SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-34

The Forecast Performance of Competing Implied Volatility Measures: The Case of Individual Stocks

Leonidas Tsiaras

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

The Forecast Performance of Competing Implied Volatility Measures: The Case of Individual Stocks

Leonidas Tsiaras*

Department of Business Studies, ASB, Aarhus University
and CREATES, Aarhus University

February, 2010

Abstract

This study examines the information content of alternative implied volatility measures for the 30 components of the Dow Jones Industrial Average Index from 1996 until 2007. Along with the popular Black-Scholes and "model-free" implied volatility expectations, the recently proposed corridor implied volatility (CIV) measures are explored. For all pair-wise comparisons, it is found that a CIV measure that is closely related to the model-free implied volatility, nearly always delivers the most accurate forecasts for the majority of the firms. This finding remains consistent for different forecast horizons, volatility definitions, loss functions and forecast evaluation settings.

JEL Classifications: C22, C53, G13, G14

Keywords: Model-Free Implied Volatility; Corridor Implied Volatility; Volatility Forecasting

---

*I would like to thank Torben G. Andersen, Esben Høg, Maria Teresa Gonzalez Perez, and seminar participants at the Center for Research in Econometric Analysis of Time Series (CREATES) and the Nordic Finance Network 2009 Research Workshop for valuable comments and suggestions. A first draft of this paper was completed during my research visit at Kellogg School of Management, Northwestern University, whose hospitality is greatly appreciated. Support from CREATES, funded by the Danish National Research Foundation, is gratefully acknowledged.
Corresponding address: Fuglesangs Alle 4, DK-8210, Aarhus V., Denmark, *e-mail*: leot@asb.dk

1 Introduction

Market efficiency suggests that volatility expectations embedded in option prices should also contain information about future volatility levels. More importantly, this information may be absent from the historical record of asset returns, because option-implied volatilities incorporate both relevant historical information as well as traders' expectations regarding future horizons. Another attractive property of implied volatilities is that they have forward-looking features by construction. Given the above, it is natural to expect that they may be useful in enhancing volatility forecast models and, indeed, empirical analysis has confirmed these theoretical predictions for a variety of markets and asset classes[1].

However, volatility information is not necessarily dispersed uniformly across the entire cross-section of option prices. Systematic differences can arise for a variety of reasons of either theoretical nature (for instance biases caused by the pricing of volatility risk) or of pragmatic substance (such as the relative illiquidity of particular contracts). For this reason, various approaches of extracting risk-neutral volatility expectations can be proposed. Essentially, alternative methodologies are motivated either by different assumptions about the asset price dynamics or by criteria concerning the relative informational efficiency of options trading at different strikes. The existence of several candidate option-implied measures has obvious implications when building forecast models.

The issue of which implied volatility to use arose naturally in early studies, since the inversion of the Black-Scholes formula did not produce a flat volatility surface. Commencing with Latane and Rendleman (1976) it was recognized that some options may be more informative than others because their observed prices are less sensitive to measurement errors or model misspecification distortions. These authors, together with Trippi (1977) and Chiras and Manaster (1978), were the first to suggest that some sort of averaging between different implied volatilities could improve the accuracy of the forecasts. Perhaps due to their ad-hoc nature, such weighting schemes have enjoyed little success as they were found to perform worst than individual implieds (Beckers, 1981; Gemmill, 1986; Fung et al, 1990). When selecting between individual implieds, on the other hand, the general consensus is to prefer at-the-money contracts since they usually trade more heavily. Along these lines, it is custom to select a single option to extract volatility information and discard all other observations.

[1]See Poon and Granger (2003) or Taylor(2005) for a survey of the related literature.

An important development towards the understanding of volatility measures implied by option prices was derived by Britten-Jones and Neuberger (2000), Carr and Madan (1998) and Demeterfi, Derman, Kamal and Zou (1999). Building on Ross (1976) and Breeden and Litzenberger (1978), these studies show that ex-ante risk-neutral expectations of future realized variance can be obtained through the fair value of a variance swap rate. Under mild assumptions, the value of this position can be well approximated using the prices of observed option contracts, resulting in a *model-free* variance estimate that does not require the existence of a particular option pricing model and remains approximately valid under a wide class of asset price dynamics, including jump-diffusive semi-martingale processes. Using these results, it becomes possible to extract a single implied volatility estimate that is consistent with the entire cross-section of observed option prices.

Despite their theoretical novelty, model-free estimates also suffer from two limitations when it comes to forecasting future volatility. The first one is that, at least for stock indices, volatility expectations will be magnified due to the presence of volatility risk-premia. While the existence of the latter per se would have mild implications if they were constant across time, their apparent time-variation (Bollerslev, Gibson and Zhou, 2006; Todorov, 2007) contaminates volatility expectations with fluctuations that are not relevant to the object of interest, but rather correspond to changes in the market pricing of volatility risk. Secondly, they do not take into account that away-from-the-money options are traded far less frequently, which implies that their relative informational efficiency is expected to be lower.

A handful of studies have recently compared the information content of at-the-money Black-Scholes and model-free implied volatility estimates (BSIV and MFIV henceforth), inspired by the appealing theoretical properties of the latter. Jiang and Tian (2005) examine the S&P 500 index (from 1988 to 1994) and their in-sample results favor volatility estimates extracted using the model-free approach. These findings are in sharp contrast with those of Andersen and Bondarenko (2007) who use S&P 500 futures data (1990-2006) and report that MFIV consistently underperforms BSIV in both in-sample and out-of-sample exercises. Further evidence, revealing some of the deficiencies of MFIV measure in the context of volatility forecasting, are put forth by Taylor et al (2007).

Selecting the most informative volatility forecast given panel of option prices remains largely an open question. It is unlikely that a unique answer will emerge, as the optimal choice may depend on the asset class or the characteristics of the underlying market. To this end, a coherent way of systematizing the range of option prices that may contain useful volatility information is proposed by Andersen and

Bondarenko (2007) who utilize the concept of *corridor implied volatility* (CIV), introduced earlier in Carr and Madan (1998). Unlike variance swaps, which pay the accumulated return variation irrespective of the price path of the underlying asset, corridor variance contracts pay the return variation that is accumulated when the asset price lies between two, pre-specified, reference asset levels or barriers. A key difference between the calculation of model-free and corridor implied volatility is that the first requires the prices of options with any attainable strike price, while for the latter only the contracts with strikes that lie inside the barriers are needed. In this way, CIV measures can potentially serve as a mechanism to alleviate volatility risk-premia fluctuations by extracting volatility expectations using a range of option prices that is less susceptible to such distortions. From a more practical perspective, a CIV approach recognizes that different parts of the risk-neutral density (whose entire support needs to be assumed so as to compute the MFIV) are estimated with different degrees of reliability. This is especially true for the part of the risk-neutral density that falls outside the range of traded option prices. The aforementioned arguments are indeed confirmed in Andersen and Bondarenko (2007), who find that certain CIV measures perform better than either the BSIV or the MFIV in terms of forecasting index volatility.

The contribution of this paper is twofold. Firstly, the forecast performance of corridor implied volatility measures is scrutinized from a cross-sectional perspective. By definition, any CIV measure will have the handicap that the there is a mismatch between the estimated quantity (risk-neutral expectation of corridor integrated variance) and the target quantity (integrated variance). Whether the advantages postulated by a CIV approach outweigh this major drawback is largely an empirical question. In this respect, the analysis of Andersen and Bondarenko (2007) is limited to a single stock index. Given that several corridor definitions are explored in their paper, it is difficult to draw a firm conclusion about the overall usefulness of these alternative volatility measures. For this purpose, this study investigates a pool of thirty individual stocks, facilitating more reliable inferences.

In addition, this paper supplements the literature that uses option prices to forecast the volatility of individual stocks. Unlike the case of equity index or currency options, studies that focus on the firm level are scarce, with Lamoureux and Lastrapes (1993) and Taylor et al (2007) being two notable exceptions. This work relates closely to the latter paper where the authors examine 149 individual firms for a four year period (1996-1999) and report that BSIV appears to be more informative than its model-free counterpart. Besides the use of additional implied volatility measures, significant improvements are also made in the empirical part of the study.

Specifically, a more recent and substantially longer dataset is explored (1996-2007), while forecasts are evaluated using realized volatility estimates constructed by accumulating intraday returns, facilitating more robust comparisons. In-sample, as well as out-of-sample, forecast evaluations are made for two distinct horizons, using different loss functions and formal tests of predictive accuracy.

The main result that emerges from the empirical forecast evaluations is that systematic differences exist in the performance of several implied volatility measures. More importantly, one particular CIV measure, which corresponds to a wide barrier width, produces more accurate forecasts for the majority of the firms irrespective of the competing alternative, including the BSIV and MFIV measures. It is noteworthy that, in contrast with Andersen and Bondarenko (2007), it is found that wide CIV measures perform better than their narrow corridor counterparts. This difference could be potentially rationalized by the dissimilar magnitudes of volatility risk-premia embedded in away-from-the-money options for stocks and stock indices, documented in Driessen et al (2009) among others. Comparisons between BSIV and MFIV forecasts produce mixed results, so that it becomes difficult to favor one of the two measures.

The paper is organized as follows. Section 2 briefly describes the dataset. Section 3 discusses the definition and estimation of the various implied volatility measures. Sample construction and forecast evaluation issues are treated in Section 4. Section 5 contains the empirical results of the study, while the main conclusions are set out in Section 6.

