

Banco de México
Documentos de Investigación

Banco de México
Working Papers

N° 2009-01

**Forecasting Exchange Rate Volatility: The Superior
Performance of Conditional Combinations of Time
Series and Option Implied Forecasts**

Guillermo Benavides
Banco de México

Carlos Capistrán
Banco de México

January 2009

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

Forecasting Exchange Rate Volatility: The Superior Performance of Conditional Combinations of Time Series and Option Implied Forecasts*

Guillermo Benavides[†]
Banco de México

Carlos Capistrán[‡]
Banco de México

Abstract: This paper provides empirical evidence that combinations of option implied and time series volatility forecasts that are conditional on current information are statistically superior to individual models, unconditional combinations, and hybrid forecasts. Superior forecasting performance is achieved by both, taking into account the conditional expected performance of each model given current information, and combining individual forecasts. The method used in this paper to produce conditional combinations extends the application of conditional predictive ability tests to select forecast combinations. The application is for volatility forecasts of the Mexican Peso–US Dollar exchange rate, where realized volatility calculated using intra-day data is used as a proxy for the (latent) daily volatility.

Keywords: Composite Forecasts, Forecast Evaluation, GARCH, Implied volatility, Mexican Peso - U.S. Dollar Exchange Rate, Regime-Switching.

JEL Classification: C22, C52, C53, G10.

Resumen: Este documento provee evidencia empírica de que combinaciones de pronósticos de volatilidad implícitos en opciones y pronósticos de series de tiempo condicionadas en información actual son estadísticamente superiores a modelos individuales, combinaciones no condicionales, y pronósticos híbridos. El desempeño superior en términos de pronóstico se obtiene tanto por tomar en cuenta el desempeño esperado de cada modelo individual condicionado en información actual, como por combinar los modelos individuales. El método utilizado en este documento para producir las combinaciones condicionales extiende la aplicación de pruebas condicionales de habilidad de predicción a la selección de combinaciones de pronósticos. La aplicación es para pronósticos de volatilidad del tipo de cambio Peso Mexicano - Dólar Estadounidense, donde la volatilidad realizada calculada utilizando datos intra-día es utilizada como una aproximación de la volatilidad diaria (latente).

Palabras Clave: Cambio de Régimen, Evaluación de Pronósticos, GARCH, Pronósticos Compuestos, Tipo de Cambio Peso Mexicano - Dólar Estadounidense, Volatilidad Implícita.

*We thank Alejandro Díaz de León, Antonio E. Noriega, Carla Ysusi, and seminar participants at the 2008 Latin American Meeting of the Econometric Society at Rio de Janeiro, the XII Meeting of CEMLA's Central Bank Researchers' Network at Banco de España, the 2008 Meeting of the Society of Nonlinear Dynamics and Econometrics at the Federal Reserve Bank of San Francisco, Banco de México, ITAM, ITESM Campus Cd. de México, and Universidad del Valle de México for helpful comments. We also thank Antonio Sibaja and Pablo Bravo for helping us with the exchange rate intraday data. Andrea San Martín, Gabriel López-Moctezuma, Luis Adrián Muñiz, and Carlos Muñoz Hink provided excellent research assistance.

[†] Dirección General de Investigación Económica. Email: gbenavid@banxico.org.mx

[‡] Dirección General de Investigación Económica. Email: ccapistran@banxico.org.mx

1 Introduction

Volatility forecasts are fundamental in several financial applications. For instance, volatility inputs are widely used for portfolio optimization, hedging, risk management, and pricing of options and other type of derivatives (Taylor, 2005). Given that financial volatility is a measure of risk, policy makers often rely on market volatility to have an idea of the vulnerability of financial markets and the economy (Poon and Granger, 2003). Furthermore, many decisions are taken anticipating what could occur in the future, thus, a forecast of the volatility of financial variables is a relevant piece of information.

There are basically two classes of models used in volatility forecasting: models based on time series, and models based on options (Poon and Granger, 2003). Among the time series models, there are models based on past volatility, such as historical averages of squared price returns, Autoregressive Conditional Heteroskedasticity-type models (ARCH-Type), such as ARCH, GARCH, and EGARCH, and stochastic volatility models. Among the options based volatility models, typically called option implied volatilities (IV), there are the Black-Scholes-type models (Black and Scholes, 1973), the model-free, and those based on hard data on volatility trading.¹

Even though several models are widely used by academics and practitioners to forecast volatility, nowadays there is no consensus about which method is superior in terms of forecasting accuracy (Poon and Granger, 2003; Taylor, 2005; Andersen et al., 2006).² Some authors believe that time series volatility forecasting models are superior because they are specifically designed to capture the persistence of volatility, a salient feature of financial volatility (Engle, 1982; Bollerslev, 1986). Others believe that implied volatility is informationally superior to forecasts based on time series models because it is the “market’s” forecast, and hence it may be based on a wider information set and also may have a forward looking component (Xu and Taylor, 1995).

In part, this has led researchers to suggest that combining a number of volatility forecasts may be preferable. Patton and Sheppard (2007), and Andersen et al. (2006), among others, highlight the importance of analyzing in more detail composite specifications for volatility forecasting. Becker and Clements (2008) show that combination forecasts of S&P 500 volatility are statistically superior to individual forecasts. Pong et al. (2004) and Benavides (2006) show that combinations of backward-looking and forward-looking forecasts can also be suc-

¹For exchange rate over-the-counter option markets investors often trade with implied volatilities. This is why there is hard data available of implied volatilities.

²Other methods to forecast financial volatility have been suggested. These are: nonparametric, neural networks, genetic programming and models based upon time change and duration. However, it has been found that these have relatively less predictive power and the number of publications using these methods is substantially lower (Poon and Granger, 2003).

cessfully used to forecast exchange rate volatility. This is intuitively appealing given that the volatility obtained from forward-looking methods may have different dynamics than the volatility obtained from backward-looking ones. Thus, combining them could be useful to incorporate features of several forecasting methods in one single forecast, aiming at obtaining a more realistic prediction of the volatility of a financial asset. In addition, combining forecasts has had a very respectable record forecasting other economic and financial variables (Timmermann, 2006).

There is some evidence in the combination literature, specifically to forecast macroeconomic variables, that time-varying combination schemes that condition the weights on current and past information may outperform linear combinations (Deutch et al., 1994; Elliott and Timmermann, 2005; Aiolfi and Timmermann, 2006; Guidolin and Timmermann, 2006). This time-varying type of forecast combinations may be particularly well suited for forecasting financial volatility due to the observed volatility clustering effect. If the dynamics of high volatility are well captured by a particular method (or set of methods), whereas the dynamics of low volatility are well captured by another method (or set of methods), then a time-varying combination of the forecasts may be the right tool to capitalize on their comparative advantage. This may be true even if one method seems to have an absolute advantage, or an absolute disadvantage. At this regard, Poon and Granger (2003) and Patton and Sheppard (2007) have suggested that more research should be carried out on this issue, given that there is almost no published research about it.

Two specific time-varying combinations captured our attention, one proposed by Deutch et al. (1994) with weights that change through time in a discrete manner, and the “Hybrid” forecast of Giacomini and White (2006) (GW), that recursively selects the best forecast. In Deutch et al.’s (1994) methodology, the appropriate combination weights can be used given an indicator function that points out the future regime, although they do not propose a method to come about the indicator function. GW, in contrast, propose a technique, based on their conditional predictive ability test (CPA), to diagnose if current information can be used to determine which forecasting model will be more accurate at a specific future date. GW explore the model-selection implications of adopting a conditional perspective with a simple example of a two-step decision rule that tests for equal performance of the competing forecasts and then –in case of rejection- uses currently available information to select the best forecast for the future date of interest.

In this paper we show that the two-step procedure proposed by GW to select forecasting methods can be extended to select forecast combinations. This results in a conditional combination of the type suggested by Deutch et al. (1994), but with the advantage that we can test if current information can be used to select the future regime. The extension is

simple, as it involves the use unconditional combinations in the second step.

In order to evaluate our proposed methodology, we first evaluate the forecasting accuracy of some of the most commonly used methods for financial volatility forecasting, as well as combinations of them, using data on the Mexican peso (MXN)–U.S. dollar (USD) exchange rate. The methods applied in this study are: 1) univariate Generalized Autoregressive Conditional Heteroskedasticity Models (GARCH; Bollerslev, 1986; Taylor, 1986); 2) model-based implied volatilities obtained using Garman and Kohlhagen’s (1983) procedure; 3) hard data of implied volatilities from quotes recorded on trading in specific over-the-counter option’s exchange rate deltas; 4) linear combinations of the aforementioned models’ forecasts; and, 5) time-varying combinations, or what we call “conditional combinations”.³

Our results indicate that statistically superior out-of-sample accuracy in terms of Mean Squared Forecast Errors (MSFEs) is achieved by conditional combinations of GARCH and (hard-data) IV exchange rate volatility forecasts. These time-varying combinations have weights that vary according to the past level of volatility, based on the fact that when the level of realized volatility is high, IV tends to perform better, whereas when the level of volatility is low, GARCH models tend to be relatively more accurate. Thus, our proposed conditional combinations take into account the comparative advantage of each, backward- and forward-looking volatility forecasting methods.

