports from developping countries: The effect of the global financial crisis." *Journal of International Financial Markets, Institutions and Money* 39: 89-101.

Christie, Andrew A. 1982. " The stochastic behavior of common stock variances: Value, leverage and interest rate effects." *Journal of Financial Economics* 10: 407-432

Clark, Peter K. 1973. "A subordinated stochastic process model with finite variance for speculative prices." *Econometrica* 41: 135–155.

Cumby, Robert, Stephen Figlewski, and Joel Hasbrouck. 1993. "Forecasting Volatilities and Correlations with EGARCH Models." *Journal of Derivatives* 1: 51-63

Day, Theodore E.,and Craig M. Lewis. 1992. "Stock market volatility and the information content of stock index options." *Journal of Econometrics* 52: 267–287

Dickey, D.A., and W.A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, 427-431.

Diebold, Francis X., 1988. *Empirical modelling of exchange rate dynamics.* (Springer Verlag: New York).

Diebold, Francis X., & Mariano, Robert. S. (1995). "Comparing predictive accuracy". *Journal of Business and Economic Statistics* 13, 253‑265.

Ding, Zhaunxin, Clive W. J. Granger, Robert F. Engle. 1993. "A long memory property of stock market returns and a new model." *Journal of Empirical Finance* 1: 83-106.

Duffee, Gregory R. 1995. "Stock returns and volatility A firm-level analysis." *Journal of Financial Economics* 37: 399-420.

Dunis, Christian., Jason Laws and Stephane Chauvin. 2003. "FX volatility forecasts and the informational content of market data for volatility." *The European Journal of Finance*, 9, 242-272.

Engle, Robert F. 1982. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica* 50: 987-1007.

Engle R.F., and Lee, G.G.J. 1999. A permanent and transitory component model of stock return volatility. In *Cointegration,Causality and Forecasting: A Festschrift in Honour of Clive W J Granger*, Engle RF, White H (eds). Oxford University Press: Oxford; 475–497.

Engle, Robert F., and Victor K. Ng. 1993. "Measuring and Testing the Impact of News on Volatility." *The Journal of Finance* 48: 1749-1778.

Fama, Eugene F. 1965. "The behavior of stock-market prices." *Journal of Business* 38: 34–105.

Figlewski, Stephen, and Xiaozu Wang. 2001. "Is the 'Leverage Effect' a Leverage Effect?" Working Paper, Department of Finance, New York University.

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

French, K., G. Schwert and R. Stambaugh, 1987. "Expected stock returns and volatility". *Journal of Financial Economics,* 19, 3-30.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. 1993. "On the Relation between the Expected Value and Volatility of the Nominal Excess Return on Stocks." *The Journal of Finance* 48: 1779-1801.

Hansen, Peter R., and Asger Lunde. 2005. "A Forecast Comparison of Volatility Models: Does Anything Beat A GARCH(1,1)?" *Journal of Applied Econometrics* 20: 873-889.

Hansen, Peter R., and Asger Lunde. 2006. "Consistent ranking of volatility models". *Journal of Econometrics* 131, 97-121.

Harvey, Andrew C., Esther Ruiz, and Neil Shephard. 1994. "Multivariate stochastic variance models." *Review of Economic Studies* 61: 247–264.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold. 1998. "Test for Forecast Encompassing." *Journal of Business & Economic Statistics* 16: 254-259.

Heynen, Ronald C., and Harry M. Kat. 1994. "Volatility prediction: a comparison of the stochastic volatility, GARCH(1, 1) and EGARCH(1, 1) models." *Journal of Derivatives* 2: 50–65.

Hsieh, D.A. 1989. "Modeling Heteroscedasticity in daily Foreign-Exchange Rates", *Journal of Business and Economic Statistics,* 7, 307 - 317.

Hu, Michael Y., Christine X. Jiang, and Christos Tsoukalas. 1997. "The European exchange rates before and after the establishment of the european monetary system." *Journal of International Financial Markets, Institutions and Money* 7: 235-253.

Huang, Fei-xue, Yan Zhao, and Tie-shan Hou. 2009. " Long Memory and Leverage Effect of Euro Exchange Rate Based on ARFIMA-FIEGARCH." Paper presented at the International Conference on Management Science & Engineering (16th), Moscow, Russia, September 14-16.