2 Data Description

Options data were obtained from the Ivy database of OptionMetrics, which provides historical option prices for equity and index options based on closing quotes at the Chicago Board of Options Exchange. For reasons of computational convenience, the analysis is restricted to the 30 components of the Dow-Jones Industrial Average (as of April 2004), which also offers some comfort against low liquidity considerations. The dataset spans a period of approximately 11 years, from January 1996 to May 2007 inclusive.

Implied volatilities, provided by OptionMetrics, are based on mid-quote option prices. When the options are European, Ivy uses the Black-Scholes formula for dividend paying assets, while for American options a binomial tree approach, that takes into account the early exercise premium, is adopted. As an attempt to filter out uninformative options data several screening criteria were applied[2].

The same source also provides the remaining inputs that are required for the computation of option prices. Continuously compounded interest rates were obtained by linearly interpolating between the two adjacent zero-coupon rates from the Yield Curve file. The Ivy DB also supplies detailed information about dividend distributions and splits for each security, while the Security Price file contains the respective spot prices. The time zero price of forward contracts for delivery at time T were computed by subtracting the present value of all dividends from time zero up to time T from the observed spot price, and subsequently multiplying by the risk-free rate that corresponds to the relevant horizon. Interest rates were assumed non-stochastic so that forward and futures prices are equal.

Intra-day return data were obtained from the Trade and Quote (TAQ) database which contains trades and quotes for all securities listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), as well as the Nasdaq National Market System (NMS). Mid-quote prices were extracted every 30-minutes, from December 1995 to July 2007 inclusive. To limit any recoding errors, screening criteria similar to those of Barndorff-Nielsen et al (2008) were applied. Finally, for days that included a dividend payment, the value of the corresponding intraday return was set to zero.

3 Construction of Implied Volatility Measures

3.1 Volatility Definitions

Let S_t denote the spot price of an asset at time t and F_t the corresponding futures price that follows a continuous-time process,

$$\frac{dF_t}{F_t} = \sigma_t dW_t \quad , \tag{1}$$

where W_t and σ_t are a Wiener and a volatility process, respectively. In the current setting the risk-free rate and the dividend yield are assumed to be deterministic[3]. Note that no restrictions are imposed for the volatility dynamics, other than that σ_t

[2]See Appendix A for further details.

[3]Note that these restrictions are not generally required (see for instance Jiang and Tian, 2005). However, in the empirical part of the study options on spot prices are going to be used so it is necessary that these conditions hold.

follows a strictly positive stochastic process. Letting $t_0 \leq t \leq T$ and fixing $t_0 = 0$, the integrated variance from time t_0 to T is then

$$IVAR(t_0, T) = \int_0^T \sigma_t^2 dt \tag{2}$$

In what follows, also assume the absence of arbitrage opportunities and the existence of a unique risk-neutral density that can be used to price European put and call options expiring at some future time T. Following Carr and Madan (1998), Demeterfi, Derman, Kamal, and Zou (1999) and Britten-Jones and Neuberger (2000), ex-ante risk-neutral expectations of future integrated variance can be obtained by computing the fair value of a portfolio that consists of options traded at a continuum of strike prices. In particular,

$$E^Q[IVAR(t_0, T)] = E^Q[\int_0^T \sigma_t^2 dt] = 2e^{rT} \int_0^\infty \frac{M_{t_0,T}(K)}{K^2} dK \quad , \tag{3}$$

where $M_{t_0,T}(K)$ equals the value of a put option maturing at time T, if the strike price K is below the current futures price, or a call option maturing at time T otherwise. Because this, option-implied, expectation does not make any assumptions about the underlying option pricing model, it is commonly referred to as *model-free implied volatility*. Similarly, by defining two positive barriers, B_1 and B_2, and the following indicator function

$$I_t(B_1, B_2) = I_t[B_1 \leq F_t \leq B_2] \tag{4}$$

one can also introduce a corridor integrated variance measure as

$$CIVAR(t_0, T) = \int_0^T \sigma_t^2 I_t(B_1, B_2) dt \tag{5}$$

The difference compared to the pure integrated variance definition is that now the return variation is only accumulated when the futures price of the underlying at time t is between the two pre-specified barrier levels. As demonstrated in Carr and Madan (1998) and Andersen and Bondarenko (2007), the corresponding risk-neutral expectation of this measure can be computed, using similar arguments as in the previous case, by estimating the value of a portfolio of options with strikes ranging from B_1 to B_2. Specifically, the risk-neutral expectation of future *corridor integrated variance* at time $t_0 = 0$ can be computed as

$$E^Q[CIVAR(t_0, T)] = E^Q[\int_0^T \sigma_t^2 I_t(B_1, B_2) dt] = 2e^{rT} \int_{B_1}^{B_2} \frac{M_{t_0,T}(K)}{K^2} dK \tag{6}$$

It is clear that when the barriers are set to $B_1 = 0$, $B_2 = +\infty$ the definitions of corridor integrated variance and integrated variance coincide. Likewise, one can also set either $B_1 = 0$ or $B_2 = +\infty$ and obtain barrier integrated variance measures as in Carr and Madan (1998) and Andersen and Bondarenko (2007). Two special cases of the barrier variance definitions that could potentially be of interest are the up-variance (UVAR) and down-variance (DVAR) measures, where return variation is accumulated only when the futures price F_t is strictly above (UVAR) or below (DVAR) the futures price at time zero.

Although the aforementioned theory is developed for the return variance, analogous measures in terms of volatilities (i.e. standard deviations) can be approximated by taking square roots in the relevant expressions. Throughout the paper the term CIV will refer to either *corridor implied variance* or *corridor implied volatility*, depending on the context. Similarly, the term MFIV will refer to either *model-free implied variance* or *model-free implied volatility*, while UVOL (DVOL) will denote either *up-variance*(*down-variance*) or *up-volatility* (*down-volatility*) expectations.

3.2 Practical Implementation

3.2.1 Risk-Neutral Density Estimation

The computation of model-free implied volatility requires the market prices of options trading at every strike for which the risk-neutral measure assigns a positive probability. Similarly, corridor implied volatility measures require option prices trading at all possible strikes within the corridor width. In practice, of course, options only trade at discrete strikes whose range is not sufficiently wide.

Given a set of observed option prices with maturity T, the problem of obtaining option prices for arbitrary strikes is equivalent to estimating the risk-neutral density (RND) of the underlying for a future date T. Numerous alternatives exist to perform the latter task[4], including parametric (mixture or flexible densities, expansion methods) or non-parametric (curve fitting, kernel regression, entropy methods) approaches and others that share some features of both (positive convolution approximation). As noted in Taylor (2005), when options data for a wide spectrum of strikes are available, as it is the usually the case for major stock indices, several methods will generally provide satisfactory results. On the other hand, options written on individual stocks are far less liquid, which renders some of the existing techniques inapplicable. As in this study RNDs will have to be computed on a daily

[4]For an overview of the related literature see Jackwerth (1999) or Taylor (2005).

basis, a highly desirable feature of the estimation technique is that it is not data intensive.

In this setting the implied volatility function approach (Shimko, 1993) appears as the most reasonable candidate, since the RND can be extracted with as few as three option prices. The essence behind this approach is to estimate a quadratic function that fits the observed volatility "smile", which in turn provides a continuous price function from which the RND can be inferred by numerical differentiation. Building on Malz (1997) and Bliss and Panigirtzoglou (2002), Taylor et al (2007) propose a variation of this approach which this paper also adopts. In particular, a quadratic implied volatility function is fitted at the volatility/delta, rather than the volatility/strike, space by minimizing the sum of weighted squared differences between observed and fitted implied volatilities. The weighting scheme uses a function of the option's delta (that can be used as a proxy for the moneyness of the contract) which imposes a smaller penalty for errors that correspond to away-from-the money options. An additional advantage of this method is that, through the non-linear mapping from the volatility/delta to the volatility/strike space, the volatility function becomes flat at extreme strike providing, thus, sensible bounds for implied volatilities.

3.2.2 Corridor Selection

In total, six different corridor-related implied volatility measures are considered. Following Andersen and Bondarenko (2007), barrier levels are determined using the inverse cumulative distribution function of the reference risk-neutral density, F_Q. Specifically, letting

$$B_1 = F_Q^{-1}(p), B_2 = F_Q^{-1}(1-p) \tag{7}$$

and using the general formula in equation (6) the first four corridor measures, CIV1-CIV4, are obtained by setting $p = 0.45, 0.35, 0.25, 0.10$ respectively. For the other two measures, DVOL and UVOL, the barriers are set to $B_1 = F_Q^{-1}(0)$, $B_2 = F_Q^{-1}(0.5)$ and $B_1 = F_Q^{-1}(0.5)$, $B_2 = F_Q^{-1}(1)$, respectively.