This study is carried out using MXN–USD intraday exchange rate data to construct a proxy for ex-post volatility to use as a benchmark for forecast evaluation purposes. As shown by Andersen and Bollerslev (1998), the intraday data can be used to form more accurate and meaningful ex-post proxies for the (latent) daily volatility than those calculated using daily data. To the best of our knowledge, high frequency data for this specific exchange rate has not been analyzed anywhere in terms of volatility forecasting.

The layout of this paper is as follows. Section 2 discusses the methodology used to obtain the individual forecasts, to combine them in linear and non-linear ways, and to evaluate forecast performance. The data and our proxies for ex-post realized volatility are presented in Section 3. Section 4 contains the empirical results. Finally, Section 5 concludes.

³Evidence about the performance of some of these methods for the case of the volatility of the peso-dollar exchange rate, using daily data, can be found in Benavides (2006).

2 Methodology

2.1 Individual models

All the forecasts produced by the individual models described in this paper are one-step-ahead forecasts (i.e., one-trading-day-ahead). A detailed description of each of the models used follows.

2.1.1 ARCH-type

The first class of models are the ARCH-type models. It is well documented that ARCH-type models can provide accurate estimates of asset price volatility. This is because these type of models capture the time-varying behavior and clustering commonly observed in the volatility of financial data. The predictive out-of-sample superiority of ARCH-type against historical models has been shown in several studies.⁴ For the case of exchange rate volatility, to mention a few, we have the following: Cumby et al., (1993), Guo (1996), and Andersen and Bollerslev (1998), among others. For other type of assets there are mix results. However, most of them show that ARCH-type do better than historical models.⁵ Most of the literature show ARCH-type models from a predictive accuracy framework. However, there is less evidence that these type of models increase the information content in simple regression analysis. For example, Manfredo et al., (2001) find, in a Mincer-Zarnowitz (MZ) regression context, that the explanatory power of out-of-sample forecasts is relatively low. In most cases, R^2 s are below 10%. Nonetheless, recent studies have shown that using ex-post realized volatility measures based on intraday data improves the explanatory power in these type of regressions. It has been documented that the coefficient of determination increases about three times compared to the one using volatility measures based only on daily data (Andersen et al., 2006).

In particular, in this paper we apply an univariate GARCH(1,1).⁶ This parsimonious model was chosen from the ARCH-family given the evidence presented in Hansen and Lunde (2005): In a forecast comparison among 330 ARCH-type models, they find no evidence that a GARCH(1,1) model is outperformed in their analysis of exchange rates.⁷ In addition, asym-

⁴Excluding ARFIMA models from the historical method.

⁵See, among others, Taylor (1986) for interest rate futures, Akgiray (1989) for stocks, and Manfredo et al. (2001) for agricultural commodities.

⁶ARCH-LM tests following Engle's (1982) methodology were carried out to corroborate that the series under study has ARCH effects. The results indicate that the series rejected the null in favor of ARCH effects. The procedure was carried out using up to seven lags.

⁷A specification search for the lag order was also carried out using information criteria. Akaike Information Criterion (AIC) and Schwartz Bayesian Information Criterion (BIC) were applied. In both cases the selected model was a GARCH(1,1).

metric volatility models (EGARCH, GJR, QGARCH, among others) were not considered given that they are not theoretically justifiable for exchange rate volatility modeling. This is because there is no proven statistical evidence that exchange volatilities have asymmetric volatility (Engle and Ng, 1993).⁸ Finally, fractionally integrated ARCH models were not analyzed due to an important drawback: in some occasions they produce a time trend in volatility, but time trends are usually not observed in volatility (Granger, 2003).

The formulae for the GARCH(1,1) are as follows (Andersen et al., 2006). The first equation is:

$$r_t = \mu_{t|t-1} + \sigma_{t|t-1}z_t \quad (1)$$

$$z_t \sim i.i.d., E(z_t) = 0, Var(z_t) = 1,$$

where r_t is the return at time t , calculated as the logarithmic difference of the exchange rate, i.e. $r_t = p_t - p_{t-1}$ where p is the natural logarithm of the exchange rate, $\mu_{t|t-1}$ is the conditional mean, $\sigma_{t|t-1}$ is the conditional variance, and z_t denotes a mean zero, variance one, serially uncorrelated error process. The GARCH(1,1) model for the conditional variance is:

$$\sigma_{t|t-1}^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1|t-2}^2, \quad (2)$$

where $\varepsilon_t \equiv \sigma_{t|t-1}z_t$ and the parameters are restricted in order to ensure that the conditional variance remains positive, $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. We use a t-distribution for the error process.

The GARCH(1,1) model is estimated applying the standard procedure, as explained in Bollerslev (1986) and Taylor (1986), but using rolling windows. Two rolling windows of fixed size were used. One contains 756 observations (approximately three years of data), and the other one contains 1526 observations (approximately six years of data). The fixed size window was chosen considering that six years of daily data is enough to capture the volatility dynamics of the series having robust coefficients with relatively low standard errors.⁹ Recursive estimation was not used because the conditional predictive ability test to be applied later can not be used when the forecasts are obtained using expanding windows (Giacomini and White (2006) rule out expanding window forecasting schemes by assumption). The parameters, including the degrees of freedom of the t-distribution, were estimated using maximum likelihood applying the Marquandt procedure.¹⁰

⁸In fact, no asymmetric volatility was found for the data considered for this study. To test for asymmetric volatility two methods were carried out. These were the Engle and Ng (1993) method and an analysis of the cross-correlations of the squared vs. non-squared standardized residuals.

⁹Notice that nowadays there is no consensus about which method to use in order to find an optimal rolling window size (Pesaran and Timmermann, 2007).

¹⁰Given that these type of models are well documented in the literature, no more details about these are given here. If more details are needed the interested reader can consult Bollerslev et al. (1994). In

2.1.2 Option-implied

It is widely known that implied volatilities from options prices are accurate estimators of the price volatility of their underlying assets (Fleming, 1998; Blair, Poon and Taylor, 2001; Taylor, 2005). The forward-looking nature of IV is intuitively appealing and theoretically different from the well-known historical backward-looking conditional volatility ARCH-type models.

Within the academic literature there is evidence that the information content of estimated IV from options could be superior to those estimated with time series approaches (see, among others, Fleming, Ostdiek and Whaley (1995) for futures market indexes; Fleming (1998), Blair, Poon and Taylor (2001) for stocks; Ederington and Guan (2002) for futures options of the S&P 500; and Manfredo, Leuthold and Irwin (2001) and Benavides (2003) for agricultural commodities). On the other hand, some research papers are skeptical about their out-of-sample forecasting accuracy (see, among others, Day and Lewis, 1992; Canina and Figlewski, 1993; and Lamoureux and Lastrapes, 1993).

In terms of forecasting exchange rate volatility, most of the literature has found that IV has both higher accuracy and information content. The evidence is supported by the works of Jorion (1995), Szakmary et al. (2003), Pong et al. (2004), and Benavides (2006).¹¹ Even though there is not a consensus about which method forecasts with higher accuracy, for exchange rate volatility forecasting there is a tendency to observe that IV has better performance compared to its counterparts (Poon and Granger, 2003).

The option IV of an underlying asset is the market's forecast of the volatility of such asset during the life of the option. IV is obtained implicitly from the options written on the underlying asset. In the present paper two types of IV are used. One of them is from price quotes expressed as IV corresponding to an option's delta and the other one is model-based.

Considering that volatility is the only unobservable variable in an option-valuation model, traders in over-the-counter option markets do the trading on volatility-quotes. This latter measurement is hard-data of IVs, which are commonly used by over-the-counter foreign exchange option traders. By a mutual agreement, once the option traders set an IV quote for a transaction it is then substituted into an option-pricing model (usually Garman and Kohlhagen's 1983 model (GK)) in order to determine the option's price in monetary terms. The final figure obtained is in monetary units, thus the buying and selling is settled in cash even though the original trade was agreed on volatility annualized percentages.

the present paper these calculations were performed using the UCSD GARCH toolbox for Matlab freely available at www.kevinshppard.com.

¹¹Pong et al. (2004) found that implied volatility forecasts performed at least as well as forecasts from historical long-memory models, specifically, Autoregressive Fractional Integrated Moving Average Models, for one and three month time horizons.

To calculate the model-based option implied volatility of an asset, an option valuation model together with inputs for that model are needed. The inputs for a typical currency option valuation model are domestic risk-free interest rate, i , foreign risk-free interest rate, i_f , time to maturity, T , spot price of the underlying asset, S , the exercise price, X , and the market price of the option. For the present study, the GK option pricing model is used. This is a model derived from the Black – Scholes formula and it is commonly used in the literature to price currency options. The GK model is:

$$\begin{aligned}
c(X, T) &= Se^{-i_f T} N(d_1) - Xe^{-iT} N(d_2), \\
p(X, T) &= Xe^{-iT} N(-d_2) - Se^{-i_f T} N(-d_1), \\
d_1 &= \frac{\ln\left(\frac{S}{X}\right) + \left(i - i_f + \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}, \\
d_2 &= d_1 - \sigma\sqrt{T},
\end{aligned} \tag{3}$$

where c is the value of the European-style call option, p is the value of the European-style put option, and $N(x)$ is the cumulative probability distribution function, which is Gaussian. σ^2 is the annualized price-return variance. The implied volatility is calculated by a backward induction technique solving for the only unobserved variable, σ^2 , in the call option price function.¹² For each trading day, these implied volatilities are derived from nearby to expiration over-the-counter one month option contracts for the MXN–USD exchange rate. Since for this paper a forecast for the next trading day is needed, we transform the annualized variance by dividing it by 252, the approximate number of trading days in one year.