Jarque, Carlos M., and Bera, Anil K., 1987. "A Test for Normality of Observations and Regression Residuals," *International Statistical Review* 55, 163-172.

Jorion, Philippe. 1995. "Predicting Volatility in the Foreign Exchange Market." *Journal of Finance* 50: 507-528.

Lamoureux, Christopher G., and Wiliam D. Lastrapes. 1990. "Persistence in variance, structural change and the GARCH model." *Journal of Business and Economic Statistics* 8: 225-234.

Lee, Keun Yeong. 1991. "Are the GARCH models best in out-of-sample performance?" *Economics Letters* 37, 305-308.

Li, Kai. 2002. "Long-memory versus option-implied volatility prediction", *Journal of Derivatives* 9, 9-25

Lopez, Jose A. 2001. “Evaluating the predictive accuracy of volatility models”, *Journal of Forecasting* 20, 87-109.

Liu, Lily, Andrew J. Patton, and Kevin Sheppard. 2015. " Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes." *Journal of Econometrics* 187: 293-311.

G. M. Ljung; G. E. P. Box, 1978. "On a Measure of a Lack of Fit in Time Series Models". *Biometrika*. 65, 297–303.

Mandelbrot, Benoit. 1963. "The Variation of certain speculative prices." *The Journal of Business* 36: 394-419.

Martens, M., 2001. “Forecasting daily exchange rate volatility using intraday returns.” *Journal of International Money and Finance* 20, 1–23.

McKenzie, Michael D., (1997). “ARCH modelling of Australian bilateral exchange rate data”, *Applied Financial Economics* 7, 147-164.

McKenzie, Michael D., (1998), “The Impact of Exchange Rate Volatility on Australian Trade Flows,” *Journal of International Financial Markets, Institutions and Money,* 8, pp.21-38.

McKenzie, Michael, and Heather Mitchell. 2002. "Generalised Asymmetric Power ARCH Modeling of Exchange Rate Volatility." *Applied Financial Economics* 12: 555-564.

McMillan, David, G., and Alan E.H. Speight (2004) “Daily Volatility Forecasts: Reassessing the Performance of GARCH Models”, *Journal of Forecasting* 23, 449–460.

McMillan, David, G., and Alan E.H. Speight (2012). “Daily FX Volatility Forecasts: Can the GARCH(1,1) Model be Beaten using High-Frequency Data?”, *Journal of Forecasting* 31, 330–343.

Mincer, Jacob A., and Victor Zarnowitz. 1969. "The Evaluation of Economic Forecasts." *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research

Nelson, Daniel B. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59: 347-370.

Pagan, Adrian R., and G William Schwert. 1990. "Alternative Models for Conditional Stock Volatility." *Journal of Econometrics* 45: 267-290.

Parkinson, Michael. 1980. "The extreme value method for estimating the variance of the rate of return." *Journal of Business* 53: 61–65.

Patton, Andrew J. 2011. "Volatility Forecast Comparison Using Imperfect Volatility Proxies." *Journal of Econometrics* 160: 246-256.

Pong, Shiuyan Pong, Mark B Shackleton, Stephen J Taylor, Xinzhong Xu (2004) Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models, *Journal of Banking and Finance*, 28, 2541–2563.

Poon, Ser-Huang, and Clive W. J. Granger. 2003. "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature* 41: 478-539

Praetz, P. D. 1969. "Australian Share Prices and the Random Walk Hypothesis." *Australian & New Zealand Journal of Statistics* 11: 123-139.

Ramasamy, Ravindram, and Shanmugam Munisamy. 2012. "Predictive Accuracy of GARCH, GJR and EGARCH Models Select Exchange Rates Application." *Global Journal of Management and Business Research* 12: 88-100

Rodríguez Villar, Maria José. 2010. "Volatility Models with Leverage Effect." PhD thesis, Universidad Carlos III de Madrid.

Ruiz, Esther. 1994. "Quasi-maximum likelihood estimation of stochastic volatility models." *Journal of Econometrics* 63: 289-306.

Schwert, G.William. 1989. "Why does stock market volatility change over time?" *Journal of Finance* 44: 1115–1153.

Taylor, Stephen J. 1980. "Conjectured Models for Trends in Financial Prices, Tests and Forecasts." *Journal of the Royal Statistical Society* 143: 338-362.