3.2.3 Further Notes

As mentioned previously, MFIV and at-the-money BSIV estimates are also included in the analysis. For the first the case, one can use equation (6) with $B_1 = F_Q^{-1}(0)$, and $B_2 = F_Q^{-1}(1)$. Regarding the case of BSIV, its value is obtained from the daily

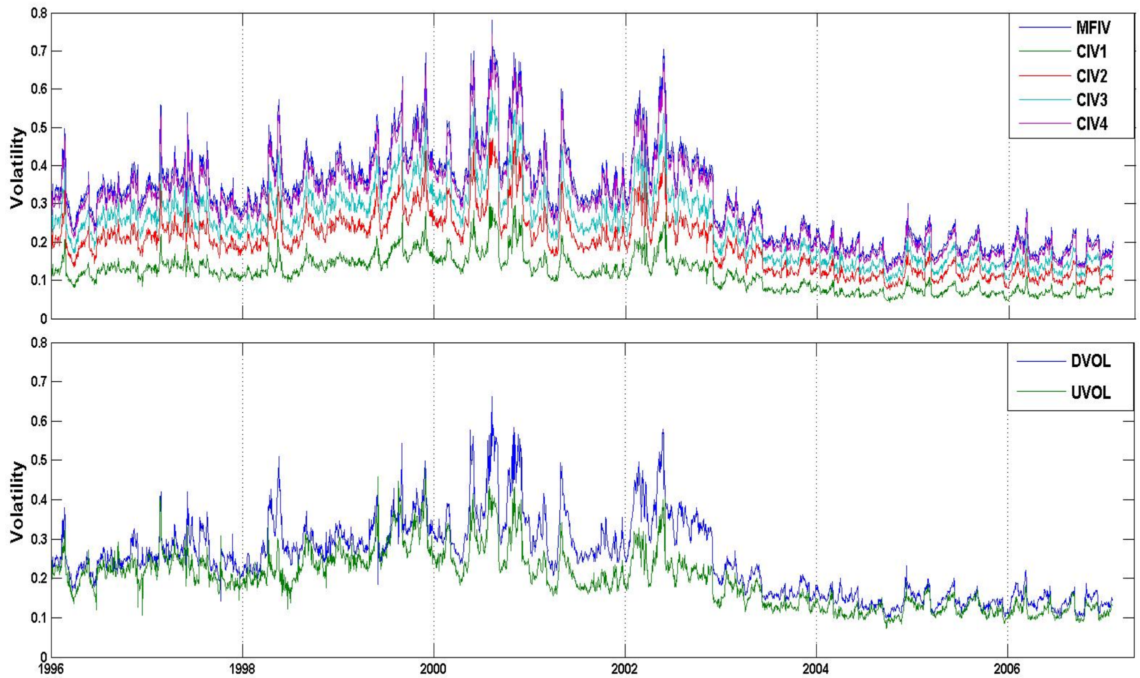

Figure 1: Time-series plots of different implied volatility measures for the IBM stock, from the beginning of 1996 until the middle of 2007. All volatility expectations are annualized and correspond to a 30-day horizon. Corridor Implied Volatility measures correspond to increasingly wider corridors as we move from CIV1 to CIV4. DVOL and UVOL refer to Up-Volatility and Down-Volatility estimates respectively, while MFIV represents the model-free implied volatility.

volatility smile using the contract whose strike price is equal to the forward price of the stock.

In order to approximate the integral in equation (6), firstly, the region for which the risk-neutral density assigns positive probabilities is determined and, subsequently, 2000 option prices having equally spaced strikes are computed.

For each day, and for each stock, two risk-neutral densities with different maturities are estimated[5]. For the monthly forecast horizon case, implied volatility expectations are matched with the forecast horizon using linear interpolation as in Carr and Wu (2009)[6]. Concerning the daily horizon case, volatility expectations

[5] As explained in Appendix A, these usually refer to the two closest maturities (after ignoring options that expire in eight days or less).

[6] After trying different interpolation methods, these authors report that the results they obtained

are computed using the following steps. Firstly, linear interpolation is used so as to compute fixed-horizon, 30 day, volatility measures. Subsequently, these are transformed into daily estimates by appropriate scaling. An alternative method would be to use only the closest to maturity options and then compute daily expectations by appropriate scaling. The reason for preferring the fixed-horizon approach was to abstract from potential liquidity effects that could emerge as the short maturity contracts were getting closer to their expiry date.

Since some of these aforementioned estimation methods involved some subjective choices, the results of the empirical section were repeated for the case where the at-the-money implied volatility was estimated using the spot, instead of forward, price as well as for the case where only the closest to maturity options were used to derive daily volatility expectations. The empirical findings, however, remained intact[7].

Figure 1 depicts the time-series behavior of various implied volatility estimates. While all measures clearly covary, their dynamics also exhibit some differences, as some appear more stable than others. Essentially, the dynamics of each series is reflecting changes in the implied volatility surface. Down-Volatility (DVOL) estimates, for instance, exceed in level their natural counterparts (UVOL) for prolonged periods of time, revealing the existence of negative skewness in risk-neutral density. Noteworthy, the CIV4 measure, that has a large corridor width, is strongly correlated with the model-free implied volatility (MFIV) measure.

4 Evaluating Volatility Forecasts

In order to assess the relative performance of competing volatility forecasts, ex-post measures of the realized quantity in question are required. As demonstrated in Andersen and Bollerslev (1998) among others, the use of one-period squared returns provides unbiased, yet very noisy, realized volatility estimates. As a result, one-period squared returns will suggest lower explanatory power for volatility forecasts, even when the latter are indeed optimal. More reliable inferences can be made when realized volatility is calculated by aggregating intra-day returns, which drastically eliminates noise in the measurement of volatility.

Formally, ex-post measures of integrated variance realized on day t, using N

were very similar.

[7] The results are available from the author upon request.

intraday returns plus the squared overnight return, can be obtained by

$$RVAR_{t,N} = \sum_{j=1}^{N} [(ln(S_{t,j}) - ln(S_{t,j-1})]^2 + [(ln(S_{t,OPEN}) - (ln(S_{t,CLOSE})]^2 \quad (8)$$

In this study $N = 13$ as intraday prices are recorded every 30-minutes, from 9:30 until 16:00. So as to remove the spurious serial correlation of high-frequency returns induced by market microstructure effects, this study follows Andersen et al (2001) in estimating demeaned pseudo-returns obtained by an applying an MA(1) filter. In a similar fashion, T-day realized variance estimates are obtained by accumulating intraday and overnight returns for the relevant horizon, while realized volatility (i.e. standard deviation) measures are obtained by taking square roots.

4.1 Sample Construction

The relative performance of the competing volatility forecasts is assessed using both in-sample and out-of-sample criteria, for two distinct horizons, i.e. monthly and daily. Both realized standard deviations as well as variances are examined. Typically, standard deviations are the object of interest in the related literature, although the theory of model-free volatilities is based on the variance of returns, so that examining the latter quantity is more natural from this perspective.

Having estimated risk-neutral densities for all days in the sample, it is now possible to construct a series of non-overlapping forecasts for the two horizons in question. While our daily forecasts are non-overlapping by definition, for the monthly horizon case this paper follows a construction methodology similar to that of Jiang and Tian (2005). Specifically, the Wednesday immediately following the expiration date of each month is chosen as the day when implied volatilities are observed. If this happens to be a non-trading day then the following Thursday is selected. If this is also a non-trading day then the preceding Tuesday is chosen[8]. The forecast horizon is then determined by the minimum of 30 calendar days and the difference between the current and the next observation date. While in doing so the forecast horizon is not always exactly 30 calendar days, the resulting forecasts are strictly non-overlapping. The other alternative would be to fix the horizon at 30 days, as in Jiang and Tian (2005), but this would result in forecasts that would occasionally overlap.

[8]Jiang and Tian (2005) use this sample construction methodology motivated by the liquidity patterns of the listed options.

Regarding the out-of-sample forecast models, the parameter estimation is done using a rolling window approach. In particular, the estimation window width is kept fixed and corresponds to 1 year and 4 years of past data, for the daily and monthly case respectively.

4.2 Ranking Volatility Forecasts

One major consideration when ranking volatility forecast models is that the target quantity is a latent variable, so its realized estimates will contain measurement errors. Inference and model comparison problems that arise in the presence of noise in the volatility proxy are analyzed in Andersen and Bollerslev (1998), Meddahi (2001) and Andersen et al (2005) among others. Focusing particularly on the issue of ranking competing volatility forecasts, Hansen and Lunde (2006) and Patton (2006) demonstrate that severe distortions can arise when noisy volatility proxies are used.

We attempt to alleviate such considerations in two ways. The first is by using 30-minute returns in order to get more precise realized volatility estimates. As Patton and Sheppard (2007) demonstrate in a simulation environment, daily realized variance proxies based on 30-minute, instead of daily, returns provide large gains in terms of consistent ranking and forecast comparison tests. In the case of monthly realized volatility estimates, it is reasonable to expect that distortions will be significantly smaller. Secondly, as it is discussed in Section 4.2.2, the loss functions employed in this study are "robust" to noise, in the sense of Patton (2006) and Hansen and Lunde (2006).

4.2.1 In-Sample Evaluation Criteria

As it is common in the literature (Canina and Figlewski, 1993; Christensen and Prabhala, 1998; Jiang and Tian, 2005; Taylor et al, 2007; Andersen and Bondarenko, 2007) the in-sample fit of the volatility forecasts for a given stock is investigated using the Mincer-Zarnowitz regression, i.e.

$$RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H} \quad , \tag{9}$$

where the superscript H indicates the forecast horizon (either 1 day or approximately 30 calendar days), the subscript t the time at which the forecast is made, RV is the realized volatility from time t up to time $t+H$ and i denotes the option-implied volatility measure (for instance BSIV or MFIV).

The resulting R^2 of this regression gives an indication of the association between volatility forecasts and volatility realizations, so that indicative comparisons between different specifications can be made. Similarly, one can also examine "encompassing-style" regressions defined as:

$$RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H} \tag{10}$$

The first lag of realized volatility is added as an explanatory variable at the right-hand side of the above equation because it may contain information that is absent from certain implied volatility measures and, thus, improve the forecasting performance of reference models[9].