2.2 Linear combinations

Another method used to forecast financial volatility is to combine different forecasting models, resulting in a composite forecast. The purpose of this method is to combine such models in order to obtain a more accurate forecast. The motivation to use a composite approach is mainly related to forecast errors. It is commonly observed that individual forecast models generally have less than perfectly correlated forecast errors. Each of the models in the composite approach is expected to add significant information to the model as a whole, given that the statistical observed difference in their estimated errors is not perfectly correlated. It is also well known that the variance of out-of-sample errors can be reduced considerably with composite forecast models (Clemen, 1989; Timmermann, 2006).

Composite approaches to forecasting started to be formally considered at least since the

¹²The calculation was performed iteratively until the pricing error was less than 0.00001. These calculations were performed using Visual Basic for Applications.

late 1960s, with the seminal work of Bates and Granger (1969). Reviews of the now extensive literature on this topic are provided by Clemen (1989) and Timmermann (2006). However, their use for forecasting volatility has been more scarce.

The majority of the research work in the literature about composite models suggests the application of linear combinations. Granger and Ramanathan (1984) consider three regressions (the original notation has been changed to adapt it to our context):

$$\begin{aligned}
 (GR1) \quad \tilde{\sigma}_{t+1}^2 &= \omega_0 + \boldsymbol{\omega}' \hat{\boldsymbol{\sigma}}_{t+1|t}^2 + \varepsilon_{t+1} \\
 (GR2) \quad \tilde{\sigma}_{t+1}^2 &= \boldsymbol{\omega}' \hat{\boldsymbol{\sigma}}_{t+1|t}^2 + \varepsilon_{t+1} \\
 (GR3) \quad \tilde{\sigma}_{t+1}^2 &= \boldsymbol{\omega}' \hat{\boldsymbol{\sigma}}_{t+1|t}^2 + \varepsilon_{t+1}, \text{ s.t. } \boldsymbol{\omega}' \boldsymbol{\iota} = 1,
 \end{aligned} \tag{4}$$

where $\tilde{\sigma}_{t+1}^2$ refers to some form of ex-post volatility measure, $\hat{\boldsymbol{\sigma}}_{t+1|t}^2$ is a K -vector of ex-ante volatility forecasts, and $\boldsymbol{\iota}$ is a $K \times 1$ vector of ones. The first and second of these regressions can be estimated by standard ordinary least squares (OLS), the only difference being that the second equation omits an intercept term. The third regression omits an intercept and can be estimated through constrained least squares. The third specification is motivated by an assumption of unbiasedness of the individual forecasts. Imposing that the weights sum to one then guarantees that the combined forecast is also unbiased.¹³ The procedure suggested by Granger and Ramanathan has two desirable characteristics, it yields a combined forecast, which is usually better than either of the individual forecasts, and the method is easy to implement.

In this paper, the three regressions are used. They are estimated recursively. In addition, for some variables it has been found that a linear combination using equal weights (i.e., simply averaging the forecasts) yields good results in terms of MSFE (Timmermann, 2006). For this reason, we also consider this linear combination, where the weights, in terms of the above notation, are equal to $\frac{1}{K}$, and $\omega_0 = 0$.

2.3 Conditional combinations

In some occasions, depending on the particular dynamics of the series to be forecasted, simple combinations as those considered above may not be flexible enough. This is more likely to be true if the performance of each of the forecasts to be combined changes conditional on current and past information.

In this context, Deutch et al. (1994), among others, pointed out that combinations using time-varying weights may have better performance than simple linear combinations as they

¹³This specification may not be efficient, however, due to the constraints imposed.

may be able to capture some of the non-linearities present in the data. A simple case of time-varying weights is that of regime switching. For two regimes, there can be a set of combination weights for one regime and a different set for the other regime. An important question arises of how to distinguish the regimes, i.e., how to determine when to switch. One approach in the combination literature has been to use observable variables, another has been to use latent variables. The case of observable variables has been studied by Deutch et al. (1994). The case of latent variables has been studied by Elliott and Timmermann (2005).¹⁴ In the spirit of Deutch et al. (1994), in this paper composite forecast models with time-varying weights are estimated, hence the present study belongs to the former case.

In particular, Deutch et al. (1994) propose combining forecasts using changing weights (the original notation has been changed to adapt it to our context):

$$\begin{aligned}\widehat{\sigma}_{t+1,t,c}^2 &= I_{(\cdot)} (\alpha_0 + \boldsymbol{\alpha}'\widehat{\sigma}_{t+1|t}^2) \\ &\quad + (1 - I_{(\cdot)}) (\beta_0 + \boldsymbol{\beta}'\widehat{\sigma}_{t+1|t}^2),\end{aligned}$$

where $\widehat{\sigma}_{t+1,t,c}^2$ is the combined forecast, and $I_{(\cdot)}$ is an indicator function. They examine several choices to construct the indicator function, such as past forecast errors or relevant economic variables in an application to inflation forecasting. Although they succeed in showing that time-varying methods can result in a substantial reduction in MSFE, they do not propose a way to select the variables used to determine the regimes (i.e., to calculate the indicator function), nor do they have a way to test if the time-varying combination is worth looking at.¹⁵

When working with switching regressions, the problem of finding the appropriate variable (or set of variables) to determine a regime is usually encountered. This problem may be even more problematic in a forecasting context, as what is needed is a variable that is able to predict the regime in the future. What we propose in this paper is to use the CPA test of Giacomini and White (2006), to select the appropriate regime. Using this technique, a variable or a set of variables can be tested to see if they can predict which forecast will have a better performance for a particular period in the future. The present paper uses this technique in order to construct an indicator function that can be used to estimate time-varying combinations as in Deutch et al. (1994).

Giacomini and White (2006) propose a two-step decision rule that uses current information to select the best forecast, between a pair of forecasts, for the future date of interest.

¹⁴The latent variable in this context refers to the variable or variables used to determine the regime, and not to the fact that the volatility is itself not observable.

¹⁵Our procedure, to be developed later, allows to test if time-varying combinations are worth looking at.

The first step performs a CPA test, where the null hypothesis is:

$$\begin{aligned} H_0 : E \left[(\hat{\sigma}_{t+1}^2 - \hat{\sigma}_{t+1|t,1}^2) - (\hat{\sigma}_{t+1}^2 - \hat{\sigma}_{t+1|t,2}^2) \mid \mathcal{F}_t \right] \\ = E \left[e_{t+1|t,1}^2 - e_{t+1|t,2}^2 \mid \mathcal{F}_t \right] = 0 \quad t = 1, 2, \dots, \end{aligned} \quad (5)$$

where $\hat{\sigma}_{t+1|t,i}^2$ is the ex-ante volatility forecast produced with model i , $e_{t+1|t,i}$ is the forecast error from model i , and \mathcal{F} denotes an information set. From (5), the following orthogonality condition can be derived:

$$E \left[h_t (e_{t+1|t,1}^2 - e_{t+1|t,2}^2) \right] = 0 \quad \text{for all } \mathcal{F}_t \text{ - measurable functions } h_t.$$

In this context, h_t is known as the test function. The test can be performed using the (out-of-sample) regression:

$$\begin{aligned} e_{t+1|t,1}^2 - e_{t+1|t,2}^2 = \delta' h_t + \epsilon_t \\ H_0 : \delta = 0. \end{aligned}$$

A rejection is interpreted as implying that h_t contains information to predict which forecast will perform better for the future date of interest. Hence, the first step consists on applying GWs conditional predictive ability test to see if it is possible to find a variable or a set of variables that can predict the future performance of each of the forecasts. The objective is to predict the forecast that will have the smaller loss for the next period.

In case of a rejection, GW's second step entails the following decision rule: use $\hat{\sigma}_{t+1,t,2}^2$ if $\hat{d}_{t+1|t} \geq 0$ and use $\hat{\sigma}_{t+1,t,1}^2$ if $\hat{d}_{t+1|t} < 0$, where $\hat{d}_{t+1|t} = \hat{\delta}' h_t$ is the predicted loss differential. Hence, the second step uses the information from the CPA tests in order to generate an indicator function that can be used to distinguish between two regimes. When the loss differential points out that one forecasts will have a better performance in the future, the indicator function selects that forecast. This is a particular type of combination with changing weights where the weights are either zero or one, and there is a variable or set of variables that are used to select the regime.

In this paper, we extend the **second** step in GW, and propose the following decision rule: use $\alpha_0 + \alpha_1 \hat{\sigma}_{t+1,t,1}^2 + \alpha_2 \hat{\sigma}_{t+1,t,2}^2$ if $\hat{d}_{t+1|t} \geq 0$ and use $\beta_0 + \beta_1 \hat{\sigma}_{t+1,t,1}^2 + \beta_3 \hat{\sigma}_{t+1,t,2}^2$ if $\hat{d}_{t+1|t} < 0$. In this context, we would expect $\hat{\sigma}_{t+1,t,2}^2$ to receive relatively more weight when $\hat{d}_{t+1|t} \geq 0$ and $\hat{\sigma}_{t+1,t,1}^2$ to receive relatively more weight when $\hat{d}_{t+1|t} < 0$. Hence, we do not go all the way, as GW, to completely switch from one forecast to the other, but maintain the advantages of combinations **within** regimes.