Taylor, Stephen J. 1986. *Modelling financial time series*. Chichester:Wiley.

Taylor, Stephen J. 2005. *Asset Price Dynamics, Volatility, and Prediction*. New Jersey: Princeton University Press.

Taylor, Stephen J., and Brian G. Kingsman. 1979. "An analysis of the variance and distribution of commodity price changes." *Australian Journal of Management* 4: 135–149.

Taylor, Stephen J., and Xinzhong Xu. 1997. “The incremental volatility information in one million foreign exchange quotations”. *Journal of Empirical Finance* 4, 317-40.

Tse, Yiu K., and Albert K. C. Tsui. 1997. "Conditional volatility in foreign exchange rates: evidence from the Malaysian ringgit and Singapore dollar." *Pacific-Basin Finance Journal* 5: 345-356.

Vilasuso, Jon. 2001. "Forecasting exchange rate volatility." *Economics Letters* 76: 59-64.

West, Kenneth. D., and Dongchul Cho. 1995. "The predictive ability of several models of exchange rate volatility." *Journal of Econometrics* 69: 367–391.

West, Kenneth D., 1996. “Asymptotic inference about predictive ability.” *Econometrica* 64, 1067–1084.

Zakoïan, Jean-Michel. 1994. "Threshold heteroskedastic models." *Journal of Economic Dynamics and Control* 18: 931–955.

Zhou, Bin. 1996. "High-frequency data and volatility in foreign-exchange rates." *Journal of Business and Economic Statistics* 14: 45–52.

Table 1: Volatility Model Specifications

Model	Variance Specification	
GARCH	$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2$	
t-GARCH		
GJR	$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma u_{t-1}^2 S_{t-1}^+$	
t-GJR		
EGARCH	$\ln(\sigma_t^2) = \omega + \gamma(z_{t-1}) + \alpha[\lvert z_{t-1}\rvert - E(\lvert z_{t-1}\rvert)] + \beta \ln(\sigma_{t-1}^2)$ where $z_t = {}^{u_t}/_{\sigma_t}$	$E(\lvert z_{t-1}\rvert) = \sqrt{\frac{2}{\pi}}$
t-EGARCH		$E(\lvert z_{t-1}\rvert) = \frac{2\sqrt{v-2}\,\Gamma\left[\frac{(v+1)}{2}\right]}{(v-1)\Gamma\left(\frac{v}{2}\right)\sqrt{\pi}}$
PGARCH	$\sigma_t^d = \omega + \alpha \lvert u_{t-1}\rvert^d + \beta \sigma_{t-1}^d$	
APGARCH	$\sigma_t^d = \omega + \alpha(\lvert u_{t-1}\rvert + \gamma u_{t-1})^d + \beta \sigma_{t-1}^d$	
TGARCH	$\sigma_t = \omega + \alpha \lvert u_{t-1}\rvert + \beta \sigma_{t-1} + \gamma \lvert u_{t-1}\rvert S_{t-1}^+$	
SV	$\ln(y_t^2) = -1.27 + \ln(\sigma_t^2) + \xi_t \qquad \text{var}(\xi_t) = \frac{\pi^2}{2}$ $\ln(\sigma_t^2) = \omega + \phi \ln(\sigma_{t-1}^2) + \eta_t \qquad \eta_t \sim NID(0, \sigma_\eta^2)$	

Notes: The variable v is the degrees of freedom, which is the parameter governing the density function of the *t* distribution. The function Γ is the gamma function, which is defined by an integral, $\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx$, $u > 0$. All other variables and parameters are as defined in the text.

Table 2: Summary Statistics of the Daily Returns

Summary statistics for daily returns on the AUD/USD exchange rates between 06/04/2010 to 29/02/2016. Mean is the mean daily return. Std. Dev. Is the standard deviation of the returns. Skew is the Skewness and Kurt is the Kurtosis of the daily returns. JB is the Jarque-Bera (1987) test for normality. ADF is the Augmented Dickey Fuller (1979) test (no trend) using the SIC criterion to select lag length. ARCH is the Engle (1982) test for ARCH effects in the variance of the returns. P-values are in brackets.