4.2.2 Out-of-Sample Evaluation Criteria

While the above OLS specifications consider forecast values against realized quantities, ranking models in this way may be misleading because the parameters of the each specification are estimated ex-post. Consequently, the resulting R^2 will incorporate a look-back bias that will overstate the actual degree of forecast accuracy. A preferred methodology is to use data known up to time t in order to estimate the parameters corresponding to equations (9) and (10) and obtain the time t forecast of the target quantity $RV_{t,t+H}$, say $RV^F_{i,t,t+H}$. According to the adopted loss function, competing models can be then evaluated according to their realized losses.

In this study, forecast comparisons are made by relying on the parametric family of "robust" loss functions defined in Patton (2006). This appears as a natural choice since, as shown in Patton (2006), such loss functions have the attractive property that they deliver consistent rankings of realized variance forecasts even when the underlying latent target quantity is replaced by a noisy, but conditionally unbiased, proxy. For a particular forecast produced by model i at time t, the employed loss functions are defined as:

$$MSE_{i,t} : L(RV_{t,t+H}, RV^F_{i,t,t+H}) = (RV_{t,t+H} - RV^F_{i,t,t+H})^2 \tag{11}$$

$$QLIKE_{i,t} : L(RV_{t,t+H}, RV^F_{i,t,t+H}) = log(RV^F_{i,t,t+H}) + \frac{RV_{t,t+H}}{RV^F_{i,t,t+H}} \tag{12}$$

[9]It should be noted that this paper does not investigate the relative performance of volatility forecasts obtained by option prices versus those constructed using solely historical return information. In order to do so, one should explicitly model the volatility process (using past returns) instead of merely relying on this naive (i.e. the lag value of RV_t) forecast.

The first loss function is the well-studied Mean-Squared-Error (MSE), while the second one is the QLIKE function discussed in Bollerslev et al (1994). One important difference between the two is that the MSE criterion treats positive and negative errors equally, while the QLIKE loss function imposes a larger penalty when the forecast underestimates the realized quantity. Given that positive volatility spikes are generally associated with bad news and that investors treat gains and losses differently, this can be considered as a desirable property.

Pair-wise comparisons of volatility forecasts are implemented using the test of Diebold and Mariano (1995). Specifically, defining the average loss differential d_T for a given loss function L as

$$d_T = \frac{1}{T}\sum_{t=1}^{T} d_{i,j,t} \quad , \tag{13}$$

$$where \;\; u_{i,t} = L(RV_{t,t+H}, RV^F_{i,t,t+H}) \;\; and \quad d_{i,j,t} = u_{i,t} - u_{j,t} \;\; , \tag{14}$$

a Diebold-Mariano test (DM henceforth) of equal predictive accuracy can be conducted as a standard t-test using the following statistic:

$$DM_T = \frac{\sqrt{T} d_T}{\sqrt{T^{-1}\hat{V}[\sqrt{T} d_T]}} \quad , \tag{15}$$

where $\hat{V}$ is a consistent estimator of asymptotic variance of $\sqrt{T} d_T$. Under the null hypothesis of equal predictive ability the test statistic has, asymptotically, a standard normal distribution[10].

5 Empirical Results

Throughout this section, only the case when volatility is the forecast quantity is analyzed. The results for the variance case point towards the same findings, so we omit this part for reasons of brevity. Similar tables as those presented in this section, along with a brief discussion of the results, can be found in Appendix B[11].

[10] In this paper the relevant hypothesis are examined using two-tailed tests. The asymptotic variance estimates are computed using the Newey-West (1986) estimator with 10 lags.

[11] One case that might be of interest for the reader is that of the out-of-sample daily regressions results, where the CIV4 measure performs worst on average than some other measures.

5.1 Monthly Horizon

5.1.1 In-Sample

We begin the discussion of the results from the monthly horizon case. Table 1 presents pair-wise comparisons between the competing alternatives. These comparisons are made using both the univariate Minzer-Zarnowitz (MZ) and the bivariate regression specifications (presented in parenthesis). The statistics in Table 1 indicate the proportion of firms for which a reference model, presented in rows, exhibited a higher R^2 than the column model.

Both MZ and encompassing regressions point towards similar findings. The CIV4 measure, which corresponds to a wide corridor, emerges as the best performer on average. When compared to other forecasts, CIV4 produces a larger R^2 in the MZ (encompassing) regressions for 73% (73%) of the firms when compared with MFIV and for 60% (67%) of the firms for the case of BSIV. The outperformance of CIV4 is even greater when the other measures are considered, with the statistics ranging from 77% to 100%. The next best performers are the MFIV and BSIV, with the MFIV being marginally better in the MZ regression and marginally worst in the encompassing regression framework.

Focusing exclusively on the corridor volatility measures, two patters emerge. The first one is that the explanatory power in the OLS regressions increases monotonically as we move towards wider corridor measures (i.e. from CIV1 to CIV4), but as we move from the wide corridor CIV4 measure towards the full MFIV measure the explanatory power decreases. The second pattern is that DVOL appears more informative than the UVOL measure. In particular, DVOL produces a higher R^2 than UVOL for 80% and 67% of the DJIA components, using univariate and bivariate regression specifications respectively.

5.1.2 Out-of-Sample

Pair-wise, out-of-sample, comparisons between all models are presented in Table 2, where each element i, j in the table corresponds to the proportion of firms for which the model in row i exhibited a smaller realized loss compared to the model in column j. Starting from the univariate regressions, CIV4 provides again the most accurate volatility forecasts on average, irrespective of the loss function. In terms of MSE (QLIKE), CIV4 produces a lower realized loss for 63% (63%) of the firms when compared to the MFIV, 67% (77%) when compared to the BS and 57% (70%) when the competing model relies on the DVOL measure. CIV4 also provides

TABLE 1
Comparison of Volatility Forecasts, In-Sample, Monthly Horizon

Univariate (Bivariate) OLS Regressions

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.57 (0.47)	0.87 (0.73)	0.8 (0.67)	0.67 (0.5)	0.27 (0.27)	0.67 (0.7)	1 (0.93)
BSIV	0.43 (0.53)	-	0.93 (0.93)	0.93 (0.8)	0.83 (0.7)	0.33 (0.4)	0.57 (0.6)	1 (0.93)
CIV1	0.13 (0.27)	0.07 (0.07)	-	0.03 (0.17)	0.03 (0.07)	0.1 (0.17)	0.4 (0.47)	0.77 (0.77)
CIV2	0.2 (0.33)	0.07 (0.2)	0.97 (0.83)	-	0.03 (0.1)	0.17 (0.23)	0.43 (0.53)	0.87 (0.8)
CIV3	0.33 (0.5)	0.17 (0.3)	0.97 (0.93)	0.97 (0.9)	-	0.23 (0.3)	0.47 (0.57)	0.9 (0.87)
CIV4	0.73 (0.73)	0.67 (0.6)	0.9 (0.83)	0.83 (0.77)	0.77 (0.7)	-	0.77 (0.77)	1 (0.93)
DVOL	0.33 (0.3)	0.43 (0.4)	0.6 (0.53)	0.57 (0.47)	0.53 (0.43)	0.23 (0.23)	-	0.8 (0.67)
UVOL	0 (0.07)	0 (0.07)	0.23 (0.23)	0.13 (0.2)	0.1 (0.13)	0 (0.07)	0.2 (0.33)	-

NOTE: Each element i, j in the table refers to the proportion of firms for which the reference forecast model in row i exhibited a higher R^2 than a competing alternative, presented in column j. The OLS models are those of Section4.2.1.

better results on average than the CIV1, CIV2, CIV3 and UVOL measures with the relevant statistics being 80 % (80%), 70% (80%), 63 % (63%) and 87 % (93%), respectively. The second best measure appears to be the MFIV, whose forecasts are found somewhat more accurate than those of the BSIV (57% for the MSE and 53% for the QLIKE criterion). MFIV also ranks favorably against the rest of the measures. The performance of narrow CIV forecasts is again poor, while UVOL delivers the worst results. Notably, DVOL is superior to its natural counterpart (UVOL) for the overwhelming majority of the firms in the sample (73% under the MSE and 80% under the QLIKE criterion).

Additional evidence are provided by the Diebold-Mariano (DM) tests of equal predictive ability[12]. From tests that are conducted at the 5% level it is evident that UVOL produces the least efficient forecasts, as the null of equal predictive ability is often rejected in favor of other alternatives. When the loss is defined by the MSE function, the percentage of firms for which UVOL is rejected ranges from 13% to 30%, while rejection rates are even higher for the case of the QLIKE, where the corresponding statistics range from 23% to 53%. CIV1 and CIV2 are also quite frequently rejected in favor of the BSIV, MFIV or broader corridor measures. It is difficult to draw a firm conclusion from the comparative evaluation of the MFIV, BSIV and CIV4 expectations because the rejection rates are generally low. The only exception is the difference between BSIV and CIV4 measures where the null, under a QLIKE loss, is rejected 17% of the time in favor of the CIV4 while the opposite is never true. Lastly, the substantial difference between the DVOL and UVOL rejection rates further confirms the significant differences in their forecasting

[12]The detailed results corresponding to these DM tests can be found in Appendix C.