Once we obtain the indicator function from GW's first step, we estimate the time-varying combination using OLS on the following regression:

$$\begin{aligned} \tilde{\sigma}_{t+1}^2 = & I_{(\hat{d}_{t+1|t} \geq 0)} (\alpha_0 + \alpha_1 \hat{\sigma}_{t+1,t,1}^2 + \alpha_2 \hat{\sigma}_{t+1,t,2}^2) \\ & + \left(1 - I_{(\hat{d}_{t+1|t} > 0)}\right) (\beta_0 + \beta_1 \hat{\sigma}_{t+1,t,1}^2 + \beta_3 \hat{\sigma}_{t+1,t,2}^2) + \varepsilon_{t+1}. \end{aligned} \quad (6)$$

2.4 Evaluation of forecasting performance

The metric employed in the present paper to test for predictive accuracy is MSFE, calculated for each type of forecast i as: $MSFE_i = P^{\frac{1}{2}} \sum_{t=1}^P (\tilde{\sigma}_t^2 - \hat{\sigma}_{t|t-1,i}^2)^2$, where P is equal to the number of out-of-sample observations. The choice of loss function considers its simplicity, its generalized use, and that it gives consistent model rankings for commonly used volatility proxies (Patton, 2006).¹⁶

The method with the smaller MSFE is considered the most accurate volatility forecasting method. In order to compare the statistical significance of the MSFEs we use Diebold and Mariano's (1995) test (the Diebold-Mariano-West test, West (1996)).

In order to assess the quality of the ex-ante volatility forecasts, we need an ex-post measure of volatility to use as a benchmark (e.g., to calculate forecast errors and as the dependent variable in combination regressions). A measure typically used in the literature of volatility forecast evaluation is the squared daily return. However, recent research has shown that daily squared returns used as a realized volatility measure are a noisy proxy for that day's true (latent) volatility (Taylor, 2005; Andersen et al., 2006).

Instead, several papers have emphasized the benefits of using intraday data to calculate a proxy for financial volatility. Parkinson (1980) was about the first to propose the use of this type of high-frequency data. Other authors coincide: Garman and Klass (1980), Taylor (1986), and Andersen and Bollerslev (1998), among others. Realized volatility constructed with intraday data has been proven to be a less noisy proxy for the (latent) daily volatility (Andersen and Bollerslev, 1998). The use of intraday data is based on the idea that, during a period of time, volatility can be estimated more efficiently as the frequency of returns increases, providing that the intraperiod returns are uncorrelated to each other. However, a trade-off exists, since having the most frequent intraday quotes, say one minute, could increase the bias due to market microstructure effects (e.g., noise from the bid-ask spread, discreteness of the price grid).

¹⁶Patton (2006) derive necessary and sufficient conditions on the functional form of the loss for the ranking of volatility forecasts to be robust to the presence of noise in the volatility proxy. He shows that MSE loss is robust. Realized volatility is one of the volatility proxies used in Patton (2006) to derive this result.

The realized variance using intraday data can be calculated as:

$$RV_t = \sum_{j=1}^{1/\Delta} [p_{t+j\Delta} - p_{t+(j-1)\Delta}]^2, \quad (7)$$

for Δ small, positive, and with $1/\Delta \gg 1$. Notice that with $\Delta = 1$ the above measure coincides with the squared daily return. Andersen and Bollerslev (1998) suggest that fixing Δ at 5 minutes is the frequency that favorably trade-offs the bias and inconsistency induced by microstructure noise and the efficiency achieved by using higher frequencies. Andersen et al. (2006) point out that: “the actual empirical evidence for a host of actively traded assets indicate that fixing the Δ somewhere between 5 minutes and 15 minutes typically works well” (p. 859). In fact, sampling prices at 5 minutes intervals is probably the most popular choice nowadays.

For the rest of this paper, we take the realized volatility calculated as the realized variance using intraday data at five minutes intervals as our ex-post proxy for volatility. Therefore, the evaluation of the forecasting performance of different models is made with respect to this benchmark. However, the exchange rate that we use is not traded as actively as the exchange rates on which the evidence for using 5 minutes intervals is based on. Therefore, we will also report a robustness exercise using as a benchmark the squared daily return.

3 Data

3.1 Daily and intradaily returns

The daily data for the spot exchange rate MXN–USD consists of daily spot prices obtained from Banco de México’s web page.¹⁷ These are daily averages of quotes offered by major Mexican banks and other financial intermediaries. The sample period is from January 2nd, 1998 to December 31st, 2007 for a total of 2,499 observations.

The intraday data are realizations of the MXN–USD exchange rate with a frequency of 5 minutes. The transactions were carried out through the Reuters electronic platform. The exact reference is Reuters Matching (RIC: MXN=D2). We considered transactions from the 96th observation to the 240th observation. This interval was chosen given that almost 95 percent of the transactions fall within this range. Weekends and holidays were excluded due to significant infrequent trading. The sample size for intraday data is from January 2nd, 2004 to December 31st, 2007. The sample size was chosen considering data availability.¹⁸

¹⁷Banco de México’s is México’s Central Bank, with web page: <http://www.banxico.org.mx>

¹⁸There were 10 missing trading days in our data base (the first two weeks in May 2004). For those days

The observations are taken at each 5-minute interval or the last observation if there is no observation at the exact time interval.

3.2 Realized volatility

The measures of ex-post volatility that we use as the target variable are calculated using equation (7). Our preferred measure uses the intraday data. But we also calculate realized volatility using daily data. Figure 1 shows both estimates for the period from January 2nd, 2004 to December 31st, 2007. As can be seen, the estimate that uses intraday data at five minute intervals captures relatively well the dynamics of the volatility estimated using daily data. However, the former is less noisy. This reflects Andersen and Bollerslev's (1998) mathematical result that the intraday estimator has lower variance (as $\Delta \rightarrow 0$, the variance decreases).

3.3 Implied volatility

The volatility implied in option data is calculated from daily over-the-counter (OTC) options for 1-month to maturity contracts of the MXN-USD exchange rate. The hard data on IV was downloaded from UBS -an international financial institution home based in Switzerland- (the ticker is 1MDNMXNUSDImplied). The sample period for the option data is from January 2nd, 2004 to December 31st, 2007, which consists of 993 daily observations. Mexican and US risk-free interest rates were obtained in order to estimate the IVs from the model-based method. Interest rates from 1-month Mexican Federal Government bonds (CETES) were downloaded from Banco de Mexico's web page. US CDs with the same maturity were downloaded from the Board of Governors of the Federal Reserve System's web page.¹⁹ Model-based IVs are calculated from January 2nd, 2004 to December 31st, 2007.

4 Empirical results

In the following exercises, the out-of-sample evaluation period is from January 2nd, 2004 to December 31st, 2007 for the individual methods, and from January 2nd, 2006 to December 31st, 2007 for all the methods when the combinations are included among the forecasting methods. In the latter, data for 2004 and 2005 are used to start the recursive estimation of the combination weights.

the daily quotes were used instead (as published by Banco de México). The missing data represent less than 1% of the total number of observations considered for the out-of-sample evaluations.

¹⁹The Federal Reserve's web page is <http://www.federalreserve.gov/>

4.1 Individual models

Figure 2 shows the forecast errors of the individual methods. The abbreviations represent the following: implied volatility model-based (ivmb), hard data on implied volatility one-week-to-maturity (iv1week) and three-week-to-maturity (iv3week), and GARCH rolling using a window of 3 years (GARCHroll3) and a window of 6 years (GARCHroll6) of daily data. From the figure it is clear that the implied volatility model-based does not provide good forecasts, compared with the others. Ivmb forecasts are very biased (systematically over-predict realized volatility) and have high variance.

Table 1 contains descriptive statistics about the performance of the individual methods. Those with negative mean errors tend to systematically over-predict realized volatility, with option implied model based showing considerably higher bias. The last column shows the MSFE ratio, using as common denominator the MSFE of the GARCH with a rolling window of 6 years given that it is the individual model that performed better in terms of MSFE. The ratio of the IV model based is extremely high, showing that indeed this method is particularly bad to forecast realized volatility. In contrast, the differences between both GARCHs and the IVs (hard data) are smaller, with apparently a small advantage for the GARCHs.²⁰

Table 2 contains the t-statistics (above the diagonal) and the p-values (below the diagonal) of pair-wise Diebold-Mariano-West tests, where we have used Newey and West's (1987) heteroskedasticity and autocorrelation consistent covariance matrix estimator. Small p-values can be interpreted as a rejection of the null hypothesis of equal predictive ability. The table shows evidence that the IV model based is statistically worse than all the other methods. If inference is conducted at the 5% level, it is clear that: (i) both IV hard data show statistically the same performance (p-value is 0.3711); (ii) the GARCH estimated using a rolling window of 6 years is not statistically superior to the one estimated using a rolling window of 3 years (p-value is 0.1164); and (iii) both IV hard data show statistically the same performance than both GARCH models (pairwise), i.e., they have statistically indistinguishable MSFEs, as the DMW predictive ability tests are not able to reject the null of equal predictive accuracy between these forecasts, with the exception of the pair GARCHroll6 and iv1week, for which there is some evidence of superior performance of the ARCH-type forecast.