Mean × (10^{-3})	Std. Dev.	Skew.	Kurt.	JB	ADF	ARCH
0.1655	0.0072	0.0813	4.916	237.3	-40.00	8.86
[0.369]				[0.000]	[0.000]	[0.003]

Min	Q1	Median	Q3	Max
-0.0356	-0.0040	0.0001	0.0042	0.0360

Table 3: Autocorrelations of Returns and Squared Returns

Autocorrelations for daily returns, y_t and squared daily returns y_t^2 on the AUD/USD exchange rates between 06/04/2010 to 29/02/2016. Q stats are the Ljung Box (1978) statistics. P-values for the autocorrelations and Q statistics are given underneath in brackets.

	Autocorrelations at lag										Q stats at lag	
	1	2	3	4	5	6	7	8	9	10	20	40
y_t	-0.0204	0.0321	-0.0429	0.0147	-0.0073	0.0069	-0.0089	0.0163	-0.0349	-0.0049	17.53	48.08
	[0.422]	[0.327]	[0.166]	[0.247]	[0.358]	[0.473]	[0.575]	[0.635]	[0.534]	[0.625]	[0.619]	[0.178]
y_t^2	0.0778	0.1797	0.0794	0.1005	0.1388	0.1304	0.0898	0.1304	0.0841	0.1453	380.10	574.31
	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Table 4: Coefficient Estimates from the Sign and Size Bias Tests

Coefficient estimates from the following regressions

$$\hat{u}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \psi_t$$

$$\hat{u}_t^2 = \phi_0 + \phi_2 S_{t-1}^- u_{t-1} + \psi_t$$

$$\hat{u}_t^2 = \phi_0 + \phi_3 S_{t-1}^+ u_{t-1} + \psi_t$$

where where $\hat{u}_t^2$ is the residual from a GARCH(1,1) model applied to the daily returns on the AUD/USD exchange rates between 06/04/2010 to 29/02/2016, S_{t-1}^- is an indicator variable that takes the value 1 if $u_{t-1} < 0$ and is zero otherwise, $S_{t-1}^+ = 1 - S_{t-1}^-$, and ψ_t is an i.i.d error term.

Bias Test	$\boldsymbol{\phi_0}$ $(\times \mathbf{10^{-5}})$	$\boldsymbol{\phi_1}$	$\boldsymbol{\phi_2}$	$\boldsymbol{\phi_3}$
Sign	5.95 (0.000)	-1.45×10^{-5} (0.006)	–	–
Negative Size	5.50 (0.000)		1.07×10^{-3} (0.085)	–
Positive Size	4.56 (0.000)		–	2.44×10^{-3} (0.000)

Table 5 - Coefficient Estimates from the GARCH and Stochastic Volatility Models

Coefficient estimates for the models given in Table 1 above applied to the daily returns on the AUD/USD exchange rates between 06/04/2010 to 29/02/2016. P-values are given in parentheses. Log L is the maximized value of the likelihood function. Kurt is the kurtosis of the standardized residusals. Q(40) is p-value of the Ljung-Box statistic (with 40 lags).

	$\mu(\times 10^{-3})$	ω	α	β	υ	γ	d	Log L	Kurt	Q(40)
GARCH	0.143	0.483	0.056	0.936				5507.8	3.658	0.875
	(0.382)	(0.051)	(<0.001)	(<0.001)						
GARCH(t)	0.093	0.461	0.054	0.938	10.095			5518.6	3.660	0.873
	(0.554)	(0.055)	(<0.001)	(<0.001)						
GJR	0.262	0.344	0.004	0.958		0.064		5516.9	3.677	0.863
	(0.105)	(0.074)	(0.737)	(<0.001)		(<0.001)				
GJR(t)	0.182	0.297	<0.001	0.960	10.79	0.067		5527.0	3.695	0.858
	(0.581)	(0.075)	(0.981)	(<0.001)		(<0.001)				
EGARCH	0.263	-0.093	0.091	0.991		0.050		5513.4	3.686	0.861
	(0.002)	(0.033)	(<0.001)	(<0.001)		(<0.001)				
EGARCH(t)	0.185	-0.089	0.090	0.991	10.56	0.052		5523.8	3.693	0.860
	(0.178)	(0.036)	(<0.001)	(<0.001)		(<0.001)				
PGARCH	0.144	0.175	0.052	0.935			2.200	5507.9	3.656	0.876
	(0.379)	(0.750)	(0.001)	(<0.001)			(<0.001)			
APGARCH	0.257	0.865	0.021	0.957		0.626	2.270	5517.1	3.672	0.866
	(0.110)	(0.791)	(0.365)	(<0.001)		(0.326)	(0.002)			
TGARCH	0.261	0.600	0.021	0.956		0.051		5512.9	3.709	0.849
	(0.098)	(0.040)	(0.039)	(0.001)		(<0.001)				
		γ	ϕ	σ_η^2						
Stochastic Volatility		-0.116	0.989	0.007				-3378.60	6.002	0.002
		(0.030)	(<0.001)	(0.015)						