TABLE 2
Comparison of Volatility Forecasts, Out-of-Sample, Monthly Horizon

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.57 (0.53)	0.8 (0.77)	0.7 (0.63)	0.6 (0.6)	0.37 (0.37)	0.63 (0.67)	0.9 (0.97)
BSIV	0.43 (0.47)	-	0.83 (0.83)	0.8 (0.8)	0.67 (0.63)	0.33 (0.23)	0.53 (0.6)	0.87 (0.93)
CIV1	0.2 (0.23)	0.17 (0.17)	-	0.13 (0.13)	0.13 (0.1)	0.2 (0.2)	0.33 (0.33)	0.83 (0.87)
CIV2	0.3 (0.37)	0.2 (0.2)	0.87 (0.87)	-	0.13 (0.13)	0.3 (0.2)	0.4 (0.47)	0.87 (0.93)
CIV3	0.4 (0.4)	0.33 (0.37)	0.87 (0.9)	0.87 (0.87)	-	0.37 (0.33)	0.5 (0.53)	0.9 (0.93)
CIV4	0.63 (0.63)	0.67 (0.77)	0.8 (0.8)	0.7 (0.8)	0.63 (0.67)	-	0.57 (0.7)	0.87 (0.93)
DVOL	0.37 (0.33)	0.47 (0.4)	0.67 (0.67)	0.6 (0.53)	0.5 (0.47)	0.43 (0.3)	-	0.73 (0.8)
UVOL	0.1 (0.03)	0.13 (0.07)	0.17 (0.13)	0.13 (0.07)	0.1 (0.07)	0.13 (0.07)	0.27 (0.2)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.6 (0.47)	0.7 (0.73)	0.63 (0.67)	0.63 (0.57)	0.23 (0.37)	0.63 (0.67)	0.9 (0.93)
BSIV	0.4 (0.53)	-	0.83 (0.77)	0.77 (0.73)	0.67 (0.67)	0.33 (0.37)	0.6 (0.67)	0.9 (0.93)
CIV1	0.3 (0.27)	0.17 (0.23)	-	0.2 (0.23)	0.17 (0.17)	0.27 (0.27)	0.4 (0.4)	0.9 (0.9)
CIV2	0.37 (0.33)	0.23 (0.27)	0.8 (0.77)	-	0.17 (0.2)	0.33 (0.3)	0.5 (0.53)	0.93 (0.93)
CIV3	0.37 (0.43)	0.33 (0.33)	0.83 (0.83)	0.83 (0.8)	-	0.33 (0.37)	0.5 (0.63)	0.9 (0.93)
CIV4	0.77 (0.63)	0.67 (0.63)	0.73 (0.73)	0.67 (0.7)	0.67 (0.63)	-	0.67 (0.73)	0.9 (0.93)
DVOL	0.37 (0.33)	0.4 (0.33)	0.6 (0.6)	0.5 (0.47)	0.5 (0.37)	0.33 (0.27)	-	0.67 (0.83)
UVOL	0.1 (0.07)	0.1 (0.07)	0.1 (0.1)	0.07 (0.07)	0.1 (0.07)	0.1 (0.07)	0.33 (0.17)	-

NOTE: Proportion of firms for which a reference model i, represented in rows, exhibited a smaller realized loss than a competing alternative, represented in columns, under the MSE (QLIKE) criterion. The model parameters are estimated by rolling regressions using alternative implied volatility (IV) specifications and monthly realized volatilities (RV).

performance.

A similar picture is given by the DM tests conducted at the 10% level. What stands out is that under the MSE (QLIKE) loss function CIV4 is significantly more accurate than the MFIV and BSIV measures for 13% (13%) and 20% (23%) of the firms respectively. The rejection rates that correspond to the opposite case are lower, with the relevant statistics being 13% (7%) for the BSIV case, while MFIV forecasts do not significantly outperform CIV4 forecasts for any of the 30 stocks. Regarding the relative performance of MFIV and BSIV, their differences are small. Bivariate forecast models, where lagged realized volatility is also included in the rolling window regressions, do not alter the findings that emerged from the univariate specifications. For a substantial majority of the firms, CIV4 again produces the smaller realized loss irrespective of the implied volatility measure included the in alternative model. Specifically, CIV4 produces smaller MSE (QLIKE) losses for 77% (63%) of the firms when the alternative model contains the MFIV and 67% (63%) when the model uses BS implied volatilities. The same holds true for the rest of the forecast models with the corresponding statistics ranging from 63% to 93%.

Further support regarding the superiority of the CIV4 when compared to the MFIV is provided by the DM statistics. For tests at conducted at the 10% level, CIV4 significantly outperforms a MFIV specification for 20% of the firms using under either a MSE or a QLIKE loss, while the respective hypothesis is never rejected in favor of the MFIV.

Forecast comparisons between BSIV and MFIV again provide mixed results. Under a MSE loss, MFIV produces better results than the BSIV forecasts for 60% of the firms, however, the DM tests do not offer further support of superior forecasting accuracy since the null of equal predictive accuracy is rejected, at the 10% level, in favor of MFIV for only one stock. On the contrary, DM test conducted at the same significance level indicate that an encompassing model that includes BSIV implied volatilities is significantly better than the MFIV measure for 13% of the firms. When the QLIKE function is employed, the performance of BSIV and MFIV models is essentially indistinguishable, since the corresponding statistics are very similar.

In agreement with the in-sample findings, CIV measures perform better as we move from CIV1 to CIV4. This can be either due to the fact that the corridor measure is getting closer to the definition of the target quantity or because further volatility information is contained in away-from-the-money options. On the other hand, MFIV is outperformed by CIV4, presumably reflecting the difficulty of estimating the tails of the risk-neutral density with acceptable accuracy.

5.2 Daily Horizon

5.2.1 In-Sample

We now turn to the daily horizon case. Table 3 contains the results of all pairwise comparisons, in terms of R^2, using the MZ and the "encompassing" forecast regressions. Since the results are generally the same irrespective of the specification, they are discussed jointly.

When judged against other models, CIV4 again produces a higher R^2 for the majority of the firms. For the MZ (encompassing) regression case the relevant statistics are 67% (70%) for the MFIV, 63% (60%) for the BSIV and 80% (77%) for the CIV3. For the rest of the forecast models, CIV4 displays a higher R^2 for the overwhelming majority of the firms, with the statistics ranging from 87% to 100%. MFIV comes as second best model in this respect, surpassing BSIV for 57% of the firms for either OLS specification.

TABLE 3
Comparison of Volatility Forecasts, In-Sample, Daily Horizon

Univariate (Bivariate) OLS Regressions

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.57 (0.57)	0.97 (0.93)	0.8 (0.8)	0.7 (0.63)	0.33 (0.3)	0.97 (0.93)	0.97 (0.93)
BSIV	0.43 (0.43)	-	1 (1)	1 (0.97)	0.97 (0.93)	0.37 (0.4)	0.87 (0.83)	0.97 (0.97)
CIV1	0.03 (0.07)	0 (0)	-	0 (0)	0 (0)	0 (0)	0.53 (0.5)	0.9 (0.87)
CIV2	0.2 (0.2)	0 (0.03)	1 (1)	-	0 (0)	0.07 (0.13)	0.7 (0.7)	0.93 (0.9)
CIV3	0.3 (0.37)	0.03 (0.07)	1 (1)	1 (1)	-	0.2 (0.23)	0.8 (0.8)	0.97 (0.93)
CIV4	0.67 (0.7)	0.63 (0.6)	1 (1)	0.93 (0.87)	0.8 (0.77)	-	0.93 (0.93)	0.97 (0.93)
DVOL	0.03 (0.07)	0.13 (0.17)	0.47 (0.5)	0.3 (0.3)	0.2 (0.2)	0.07 (0.07)	-	0.73 (0.73)
UVOL	0.03 (0.07)	0.03 (0.03)	0.1 (0.13)	0.07 (0.1)	0.03 (0.07)	0.03 (0.07)	0.27 (0.27)	-

NOTE: Each element i, j in the table refers to the proportion of firms for which the reference forecast model in row i exhibited a higher R^2 than a competing alternative, presented in column j. The OLS models are those of Section 4.2.1.

The main patterns discussed in the monthly horizon case survive in the daily setting as well. Specifically, volatility expectations conditioned on narrow corridors perform rather poorly, while DVOL, although also exhibiting quite poor results, is again better that UVOL for 73% of the firms.

5.2.2 Out-of-Sample

Out-of-sample results of the rolling regressions, presented in Table 4, continue to offer support for the CIV4 measure. The proportion of firms for which CIV4 produces lower losses compared to the BSIV model is still above 50%, although the differences are not large. Particularly, CIV4-based forecasts are more accurate for 57% of the firms under both the MSE and the QLIKE loss functions at the univariate regressions, while the corresponding proportion for the encompassing specifications is 53% and 63%, for the MSE and the QLIKE respectively. Tests for equal predictive ability usually indicate very low rejection rates in favor of either model, although the number of firms for which the null is rejected in favor of the CIV4 forecasts is never lower than the number of significant rejections that correspond to the opposite case. A similar picture is found when CIV4 is compared to the MFIV. This time, the proportion of firms for which CIV4 is more accurate ranges from 67% to 73% for different OLS models and loss functions. The DM statistics, however, display very low rejection rates.