Our results confirm that indeed it is difficult to choose between forecasts produced using time series methods and those obtained from financial markets (option implieds).

²⁰We also evaluated the performance of historical volatility forecasts, using different estimation windows, from 5 days to 252 days, but in all cases they perform poorly, even when compared to the option implied model based forecasts.

4.2 Linear Combinations

The descriptive statistics about the performance of the linear combinations are presented in Table 3. All the combinations include forecasts from the five different methods reported in Table 1. As a reference, the statistics for the GARCH model estimated using 6 years windows are also presented, although in contrast to tables 1 and 2, here the analysis is performed using only forecasts for 2006 and 2007, since the data for 2004 and 2005 is used to start the (recursive) estimation of the combination weights. As can be seen, even the combination that includes a constant (GR1) has a negative mean error, although smaller than that of the other combinations, as expected. Looking at the MSFE ratios, the combination with equal weights does not produce good results. In addition, three of the combinations present a MSFE ratio smaller than 1, which indicates that their MSFEs are smaller than that of the best individual model (GARCHroll6) calculated over the same reduced sample, confirming that combining forecasts is a procedure that improves forecasting accuracy. However, the results of the DMW tests applied to the linear combinations (taken the GARCHroll6 as the benchmark) show that only GR1 and GR2 are able to outperform the best individual model at the 5% significant level.

It is interesting that the best two combinations are GR1, which is the less restrictive of the Granger-Ramanathan regressions, and GR2, the combination that does not include a constant. The latter result is a consequence of: (i) the constant in most of the regressions is close to zero; and (ii) in general, the estimated weights for the GARCH forecasts are usually negative (around -0.5), while the weights estimated for the IVs are usually positive (around 1), which offsets the bias of the individual forecasts. Indeed, the gains from combination arise from a logic similar to the logic behind diversification when forming investment portfolios. In this case, because of the negative sign of the weights, the best “portfolio” usually goes short on the GARCHs models. The heterogeneity in the weights also explains the poor performance of the mean forecasts in this context, and opens the door for the use of time-varying combination schemes that can make better use of it.

4.3 Conditional Combinations

In this section, we first proceed with the conditional evaluation of the individual forecasts (the first step in GW). Then, we use these results to form an indicator function and then to estimate the time-varying combinations, following equation (6). Finally, we evaluate the performance of the proposed time-varying composite forecasts.

Table 4 presents the results of the conditional predictive ability test for the individual

forecasts, using forecasts from January 2, 2004 to December 30, 2005.²¹ The table presents, below the diagonal, the p-values of GW tests using just a constant as an instrument. These tests are very similar to the DMW test, in the sense that they can be interpreted as unconditional tests. In this case, at the 1% level there are four null hypotheses that cannot be rejected, those corresponding to the pairs GARCHroll6-GARCHroll3, iv1week-GARCHroll6, iv1week-GARCHroll3, and iv3weeks-GARCHroll3.²² The interpretation of these results is that, unconditionally, these pairs have MSFEs that are statistically equal. Above the diagonal, the table presents p-values of CPA tests using as instruments lagged values of the realized volatility, as this variable easily suggests itself due to the volatility clustering, and lagged values of the estimated loss differential, d_t . In this case, the null of equal conditional predictive accuracy is rejected for only two of these four pairs. For the pair GARCHroll6-iv1week at the 10% level, and for the pair iv3weeks-GARCHroll3 at the 1% level. These results indicate that for these pairs the instruments chosen can be used to predict which method will be more accurate the next day.²³

The important question is how to choose the appropriate rule and the variables to form the indicator function. GW use one lag of the variable to be forecast to predict the loss differential in the future. In accordance to the results presented in Table 4, we use the following regression:

$$d_{t|t-1} = \gamma_0 + \sum_{i=1}^k \gamma_i RV_{t-i} + \sum_{j=1}^p \phi_j d_{t-j|t-1-j} + \varepsilon_t, \quad (8)$$

where $d_{t|t-1}$ represents the loss differential between each of the two pairs resulting from the application of the CPA tests (e.g., the GARCHroll6 and the iv1week). In contrast to GW, we use lags of the realized volatility as well as lags of the loss differential. This equation is basically a forecasting equation that, using variables known at t , can be used to predict if the loss differential will be positive or negative at $t + 1$.

The steps that we followed to produce the conditional combinations are:

1. In the estimation sample (2004 and 2005), we estimate regression (8), selecting the number of lags using AIC.²⁴ The data corresponding to years 2006 and 2007 are left out for the out-of-sample evaluation.

²¹Data for 2006 and 2007 are left out for out-of-sample evaluation.

²²The difference between these results and the DMW results presented in Table 2 are mainly due to the different samples employed. In Table 2 the sample is January 2, 2004 to December 31, 2007, whereas in Table 4 the sample is January 2, 2004 to December 30, 2005.

²³The results are robust to the use of different lags, and to the use of only a subset of the instruments.

²⁴The results are robust to the use of BIC instead.

2. We form a dummy variable as follows:

$$D_t = \begin{cases} 1 & \text{if } \hat{d}_{t|t-1} \geq 0, \\ 0 & \text{if } \hat{d}_{t|t-1} < 0. \end{cases}$$

3. Next, we use the dummy variable to estimate regression (6) lagged one period:

$$RV_t = \delta_0 + \delta_1 D_t + \delta_2 \hat{\sigma}_{t|t-1,i}^2 + \delta_3 D_t \hat{\sigma}_{t|t-1,i}^2 + \delta_4 \hat{\sigma}_{t|t-1,j}^2 + \delta_5 D_t \hat{\sigma}_{t|t-1,j}^2 + \varepsilon_t,$$

using data for 2004 and 2005, for the pair of forecasts i, j (e.g., GARCHroll6 and iv1w).

4. To calculate the combined forecast for the first day of 2006, we need to forecast $\hat{d}_{t+1|t}$. To do that we use the regression estimated in step 1 above but updating the right-hand variables so that the most recent be at t .

5. Then, we use $\hat{d}_{t+1|t}$ to define the value for the dummy D_{t+1} .

6. Finally, the conditionally combined forecast for the first day of 2006 is:

$$\hat{\sigma}_{t+1|t,c}^2 \text{ (Conditional)} = \hat{\delta}_0 + \hat{\delta}_1 D_{t+1} + \hat{\delta}_2 \hat{\sigma}_{t+1|t,i}^2 + \hat{\delta}_3 D_{t+1} \hat{\sigma}_{t+1|t,i}^2 + \hat{\delta}_4 \hat{\sigma}_{t+1|t,j}^2 + \hat{\delta}_5 D_{t+1} \hat{\sigma}_{t+1|t,j}^2.$$

7. To calculate the forecasts for the next days (2nd day of 2006, 3rd day of 2006, ..., last available day of 2007) we repeat steps 1 to 6 but adding one observation at a time (1st day of 2006, 2nd day of 2006, ..., one before the last available day of 2007), in a recursive manner. Always making sure we use observations up to time t to forecast at time $t + 1$.

It turns out that when the estimated loss differential is expected to be positive, this reflects periods of high volatility, whereas when the loss differential is expected to be negative, this reflects periods of low volatility. Figure 3 shows that if we separate the returns, and calculate the empirical distribution of the returns when the loss differential is predicted to be positive, and the empirical distribution of the returns corresponding to dates when the loss differential is predicted to be negative, indeed, positive expected loss differentials are associated with periods of high volatility.

When we use the indicator function to construct hybrid forecasts in the spirit of GW, the rule implies that the GARCHs should be used when this difference is predicted to be negative (i.e., when we predict that the loss function from the GARCH model will be smaller than the loss from the IV), and the IVs hard-data should be used when it is positive. This result is in line with what can be observed in Figure 4. The option implied can follow the realized

volatility in periods of high volatility, but it over-predicts during low volatility episodes, in particular with respect to what the GARCH does.

Table 5 presents the MSFE ratios, with respect to the MSFEs of the best individual method, the GARCHroll6, the best linear combination, GR1, the hybrid forecast, and the conditional combination proposed here (all pairwise), calculated over the same sample. Notice that in contrast to results presented in Table 3, the GR1 combination method reported in Table 5 only combines the corresponding pair of forecasts. The MSFE of the time-varying combinations are smaller than those of any other forecasting method considered in Table 5. In addition, DMW tests confirm that the conditional combinations have MSFEs that are statistically superior to the MSFE of the GARCHroll6, and in one of the two cases is even statistically superior than the MSFE of the best linear combination. The overall finding is that the conditional combinations are superior in terms of MSFE.