Table 6: Summary Statistics for Ex-post Volatility

Summary statistics for three ex-post measures of volatility for the AUD/USD exchange rates between 01/03/2016 to 15/07/2016. Mean is the mean value. Std. Dev. is the standard deviation. of the returns. Skew is the Skewness and Kurt is the Kurtosis. ADF is the Augmented Dickey Fuller (1979) test (no trend) using the SIC criterion to select lag length. Q(40) is the Ljung-Box statistic (with 40 lags). P-values are in brackets.

	Mean $\times (10^{-3})$	Std. Dev. $\times (10^{-3})$	Skew.	Kurt.	ADF Test	Q(40)
Squared returns	0.0688	0.0996	2.695	11.26	-9.474	35.16
	[0.000]				[0.000]	[0.688]
Range (High-Low)	0.0644	0.0836	5.729	43.67	-9.427	20.49
	[0.000]				[0.000]	[0.996]
Realized variance	0.0700	0.0777	6.544	52.87	-9.119	23.09
	[0.000]				[0.000]	[0.985]

Table 7: Coefficient Estimates, R^2 and unbiasedness test statistics from Mincer-Zarnowitz Regressions

Coefficient estimates (p-values in parentheses below) from the regression equations $\boldsymbol{\sigma_t^2 = \alpha + \beta\hat{\sigma}_t^2 + \varepsilon_t}$, where $\boldsymbol{\sigma_t^2}$ is one of the three measures of ex-post volatility and $\boldsymbol{\hat{\sigma}_t^2}$ is the forecast of this volatility provided by one of the models in Table 1 (and using coefficients estimates from Table 5). $\boldsymbol{R^2}$is the coefficient of determination of the regression, and "rank" is the rank of the model's forecast ordered by $\boldsymbol{R^2}$. Wald p-value is the p-value of the joint hypothesis that $\boldsymbol{\alpha = 0, \beta = 1}$, that is, that the forecast is unbiased. SV is the stochastic volatility model.

	Squared Returns				High-low				Realized Variance			
Model	$\alpha \times 10^{-4}$ (p-value)	β (p-value)	R^2 (rank)	Wald p-value $\alpha = 0, \beta = 1$	$\alpha \times 10^{-4}$ (p-value)	β (p-value)	R^2 (rank)	Wald p-value $\alpha = 0, \beta = 1$	$\alpha \times 10^{-4}$ (p-value)	β (p-value)	R^2 (rank)	Wald p-value $\alpha = 0, \beta = 1$
GARCH	1.592 (0.013)	-1.376 (0.147)	0.0215 (5)	0.0438	1.407 (0.009)	-1.162 (0.144)	0.0218 (8)	0.0268	1.027 (0.040)	-0.499 (0.503)	0.0046 (9)	0.1165
t-GARCH	1.666 (0.011)	-1.485 (0.128)	0.0237 (4)	0.0393	1.455 (0.008)	-1.232 (0.132)	0.0232 (7)	0.0258	1.058 (0.040)	-0.544 (0.47i97)	0.0052 (7)	0.1165
GJR	1.222 (0.003)	-0.853 (0.174)	0.0190 (7)	0.0121	1.309 (0.000)	-1.062 (0.042)	0.0419 (4)	0.0006	1.117 (0.001)	-0.666 (0.173)	0.0190 (4)	0.0026
t-GJR	1.183 (0.003)	-0.787 (0.189)	0.0177 (8)	0.0114	1.267 (0.000)	-0.992 (0.047)	0.0400 (5)	0.0005	1.097 (0.000)	-0.633 (0.176)	0.0188 (5)	0.0020
EGARCH	1.648 (0.002)	-1.564 (0.062)	0.0354 (3)	0.0081	1.615 (0.000)	-1.581 (0.024)	0.0515 (3)	0.0014	1.282 (0.002)	-0.948 (0.149)	0.0214 (3)	0.0079
t-EGARCH	1.640 (0.002)	-1.552 (0.056)	0.0371 (2)	0.0063	1.614 (0.000)	-1.582 (0.020)	0.0548 (1)	0.0009	1.289 (0.001)	-0.960 (0.131)	0.0234 (1)	0.0057
PGARCH	1.546 (0.014)	-1.308 (0.161)	0.0201 (6)	0.0469	1.388 (0.008)	-1.136 (0.147)	0.0216 (9)	0.0260	1.030 (0.036)	-0.504 (0.491)	0.0049 (8)	0.1063
APGARCH	1.146 (0.004)	-0.7188 (0.222)	0.0153 (9)	0.0142	1.241 (0.000)	-0.938 (0.056)	0.0370 (6)	0.0006	1.070 (0.001)	-0.581 (0.206)	0.0165 (6)	0.0026
TGARCH	1.731 (0.002)	-1.716 (0.050)	0.0390 (1)	0.0067	1.661 (0.000)	-1.674 (0.022)	0.0527 (2)	0.0015	1.307 (0.002)	-0.999 (0.145)	0.0217 (2)	0.0084
SV	1.182 (0.037)	-0.820 (0.372)	0.0082 (10)	0.0997	0.971 (0.041)	-0.544 (0.480)	0.0051 (10)	0.1221	0.747 (0.091)	-0.078 (0.913)	0.0001 (10)	0.1514