A contest for the second best performer is between models containing the BSIV and the MFIV measures but, once more, it is impossible to draw a clear winner between the two. Forecast specifications that contain BSIV deliver better forecasts

TABLE 4
Comparison of Volatility Forecasts, Out-of-Sample, Daily Horizon

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.47 (0.53)	0.97 (1)	0.83 (0.87)	0.7 (0.77)	0.27 (0.33)	1 (1)	1 (1)
BSIV	0.53 (0.47)	-	1 (1)	1 (1)	0.93 (0.9)	0.47 (0.37)	1 (1)	1 (1)
CIV1	0.03 (0)	0 (0)	-	0 (0)	0 (0)	0 (0)	0.47 (0.5)	0.77 (0.8)
CIV2	0.17 (0.13)	0 (0)	1 (1)	-	0 (0)	0.1 (0.1)	0.7 (0.77)	0.87 (0.93)
CIV3	0.3 (0.23)	0.07 (0.1)	1 (1)	1 (1)	-	0.27 (0.13)	0.93 (0.93)	1 (0.97)
CIV4	0.73 (0.67)	0.53 (0.63)	1 (1)	0.9 (0.9)	0.73 (0.87)	-	1 (1)	1 (1)
DVOL	0 (0)	0 (0)	0.53 (0.5)	0.3 (0.23)	0.07 (0.07)	0 (0)	-	0.63 (0.77)
UVOL	0 (0)	0 (0)	0.23 (0.2)	0.13 (0.07)	0 (0.03)	0 (0)	0.37 (0.23)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.4 (0.57)	0.93 (0.97)	0.73 (0.9)	0.57 (0.77)	0.3 (0.37)	0.97 (1)	0.97 (1)
BSIV	0.6 (0.43)	-	1 (1)	1 (0.97)	0.87 (0.9)	0.43 (0.33)	0.97 (1)	1 (1)
CIV1	0.07 (0.03)	0 (0)	-	0 (0)	0 (0)	0.03 (0)	0.43 (0.5)	0.83 (0.9)
CIV2	0.27 (0.1)	0 (0.03)	1 (1)	-	0 (0)	0.07 (0.1)	0.7 (0.73)	0.9 (0.93)
CIV3	0.43 (0.23)	0.13 (0.1)	1 (1)	1 (1)	-	0.27 (0.13)	0.9 (0.93)	0.97 (0.97)
CIV4	0.7 (0.63)	0.57 (0.67)	0.97 (1)	0.93 (0.9)	0.73 (0.87)	-	1 (1)	0.97 (1)
DVOL	0.03 (0)	0.03 (0)	0.57 (0.5)	0.3 (0.27)	0.1 (0.07)	0 (0)	-	0.63 (0.7)
UVOL	0.03 (0)	0 (0)	0.17 (0.1)	0.1 (0.07)	0.03 (0.03)	0.03 (0)	0.37 (0.3)	-

NOTE: Proportion of firms for which a reference model i, represented in rows, exhibited a smaller realized loss than a competing alternative, represented in columns, under the MSE (QLIKE) criterion. The model parameters are estimated by rolling regressions using alternative implied volatility(IV) specifications and daily realized volatilities (RV).

for the majority of the firms under a MSE loss (53% and 60% for the univariate and bivariate specifications respectively) but the ranking is reversed under a QLIKE loss, with the statistics favoring the MFIV (53% and 60% for the univariate and bivariate specifications respectively).

6 Conclusion

In this paper the forecasting performance of different option-implied volatility measures was evaluated and compared for the case of the DJIA components. Along with the popular Black-Scholes and “model-free” implied volatility expectations, the recently suggested corridor implied volatility (CIV) measures were included in the analysis. In contrast to model-free implied volatility, corridor volatility contracts accumulate return variation only when the futures price lies within two, pre-specified, barrier levels. For this reason, only options that have strike prices within the corridor range are required to compute the corresponding option-implied expectations.

As the corridor becomes wide enough so as to contain all the asset levels for which the risk-neutral measure assigns a positive probability, the definitions of corridor and model-free implied volatility coincide.

Along with at-the-money Black-Scholes and model-free implied volatilities, six corridor-related measures were included in the empirical analysis. Using realized volatility or variance as the target quantity, univariate and bivariate, regression-based, forecast models were evaluated for two distinct horizons (daily and monthly). Pair-wise comparisons were made using three different ranking criteria (R^2, MSE, QLIKE). Ex-post realizations of return variation were computed using intra-day data, so as to alleviate the effect of measurement errors.

When the forecast quantity was realized volatility, a particular CIV measure with a wide corridor width delivered the most accurate forecasts for the majority of the firms, in all settings and for all possible pair-wise comparisons. When the forecast quantity was realized variance, the same was true in 9 out of the 12 possible settings. Formal tests of predictive accuracy, although still favoring the same measure, did not clearly establish its superiority. This is perhaps natural, given that most measures share a similar information set. As far as model-free and Black-Scholes forecasts are concerned, it was impossible to draw a clear winner.

The results of this paper indicate that systematic differences exist in the performance of several CIV measures. For the sample of the 30 DJIA stocks, CIV forecasts became increasingly more accurate as the width of the corridor increased. When, however, the corridor became so wide so as to include extreme strikes, and hence equal the full model-free volatility estimate, their forecast accuracy weakened. This is not surprising, given that the tails of the risk-neutral density are notoriously difficult to estimate.

The results of this study are in line with the work of Andersen and Bondarenko (2007) who also favor a CIV approach in their S&P 500 study. Nevertheless, it is important to consider other asset classes and markets in the analysis in order to firmly establish the overall usefulness of such measures in the context of volatility forecasting. Given that the determinants of option premiums, as well the liquidity characteristics of the underlying markets, exhibit substantial differences from asset to asset, further empirical studies are required.

References

[1] Andersen, T.G., Bollerslev, T., Diebold, F.X. and H. Ebens, 2001, The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics*, 61, 43-76.

[2] Andersen, T.G., Bollerslev, T. and Francis X. Diebold, 2008, Parametric and Nonparametric Volatility Measurement, *Handbook of Financial Econometrics*, Elsevier Science B.V., Amsterdam, forthcoming.

[3] Andersen, T.G., Bollerslev, T. and Meddahi, N., 2005, Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities, *Econometrica*, 73, 1, 279-296.

[4] Andersen, T.G. and Tim Bollerslev, 1998, Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, 39, 885-905.

[5] Andersen, T.G. and Oleg Bondarenko, 2007, Construction and Interpretation of Model-Free Implied Volatility, Working Paper, Northwestern University.

[6] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, Asger and N. Shephard, 2008, Multivariate Realised Kernels: Consistent Positive Semi-definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading, University of Aarhus.

[7] Beckers, S., 1981, Standard Deviations Implied in Options Prices as Predictors of Future Stock Price Variability, *Journal of Banking and Finance*, 5, 363-381.

[8] Bertsimas, Dimitris and Ioana Popescu, 2002, On the Relation between Option and Stock Prices: A Convex Optimization Approach, *Operations Research*, 50, 358-374.

[9] Black, F. and M. Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81, 637-659.

[10] Bliss, R. and N. Panigirtzoglou, 2002, Testing the Stability of Implied Probability Density Functions, *Journal of Banking and Finance*, 26, 381-422.

[11] Bollerslev, T., Engle, R. F. and D. Nelson, 1994, ARCH Models, *Handbook of Econometrics*, Vol. IV, Elsevier Science B.V., 2961-3038.

[12] Bollerslev, T., Gibson, M., and Hao Zhou, 2006, Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities, Working Paper, Federal Reserve Board.

[13] Breeden, D., and R. Litzenberger, 1978, Prices of State Contingent Claims Implicit in Options Prices, *Journal of Business*, 51, 621-652.

[14] Britten-Jones, M. and A. Neuberger, 2000, Option Prices, Implied Price Processes, and Stochastic Volatility, *Journal of Finance*, 55, 839-866.

[15] Canina, L., and S. Figlewski, 1993, The Information Content of Implied Volatility, *Review of Financial Studies*, 6, 659-681.

[16] Carr, P., and Dilip Madan, 1998, Towards a Theory of Volatility Trading, Volatility, *Risk Publications*, R. Jarrow, Ed., 417-427.

[17] Carr, P., and Liuren Wu, 2009, Variance Risk Premiums, *Review of Financial Studies*, 22, 1311-1341.

[18] Chiras, D.P. and S. Manaster, 1978. The Information Content of Option Prices and a Test for Market Efficiency. *Journal of Financial Economics*, 6, 213-234.

[19] Christensen, B.J., and N.R. Prabhala, 1998, The Relation between Implied and Realized Volatility, *Journal of Financial Economics*, 59, 125-150.

[20] Demeterfi, K., Derman, E., Kamal, M., and J. Zou, 1999, A Guide to Volatility and Variance Swaps, *Journal of Derivatives*, 6, 9-32.

[21] Diebold, F.X., and R.S. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 3, 253-263.

[22] Ederington, L.H. and Wei Guan, 2002, Measuring Implied Volatility: Is an Average Better? Which Average? *Journal of Futures Markets*, 22, 811-837.

[23] Fung, H.G., Lie C.J, and A. Moreno, 1990, The Forecasting Performance of the Implied Standard Deviation in Currency Options, *Managerial Finance*, 16, 3, 24-29.

[24] Gemmill, Gordon, 1986, The Forecasting Performance of Stock Options on the London Traded Options Markets, *Journal of Business Finance and Accounting*, 13, 4, 535-546.

[25] Hansen, P.R. and Asger Lunde, 2006, Consistent Ranking of Volatility Models, *Journal of Econometrics*, 131, 97-12.