As a robustness check, we repeated the exercise but using the squared daily volatility as our proxy for the (latent) daily volatility. As expected given that this is a noisier measure than the one using intra-day data, the MSFE for all the forecasts and forecast combinations increased. The extra noise certainly makes it more difficult to tell apart the performance of different forecasts. However, in this case we also find that the MSFEs of the conditional combinations are smaller than those of the other methods considered. Table 6 shows the results in the same format than the one used in Table 5. The first thing to notice is that one of the pairs used for the combination changed with respect to Table 5. Now the first pair is formed by the GARCHroll3 and the iv1w, instead of the GARCHroll6 and the iv1w. This is a direct result of the CPA tests (not reported) and the fact that now the individual model with the smaller MSFE is the GARCHroll3. Nonetheless, the qualitative results are the same. The MSFE ratios of the conditional combinations are the smaller ratios and the statistical tests of predictive accuracy indicate that they have a very good performance. Furthermore, the fact that the pairs detected by the CPA tests always include a member from each type of forecasts indicates that the conditional combinations are taking into consideration the comparative advantages of each type.

5 Summary and Conclusions

The on-going debate regarding which is the most accurate model to forecast volatility of price returns of financial assets has led to a substantial amount of research. Many have compared ARCH-type models against option implied volatilities. Albeit the majority of the literature advocates the use of option implied volatilities as the most accurate alternative to forecast price returns volatilities, there is still no consensus in terms of finding one unique superior

model. This is mainly because the statistical evaluation of the forecasts has generally shown that the competing models either perform worse, or have equal statistical accuracy. More recently, following an extensive literature in macroeconomic forecasting, composite forecasts have been shown to be a good way to improve forecasting performance.

In the present paper, the aforementioned volatility forecast models are compared with each other in order to find the most accurate volatility forecasting method for the volatility of the daily spot returns of the Mexican peso – US dollar exchange rate. Intraday exchange rate data with 5 minute time intervals is used to construct the realized volatility used as our preferred measure of ex-post volatility. The results show that, when comparing single forecasts, there is almost no evidence to reject the null of equal predictive ability between the ARCH-type and (hard data) implied volatilities. Among linear combinations, it is found that, in general, combining results in smaller MSFE with respect to individual forecasts.

However, the best forecast overall is obtained when one ARCH-type forecast and one forecast from the option implieds are combined using time-varying weights. By using past volatility as well as past forecasting performance to predict the regime, a conditional combination that assigns relatively more weight to the option implied during high volatility regimes, and relatively less weight to it during low volatility regimes, has a MSFE that is about 50% lower than the MSFE of the best individual model, and about 25% lower than the MSFE of the best (unconditional) combination, a clear improvement in forecast accuracy.

Future research should look at conditional combinations involving multiple forecasts, and should also explore the conditional combination of forecasts for other variables such as inflation, interest rates, and volatility of other assets.

References

- [1] Aiolfi, M. and Timmermann A. (2006). Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics*, 135, 31-53.
- [2] Andersen, T.G., Bollerslev, T., Christoffersen, P. F., and Diebold, F.X. (2006). Volatility and Correlation Forecasting. *Handbook of Economic Forecasting*. Edited by G. Elliott., C.W.J. Granger., and A. Timmermann. Amsterdam: North Holland.
- [3] Andersen, T.G., and Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 39, 885-905.
- [4] Akgiray, V. (1989). Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business*, 62, 55-80.
- [5] Bates, J.M. and Granger, C.W.J. (1969). The Combination of Forecasts. *Operations Research Quarterly*, 20, 451-468.
- [6] Becker, R. and Clemens, A.E. (2008). Are Combination Forecasts of S&P 500 Volatility Statistically Superior? *International Journal of Forecasting*, 24, 122-133.
- [7] Benavides, G. (2003). Price Volatility Forecasts for Agricultural Commodities: An Application of Historical Volatility Models, Option Implieds and Composite Approaches for Futures Prices of Corn and Wheat. *Working Paper, Lancaster University*. url: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=611062
- [8] Benavides, G. (2006). Volatility Forecasts for the Mexican Peso - US Dollar Exchange Rate: An Empirical Analysis of GARCH, Option Implied and Composite Forecasts Models. *Working Paper 2006-04, Banco de México*. url: <http://www.banxico.org.mx/documents/%7B7F1FAAE1-FEFC-EB00-B0C8-55420567F74F%7D.pdf>
- [9] Black, F. and Scholes, M. S. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81, May-June, 637-654.
- [10] Blair, B. J., Poon, S. and Taylor, S. J. (2001). Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns. *Journal of Econometrics*, 105, 5-26.
- [11] Bollerslev, T. P. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- [12] Bollerslev, T. P., Engle, R. F. and Nelson D. B. (1994). ARCH Models. *Handbook of Econometrics*. Vol 4. Edited by R. F. Engle and D. L. McFadden. Amsterdam: Elsevier.
- [13] Canina, L and Figlewski, S. (1993) The Informational Content of Implied Volatility. *Review of Financial Studies*, 6, 659-681.

- [14] Clemen, R. T. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, 5, 559-583.
- [15] Cumby, R., Figlewski, S. and Hasbrouck, J. (1993). Forecasting Volatilities and Correlations with EGARCH Models. *Journal of Derivatives*, 1, 51-63.
- [16] Day, T. E. and Lewis, C. M. (1992). Stock Market Volatility and the Information Content of Stock Index Options. *Journal of Econometrics*, 52, 267-287.
- [17] Deutsch, M., Granger, C.W.J. and Terasvirta, T. (1994). The Combination of Forecasts Using Changing Weights. *International Journal of Forecasting*, 10, 47-57.
- [18] Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253-263.
- [19] Ederington, L. and Guan, W. (2002). Is Implied Volatility an Informationally Efficient and Effective Predictor of Future Volatility? *Journal of Risk*, 4, 3.
- [20] Elliott, G. and Timmermann, A. (2005). Optimal Forecast Combination Weights under Regime Switching. *International Economic Review*, 46, 1081-1102.
- [21] Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica*, 50, 987-1008.
- [22] Engle, R.F and Ng, V.K. (1993). Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, 48, 1749-1778.
- [23] Fleming, J. (1998). The Quality of Market Volatility Forecasts Implied by S&P 100 Index Option Prices. *Journal of Empirical Finance*, 5, 317-345.
- [24] Fleming, J., Ostdiek, B. and Whaley, R.E. (1995). Predicting Stock Market Volatility : A New Measure. *Journal of Futures Markets*, 15, 265-302.
- [25] Garman, M. B and Klass, M. J. (1980). On the Estimation of Security Price Volatilities from Historical Data. *Journal of Business*, 53, 67-78.
- [26] Garman, M.B. and Kohlhagen, S.W. (1983). Foreign Currency Option Values. *Journal of International Money and Finance*, 2, 231-237.
- [27] Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74, 1545-1578.
- [28] Granger, C.W.J. and Ramanathan, R. (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting*, 3, 197-204.
- [29] Granger, C.W.J. (2003). Long Memory Process-An Economist's Viewpoint. *Advances in Statistics, Combinations, and Related Areas*, edited by C. Gulats, et al., 100-111. World Scientific Publishers.

- [30] Guidolin, M. and Timmermann, A. (2006). Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach. *Federal Reserve Bank of St. Louis, Working Paper Series*. url: <http://research.stlouisfed.org/wp/2005/2005-059.pdf>
- [31] Guo, D. (1996). The Predictive Power of Implied Stochastic Variance from Currency Options. *Journal of Futures Markets*, 16, 915-942.
- [32] Hansen, P.R. and Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20, 873-889.
- [33] Jorion, P. (1995). Predicting Volatility in the Foreign Exchange Market. *The Journal of Finance*, 50, 507-528.
- [34] Lamoureux, C. G. and Lastrapes, W. D. (1993). Forecasting Stock Return Variance: Toward an Understanding of Stochastic Implied Volatilities. *The Review of Financial Studies*., 6, 293-326.
- [35] Manfredo, M. Leuthold, R. M. and Irwin, S. H. (2001). Forecasting Cash Price Volatility of Fed Cattle, Feeder Cattle and Corn: Time Series, Implied Volatility and Composite Approaches. *Journal of Agricultural and Applied Economics*, 33, 523-538.
- [36] Newey, W. and West, K. (1987). A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55, 703-708.
- [37] Parkinson, M. (1980) The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Financial Economics*, 18, 199-228.
- [38] Patton, A. J. (2006) Volatility Forecast Comparison using Imperfect Volatility Proxies. *Quantitative Finance Research Centre, University of Technology Sydney*, Research Paper 175.
- [39] Patton, A. J. and Sheppard, K. (2007). Evaluating Volatility and Correlation Forecasts. *Handbook of Financial Time Series*. Edited y T.G. Andersen, R.A., Davis, J.-P. Kreiss and T. Mikosch. Springer Verlag (forthcoming).
- [40] Pesaran, M.H. and Timmermann, A. (2007). Selection of Estimation Window in the Presence of Breaks. *Journal of Econometrics*, 137, 134-161.
- [41] Poon, S-H. and Granger, C.W.J. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, 41, 478-539.
- [42] Pong, S., Shackleton, M., Taylor, S. and Xu, X. (2004). Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models. *Journal of Banking and Finance*.
- [43] Szakmary, A, Ors, E., Kim, J. K. and Davidson III, W. D. (2003). The Predictive Power of Implied Volatility: Evidence from 35 Futures Markets. *Journal of Banking and Finance*, 27, 2151-2175.