Table 8: Loss Function Values and Ranks

The estimated values of the loss functions given in equations (8) to (13) for forecasts of daily volatility of the AUD/USD exchange rate between 01/03/2016 to 15/07/2016, usingeach of the models in Table 1 (normally distributed errors) against each of the three measures of ex-post volatility given in Table 5. SV is the stochastic volatility model. The rank of the loss function value, ordered across the models, are given in parentheses below the loss function value.

	Squared Returns						High-low						Realized variance					
	MSE1 $\times 10^{-5}$	MSE2 $\times 10^{-9}$	QLIKE	R2LOG	MAE1 $\times 10^{-3}$	MAE2 $\times 10^{-5}$	MSE1 $\times 10^{-5}$	MSE2 $\times 10^{-9}$	QLIKE	R2LOG	MAE1 $\times 10^{-3}$	MAE2 $\times 10^{-5}$	MSE1 $\times 10^{-5}$	MSE2 $\times 10^{-9}$	QLIKE	R2LOG	MAE1 $\times 10^{-3}$	MAE2 $\times 10^{-5}$
GARCH	2.678 (3)	10.28 (3)	−8.525 (1)	5.779 (4)	4.197 (5)	6.543 (5)	1.261 (2)	7.281 (2)	−8.608 (2)	0.748 (2)	2.576 (4)	4.369 (4)	0.853 (2)	6.129 (1)	−8.538 (1)	0.358 (2)	1.815 (2)	3.326 (2)
GJR	2.758 (6)	10.57 (6)	−8.491 (7)	5.839 (6)	4.212 (6)	6.590 (6)	1.375 (6)	7.703 (6)	−8.548 (6)	0.797 (6)	2.657 (6)	4.517 (6)	0.969 (6)	6.630 (6)	−8.475 (7)	0.414 (6)	1.955 (6)	3.550 (6)
EGARCH	2.706 (5)	10.49 (5)	−8.496 (4)	5.773 (3)	4.159 (3)	6.482 (3)	1.307 (5)	7.513 (5)	−8.569 (5)	0.760 (5)	2.570 (3)	4.356 (3)	0.907 (5)	6.463 (5)	−8.497 (5)	0.377 (5)	1.844 (5)	3.364 (5)
PGARCH	2.678 (2)	10.28 (2)	−8.525 (2)	5.782 (5)	4.196 (4)	6.541 (4)	1.263 (3)	7.290 (3)	−8.606 (3)	0.748 (3)	2.577 (5)	4.371 (5)	0.866 (3)	6.229 (3)	−8.537 (2)	0.360 (3)	1.823 (3)	3.339 (4)
APGARCH	2.781 (7)	10.59 (7)	−8.492 (6)	5.875 (7)	4.242 (7)	6.645 (7)	1.396 (7)	7.749 (7)	−8.547 (7)	0.812 (7)	2.693 (7)	4.583 (7)	0.978 (7)	6.647 (7)	−8.477 (6)	0.420 (7)	1.977 (7)	3.593 (7)
TGARCH	2.694 (4)	10.48 (4)	−8.494 (5)	5.753 (2)	4.141 (2)	6.451 (2)	1.297 (4)	7.492 (4)	−8.570 (4)	0.753 (4)	2.553 (2)	4.325 (2)	0.902 (4)	6.453 (4)	−8.498 (4)	0.373 (4)	1.82 (4)	3.337 (3)
SV	2.595 (1)	10.23 (1)	−8.525 (3)	5.583 (1)	4.053 (1)	6.310 (1)	1.194 (1)	7.179 (1)	−8.610 (1)	0.690 (1)	2.427 (1)	4.122 (1)	0.827 (1)	6.214 (2)	−8.530 (3)	0.331 (1)	1.672 (1)	3.088 (1)