[26] Jackwerth, J.C, 1999, Option Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review, *Journal of Derivatives*, 6, 66-82.

[27] Jiang, G., and Y. Tian, 2005, The Model-Free Implied Volatility and Its Information Content, *Review of Financial Studies*, 18, 1305-1342.

[28] Lamoureux, C.G. and W.D. Lastrapes, 1993, Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities, *Review of Financial Studies*, 6, 293-326.

[29] Latane, H.A. and R.J. Rendleman, 1976. Standard Deviations of Stock Price Ratios Implied in Option Prices, *Journal of Finance*, 31, 369-381.

[30] Malz, A., 1997, Estimating the Probability Distribution of the Future Exchange Rate from Option Prices, *Journal of Derivatives*, 20-36.

[31] Meddahi, N., 2001, A Theoretical Comparison between Integrated and Realized Volatilities, manuscript, University of Montreal.

[32] Patton, A.J., 2006, Volatility Forecast Comparison Using Imperfect Volatility Proxies, Research Paper 175, Quantitative Finance Research Center, University of Technology Sydney.

[33] Patton, A.J. and Kevin Sheppard, 2007, Evaluating Volatility and Correlation Forecasts, Working Paper, University of Oxford.

[34] Poon, S-H., and C.W.J. Granger, 2003, Forecasting Volatility in Financial Markets, *Journal of Economic Literature*, 41, 478-539.

[35] Ross, S., 1976, Options and Efficiency, *Quarterly Journal of Economics*, 90, 75-89.

[36] Taylor, S.J., Yadav, P.K. and Yuanyuan Zhang, 2007, The Information Content of Implied Volatilities and Model-Free Volatility Expectations: Evidence from Options Written on Individual Stocks, Working Paper, Lancaster University.

[37] Taylor, S.J., 2005, Asset Price Dynamics, Volatility, and Prediction, Princeton University Press.

[38] Todorov, V., 2007, Variance Risk Premium Dynamics, Working Paper, Northwestern University.

[39] Trippi, Robert, 1977, A Test of Option Market Efficiency Using a Random-walk Valuation Model, *Journal of Economics and Business*, 29, 93-98.

[40] Shimko, D., 1993, Bounds on Probability, *Risk*, 6, 33-37.

[41] Newey, Whitney and Kenneth West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 3, 703-708.

Appendix

A. Construction of the Dataset

To construct the options dataset for each day, the following screening criteria are used:

1) Option contracts whose time-to-maturity is 8 days or less are ignored.

2) Option prices that violate standard no arbitrage restrictions are excluded from the dataset. Monotonicity and convexity restrictions are checked using the price bounds of Bertsimas and Popescu (2002).

3) Only options with strictly positive ask prices and strictly positive bid-ask spreads are considered.

4) Contracts whose absolute value of delta (provided by OptionMetrics) is above 0.75 are discarded. This is because, as the early exercise premium is increasing with moneyness, the provided implied volatilities are more susceptible to approximation errors. Moreover, deep in-the-money options are usually the least liquid.

5) Using a similar argumentation, when both puts and calls with the same strike and maturity exist, only the implied volatility of the out-of-money contract is used.

6) From each daily panel, those options that have the two closest maturities are selected. When one of these two contract sets has less than 3 options, then it is replaced by the next closest maturity set. If again, at least one of the contract sets has less than 3 options, then no data are extracted for this particular date and the implied measures are replaced with their last known values.

7) Finally, when for a given maturity at least four contracts with non-zero volume exist, options that have zero volume are eliminated. The reason is that the prices of such options volume are likely to be stale.

B. Forecast Comparison Results, Variance

Tables A1 to A3 contain the results of the pair-wise comparisons when the target quantity is realized variance. Table A1 corresponds to the in-sample regression case, where the competing alternatives are ranked according to the R^2 criterion. Except for the daily horizon case where the statistics for the CIV4 and BSIV measures are equal, CIV4 outperforms other alternatives irrespective of the setting (i.e. daily or monthly horizon, univariate or bivariate regression).

Out-of-sample comparisons in terms of MSE or QLIKE losses for the case of monthly variance forecasts (Table A2) also favor CIV4. While CIV4 and MFIV exhibit equal predictive ability in the univariate models under MSE loss, CIV4 produces the most accurate forecasts for the majority of the firms for any other comparison (univariate or bivariate models, MSE or QLIKE).

The only setting at which the CIV4 measure is somewhat problematic is the, out-of-sample, daily variance forecasting exercise (Table A3). Specifically, univariate models that include BSIV produce more accurate forecasts for 57% of the firms, while in the bivariate regression framework BSIV is more accurate for 60% of the firms under MSE but less accurate for 63% of the firms under QLIKE. Furthermore, MFIV (marginally) outperforms CIV4 in the bivariate regressions in terms of QLIKE losses, while the bivariate models that contain CIV3 or CIV4 have equal statistics under MSE loss. For all other comparisons CIV4 emerges as the best performer.

TABLE A1
Comparison of Variance Forecasts, In-Sample

Panel A. Daily Horizon, Univariate (Bivariate) OLS Regressions

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.53 (0.53)	0.73 (0.7)	0.6 (0.6)	0.57 (0.57)	0.3 (0.27)	0.9 (0.83)	0.77 (0.77)
BSIV	0.47 (0.47)	-	0.9 (0.9)	0.87 (0.87)	0.7 (0.67)	0.5 (0.5)	0.8 (0.77)	0.8 (0.8)
CIV1	0.27 (0.3)	0.1 (0.1)	-	0.07 (0.07)	0.07 (0.07)	0.2 (0.2)	0.6 (0.6)	0.7 (0.73)
CIV2	0.4 (0.4)	0.13 (0.13)	0.93 (0.93)	-	0.13 (0.17)	0.3 (0.3)	0.7 (0.7)	0.73 (0.73)
CIV3	0.43 (0.43)	0.3 (0.33)	0.93 (0.93)	0.87 (0.83)	-	0.4 (0.4)	0.73 (0.73)	0.73 (0.73)
CIV4	0.7 (0.73)	0.5 (0.5)	0.8 (0.8)	0.7 (0.7)	0.6 (0.6)	-	0.9 (0.83)	0.77 (0.77)
DVOL	0.1 (0.17)	0.2 (0.23)	0.4 (0.4)	0.3 (0.3)	0.27 (0.27)	0.1 (0.17)	-	0.53 (0.53)
UVOL	0.23 (0.23)	0.2 (0.2)	0.3 (0.27)	0.27 (0.27)	0.27 (0.27)	0.23 (0.23)	0.47 (0.47)	-

Panel B. Monthly Horizon, Univariate (Bivariate) OLS Regressions

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.67 (0.7)	0.77 (0.77)	0.73 (0.73)	0.7 (0.67)	0.47 (0.47)	0.6 (0.6)	0.9 (0.9)
BSIV	0.33 (0.3)	-	0.9 (0.87)	0.83 (0.77)	0.7 (0.7)	0.33 (0.33)	0.53 (0.57)	0.9 (0.87)
CIV1	0.23 (0.23)	0.1 (0.13)	-	0.1 (0.13)	0.1 (0.13)	0.2 (0.23)	0.33 (0.37)	0.67 (0.63)
CIV2	0.27 (0.27)	0.17 (0.23)	0.9 (0.87)	-	0.07 (0.17)	0.23 (0.23)	0.4 (0.4)	0.77 (0.73)
CIV3	0.3 (0.33)	0.3 (0.3)	0.9 (0.87)	0.93 (0.83)	-	0.3 (0.3)	0.5 (0.5)	0.87 (0.83)
CIV4	0.53 (0.53)	0.67 (0.67)	0.8 (0.77)	0.77 (0.77)	0.7 (0.7)	-	0.6 (0.6)	0.93 (0.9)
DVOL	0.4 (0.4)	0.47 (0.43)	0.67 (0.63)	0.6 (0.6)	0.5 (0.5)	0.4 (0.4)	-	0.77 (0.73)
UVOL	0.1 (0.1)	0.1 (0.13)	0.33 (0.37)	0.23 (0.27)	0.13 (0.17)	0.07 (0.1)	0.23 (0.27)	-

NOTE: Proportion of firms for which a reference model i, represented in rows, exhibited a higher R^2 than a competing alternative, represented in columns.