- [44] Taylor, S. J. (1986). *Modeling Financial Time Series*. Wiley.
- [45] Taylor, S. J. (2005). *Asset Price Dynamics, Volatility, and Prediction*. Princeton University Press.
- [46] Timmermann, A. (2006). Forecast Combinations. *Handbook of Economic Forecasting*. Edited by G. Elliott., C.W.J. Granger., and A. Timmermann. Amsterdam: North Holland.
- [47] West, K.D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica*, 64, 1067-1084.
- [48] Xu, X. and Taylor, S. J. (1995). Conditional Volatility and the Informational Efficiency of the PHLX Currency Options Market. *Journal of Banking and Finance*, 19, 803-821.

**Table 1. Individual Forecast Evaluation
(January 2, 2004 - December 31, 2007)**

Forecast	Mean Error	MSFE Ratio
GARCHroll6	-2.00E-06	1.00
GARCHroll3	1.24E-07	1.06
iv1week	-7.05E-06	1.33
iv3weeks	-6.79E-06	1.19
ivmb	-5.37E-05	155.72

This table reports the Mean Error and Mean Square Forecast Error (MSFE) Ratio of the volatility forecasting models for the volatility of the daily spot returns for the Mexican peso - USD exchange rate. Results are for out-of-sample forecasts. The realized volatility used is the annualized ex-post intraday spot return volatility.

Abbreviations of the models are as follows: GARCHroll6 and GARCHroll3 denote the GARCH(1,1) model estimated using rolling windows of six and three years respectively and a t-distribution. iv1week and iv3week are the hard data option implied volatility with time-to-maturity of one and three weeks, respectively. ivmb is the model-based option implied volatility for options with one month to maturity estimated with the Garman and Kohlhagen (1983) model.

**Table 2. Diebold-Mariano-West Tests
(January 2, 2004 - December 31,2007)**

	P-Values \ T-Stats				
	GARCHroll6	GARCHroll3	iv1week	iv3weeks	ivmb
GARCHroll6	–	-1.57	-1.99	-1.84	-3.82
GARCHroll3	0.1164	–	-1.61	-1.11	-3.82
iv1week	0.0474	0.1079	–	0.89	-3.81
iv3weeks	0.0657	0.2679	0.3711	–	-3.81
ivmb	0.0001	0.0001	0.0001	0.0001	–

This table reports the results of pair-wise Diebold-Mariano-West tests of equal predictive ability using MSFE as loss function. Above the diagonal we present the Diebold-Mariano statistic, below the diagonal we present the corresponding p-values. In every case, the difference is calculated as the row forecast minus the column forecast.

Abbreviations as in Table 1.

**Table 3. Evaluation of Forecast Combinations
(January 2, 2006 - December 31, 2007)**

Combination	Mean Error	MSFE Ratio	DMW Test
Equal weights	-1.23E-05	6.48	0.0028
GR1	-1.95E-06	0.66	0.0266
GR2	-2.30E-06	0.68	0.0337
GR3	-4.54E-06	0.79	0.1664
GARCHroll6	-2.29E-06	1.00	x

This table reports the Mean Error, Mean Square Forecast Error (MSFE) Ratio and p-values of the pair-wise Diebold-Mariano-West tests of the volatility forecasting models for the daily spot returns for the Mexican peso - USD exchange rate. The realized volatility used is the annualized ex-post intraday spot return volatility.

All the combined forecast include: Implied Volatility model based, Implied Volatility one week to maturity, Implied Volatility three weeks to maturity, GARCH rolling with three year window and a t-distribution, and GARCH rolling with six year window with a t-distribution.

Abbreviations of the models are as follows: Equal Weights refers to the simple average of the forecasts. GR = Granger-Ramanathan-type model. The first GR combination lacks a constant and the weights add to one, the second lacks a constant, and the third has a constant and the weights are not restricted. GARCHroll6 stands for the GARCH(1,1) model estimated using a rolling window of six years.

Table 4. Giacomini-White Tests
(January 2, 2004 - December 30, 2005)

P-Values. Instruments are: Constant \ Lags of realized volatility, lags of \widehat{d}					
	GARCHroll6	GARCHroll3	iv1week	iv3weeks	ivmb
GARCHroll6	–	0.1380	0.0919	0.0051	0.0098
GARCHroll3	0.1915	–	0.2600	0.0075	0.0099
iv1week	0.1728	0.5119	–	0.0000	0.0100
iv3weeks	0.0013	0.0252	0.0000	–	0.0101
ivmb	0.0000	0.0000	0.0000	0.0000	–

This table reports p-values of pair-wise Giacomini and White (2006) tests of conditional predictive ability. Below the diagonal the instrument used is just a constant. Above the diagonal the instruments are five lags of the realized volatility and five lags of the loss differential. The realized volatility used is the annualized ex-post intraday spot return volatility.

Abbreviations as in Table 1.

**Table 5. Evaluation of Time-Varying Combinations
(January 2, 2006 - December 31, 2007)**

Forecasts	Combination	Mean Error	MSFE Ratio	DMW^{a/}	DMW^{b/}
GARCHroll6	Hybrid	-4.18E-06	1.31	0.1249	0.0502
and	Conditional	-2.31E-06	0.54	0.0554	0.1796
iv1w	GR1	-2.02E-06	0.69	–	X
GARCHroll3	Hybrid	-9.64E-07	1.10	0.1009	0.0254
and	Conditional	-1.63E-06	0.52	0.0408	0.0705
iv3w	GR1	-1.05E-06	0.74	–	X
GARCHroll6		-2.29E-06	1.00	X	–

a/ The benchmark is GARCHroll6

b/ The benchmark is GR1

This table reports the Mean Error, Mean Square Forecast Error (MSFE) Ratio and p-values of pair-wise Diebold-Mariano-West tests.

Abbreviations of the models are as follows: GARCHroll6, GARCHroll3, iv1week, and iv3week as in Table 1. GR1 as in Table 3.

**Table 6. Evaluation of Time-Varying Combinations
using Squared Daily Returns as the Proxy for the Daily Volatility
(January 2, 2006 - December 31, 2007)**

Forecasts	Combination	Mean Error	MSFE Ratio	DMW ^a /	DMW ^b /
GARCHroll3	Hybrid	-5.11E-06	0.95	0.5562	0.0288
and	Conditional	-8.61E-07	0.80	0.1014	0.4127
iv1w	GR1	-8.01E-07	0.85	–	X
GARCHroll3	Hybrid	-1.42E-07	1.01	0.2134	0.0984
and	Conditional	9.33E-07	0.83	0.0619	0.1200
iv3w	GR1	9.96E-07	0.92	–	X
GARCHroll3		8.72E-08	1.00	X	–

Notes and abbreviations as in Table 5, except that here the benchmark from the individual models is the GARCHroll3.

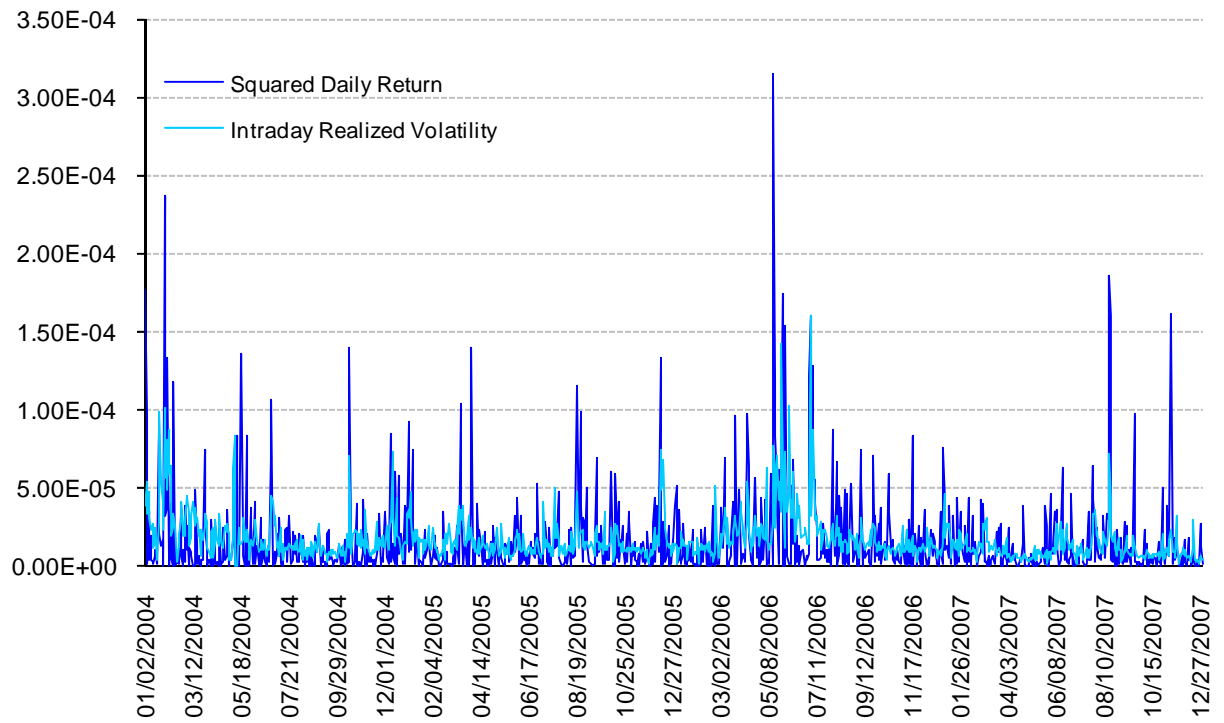


Figure 1: Exchange rate MXN—USD square daily return and intraday realized volatility.