Table 9: HLN Forecast Comparison Tests

P-values of the HLN (1998) forecast comparison test applied to forecasts from models drawn from Table 1 for each of the three ex-post volatility measures. The test is conducted using both an MSE and an MAE criterion. SV is the stochastic volatility model.

	Squared returns		High-low		Realized variance	
Model comparison	MSE	MAE	MSE	MAE	MSE	MAE
SV Vs. GARCH	0.814	0.260	0.457	0.131	0.970	0.052
SV Vs. EGARCH	0.267	0.578	0.073	0.384	0.011	0.121
SV Vs. PGARCH	0.822	0.297	0.477	0.148	0.917	0.035
SV Vs. TGARCH	0.242	0.633	0.066	0.427	0.00	0.130

Figure 1: News Impact Curves estimated from the model parameters

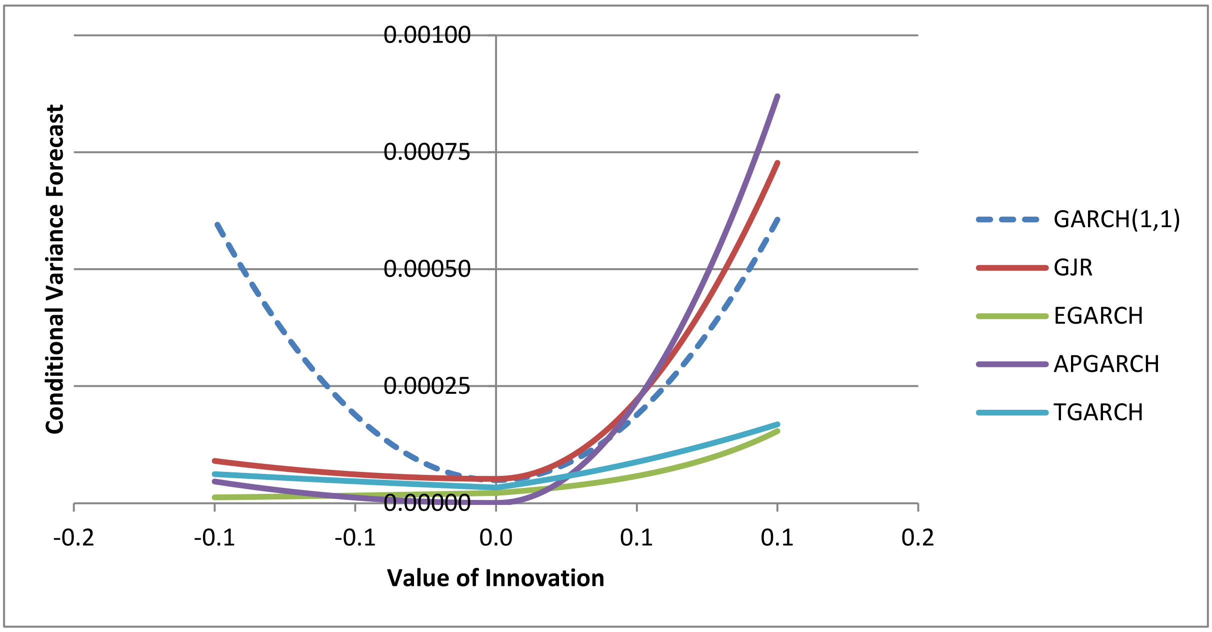

Figure 2: Evolution of the alternative estimators of ex-post volatility