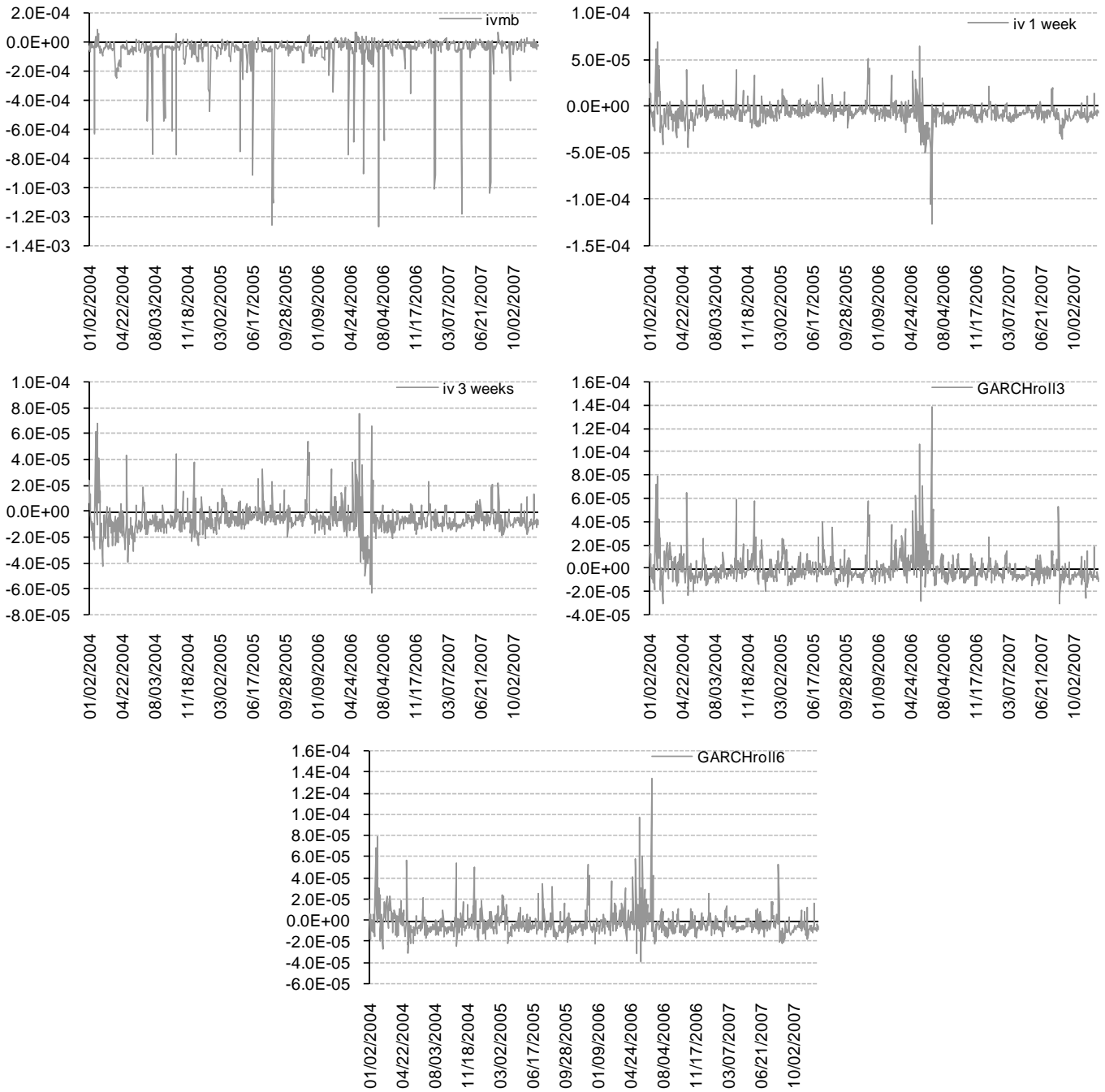


Figure 2: Forecast errors for the individual models.

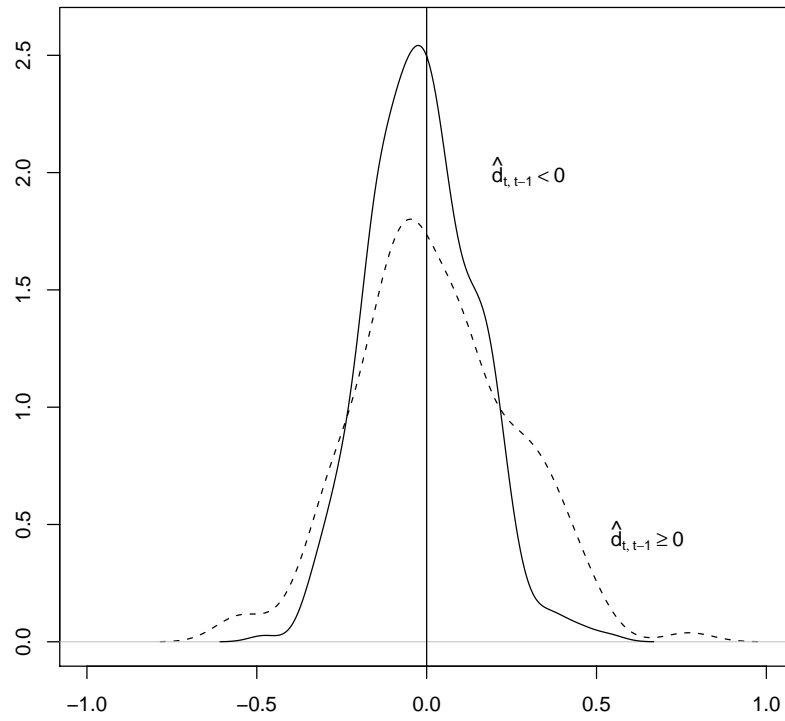


Figure 3. Indicator function times daily returns
January 2, 2006 - December 31, 2007.

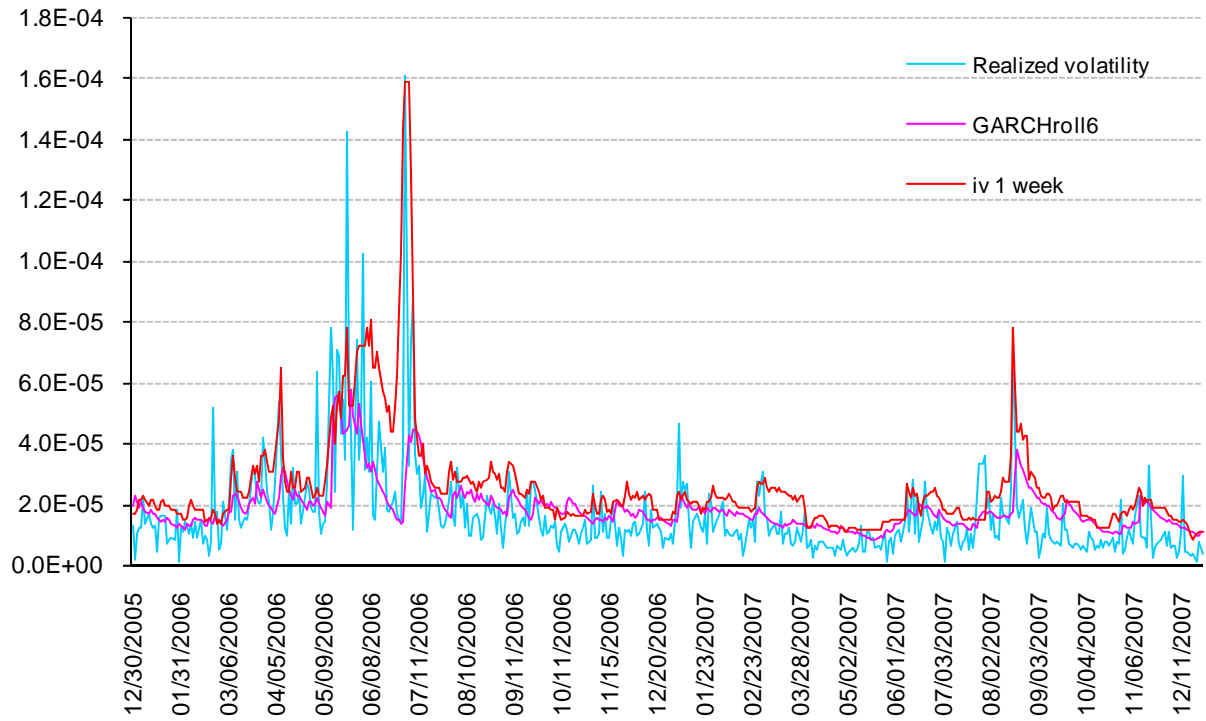


Figure 4. Realized volatility and forecasts from GARCH rolling method and hard data option implied 1 week.