TABLE A2

Comparison of Variance Forecasts, Out-of-Sample, Monthly Horizon

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.67 (0.53)	0.7 (0.7)	0.7 (0.67)	0.67 (0.6)	0.5 (0.4)	0.7 (0.6)	0.8 (0.9)
BSIV	0.33 (0.47)	-	0.77 (0.77)	0.77 (0.77)	0.67 (0.7)	0.4 (0.43)	0.57 (0.57)	0.73 (0.9)
CIV1	0.3 (0.3)	0.23 (0.23)	-	0.17 (0.2)	0.2 (0.27)	0.2 (0.27)	0.4 (0.3)	0.57 (0.8)
CIV2	0.3 (0.33)	0.23 (0.23)	0.83 (0.8)	-	0.23 (0.23)	0.27 (0.27)	0.43 (0.4)	0.63 (0.83)
CIV3	0.33 (0.4)	0.33 (0.3)	0.8 (0.73)	0.77 (0.77)	-	0.37 (0.27)	0.53 (0.4)	0.7 (0.93)
CIV4	0.5 (0.6)	0.6 (0.57)	0.8 (0.73)	0.73 (0.73)	0.63 (0.73)	-	0.7 (0.63)	0.8 (0.9)
DVOL	0.3 (0.4)	0.43 (0.43)	0.6 (0.7)	0.57 (0.6)	0.47 (0.6)	0.3 (0.37)	-	0.7 (0.8)
UVOL	0.2 (0.1)	0.27 (0.1)	0.43 (0.2)	0.37 (0.17)	0.3 (0.07)	0.2 (0.1)	0.3 (0.2)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.6 (0.53)	0.73 (0.63)	0.73 (0.67)	0.7 (0.57)	0.47 (0.37)	0.67 (0.7)	0.77 (0.9)
BSIV	0.4 (0.47)	-	0.8 (0.7)	0.77 (0.63)	0.77 (0.67)	0.4 (0.37)	0.6 (0.67)	0.73 (0.9)
CIV1	0.27 (0.37)	0.2 (0.3)	-	0.27 (0.3)	0.3 (0.3)	0.27 (0.3)	0.43 (0.4)	0.53 (0.83)
CIV2	0.27 (0.33)	0.23 (0.37)	0.73 (0.7)	-	0.2 (0.3)	0.27 (0.33)	0.47 (0.43)	0.7 (0.87)
CIV3	0.3 (0.43)	0.23 (0.33)	0.7 (0.7)	0.8 (0.7)	-	0.33 (0.37)	0.5 (0.53)	0.7 (0.87)
CIV4	0.53 (0.63)	0.6 (0.63)	0.73 (0.7)	0.73 (0.67)	0.67 (0.63)	-	0.67 (0.73)	0.73 (0.9)
DVOL	0.33 (0.3)	0.4 (0.33)	0.57 (0.6)	0.53 (0.57)	0.5 (0.47)	0.33 (0.27)	-	0.63 (0.8)
UVOL	0.23 (0.1)	0.27 (0.1)	0.47 (0.17)	0.3 (0.13)	0.3 (0.13)	0.27 (0.1)	0.37 (0.2)	-

NOTE: Proportion of firms for which a reference model i, represented in rows, exhibited a smaller realized loss than a competing alternative, represented in columns, under the MSE (QLIKE) criterion. The model parameters are estimated by rolling regressions using alternative implied variance (IV) specifications and monthly realized variances (RV).

TABLE A3

Comparison of Variance Forecasts, Out-of-Sample, Daily Horizon

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.43 (0.43)	0.77 (0.77)	0.6 (0.6)	0.47 (0.47)	0.4 (0.4)	0.93 (0.93)	0.83 (0.83)
BSIV	0.57 (0.57)	-	0.83 (0.83)	0.83 (0.83)	0.73 (0.73)	0.57 (0.57)	0.83 (0.83)	0.83 (0.83)
CIV1	0.23 (0.23)	0.17 (0.17)	-	0.17 (0.17)	0.1 (0.1)	0.17 (0.17)	0.6 (0.6)	0.6 (0.6)
CIV2	0.4 (0.4)	0.17 (0.17)	0.83 (0.83)	-	0.17 (0.17)	0.33 (0.33)	0.73 (0.73)	0.67 (0.67)
CIV3	0.53 (0.53)	0.27 (0.27)	0.9 (0.9)	0.83 (0.83)	-	0.47 (0.47)	0.8 (0.8)	0.8 (0.8)
CIV4	0.6 (0.6)	0.43 (0.43)	0.83 (0.83)	0.67 (0.67)	0.53 (0.53)	-	0.93 (0.93)	0.83 (0.83)
DVOL	0.07 (0.07)	0.17 (0.17)	0.4 (0.4)	0.27 (0.27)	0.2 (0.2)	0.07 (0.07)	-	0.57 (0.57)
UVOL	0.17 (0.17)	0.17 (0.17)	0.4 (0.4)	0.33 (0.33)	0.2 (0.2)	0.17 (0.17)	0.43 (0.43)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.5 (0.57)	0.73 (0.63)	0.57 (0.57)	0.47 (0.63)	0.47 (0.53)	0.93 (0.7)	0.8 (0.6)
BSIV	0.5 (0.43)	-	0.83 (0.5)	0.73 (0.53)	0.7 (0.63)	0.6 (0.37)	0.87 (0.77)	0.87 (0.63)
CIV1	0.27 (0.37)	0.17 (0.5)	-	0.07 (0.37)	0.1 (0.47)	0.23 (0.4)	0.67 (0.47)	0.63 (0.6)
CIV2	0.43 (0.43)	0.27 (0.47)	0.93 (0.63)	-	0.2 (0.47)	0.3 (0.4)	0.7 (0.67)	0.83 (0.67)
CIV3	0.53 (0.37)	0.3 (0.37)	0.9 (0.53)	0.8 (0.53)	-	0.5 (0.33)	0.8 (0.63)	0.83 (0.6)
CIV4	0.53 (0.47)	0.4 (0.63)	0.77 (0.6)	0.7 (0.6)	0.5 (0.67)	-	0.93 (0.73)	0.83 (0.6)
DVOL	0.07 (0.3)	0.13 (0.23)	0.33 (0.53)	0.3 (0.33)	0.2 (0.37)	0.07 (0.27)	-	0.5 (0.7)
UVOL	0.2 (0.4)	0.13 (0.37)	0.37 (0.4)	0.17 (0.33)	0.17 (0.4)	0.17 (0.4)	0.5 (0.3)	-

NOTE: Proportion of firms for which a reference model i, represented in rows, exhibited a smaller realized loss than a competing alternative, represented in columns, under the MSE (QLIKE) criterion. The model parameters are estimated by rolling regressions using alternative implied variance (IV) specifications and daily realized variances (RV).

C. Diebold-Mariano Tests

TABLE B1
Diebold-Mariano Tests, Monthly Horizon, Volatility, MSE

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.03 (0.03)	0.2 (0.33)	0.13 (0.2)	0.03 (0.17)	0 (0)	0.07 (0.23)	0.3 (0.43)
BSIV	0.03 (0.1)	-	0.23 (0.4)	0.2 (0.37)	0.17 (0.27)	0.07 (0.13)	0.1 (0.17)	0.27 (0.43)
CIV1	0.03 (0.07)	0 (0.03)	-	0 (0.03)	0 (0.07)	0.03 (0.07)	0.13 (0.13)	0.17 (0.23)
CIV2	0.03 (0.1)	0.03 (0.07)	0.4 (0.43)	-	0.03 (0.07)	0.07 (0.1)	0.13 (0.17)	0.2 (0.33)
CIV3	0 (0.1)	0 (0.03)	0.23 (0.37)	0.17 (0.33)	-	0.03 (0.1)	0.1 (0.17)	0.17 (0.37)
CIV4	0.07 (0.13)	0.03 (0.2)	0.2 (0.33)	0.17 (0.3)	0.07 (0.23)	-	0.1 (0.3)	0.27 (0.47)
DVOL	0.07 (0.1)	0.07 (0.07)	0.07 (0.17)	0.07 (0.13)	0.07 (0.1)	0.07 (0.07)	-	0.13 (0.23)
UVOL	0 (0.03)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.03)	0.03 (0.03)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0 (0.03)	0.13 (0.2)	0 (0.17)	0 (0.1)	0 (0)	0.07 (0.2)	0.23 (0.33)
BSIV	0.07 (0.13)	-	0.2 (0.43)	0.2 (0.3)	0.07 (0.2)	0.03 (0.1)	0.13 (0.2)	0.23 (0.33)
CIV1	0.07 (0.07)	0.07 (0.07)	-	0 (0.07)	0.07 (0.07)	0.07 (0.07)	0.1 (0.13)	0.1 (0.2)
CIV2	0.03 (0.07)	0.03 (0.07)	0.27 (0.33)	-	0.07 (0.07)	0.07 (0.07)	0.1 (0.17)	0.13 (0.27)
CIV3	0.03 (0.07)	0 (0.03)	0.27 (0.4)	0.2 (0.4)	-	0.03 (0.1)	0.1 (0.2)	0.2 (0.23)
CIV4	0.07 (0.2)	0.07 (0.13)	0.17 (0.3)	0.1 (0.23)	0.07 (0.17)	-	0.1 (0.17)	0.17 (0.37)
DVOL	0.03 (0.03)	0.07 (0.07)	0.07 (0.1)	0.07 (0.07)	0.07 (0.07)	0.07 (0.07)	-	0.13 (0.17)
UVOL	0.03 (0.03)	0 (0)	0 (0)	0 (0)	0 (0)	0.03 (0.03)	0.03 (0.03)	-

NOTE: Proportion of firms for which the null hypothesis of equal predictive ability between model i (presented in rows) and model j (presented in columns) was rejected in favor of the row model. The realized losses are based on the MSE criterion, while the relevant p-values are computed using the two-tailed test of Diebold and Mariano (1995). The test is conducted at the 5% (10%) level. The asymptotic variance of the statistic is computed using the Newey-West estimator with 10 lags. The parameters of each model are estimated by rolling regressions using 4 years of historical data.

TABLE B2
Diebold-Mariano Tests, Monthly Horizon, Volatility, QLIKE

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.07 (0.1)	0.27 (0.4)	0.17 (0.27)	0.07 (0.13)	0 (0)	0.1 (0.17)	0.57 (0.6)
BSIV	0 (0.07)	-	0.37 (0.43)	0.23 (0.4)	0.13 (0.2)	0 (0.07)	0.1 (0.2)	0.43 (0.57)
CIV1	0.03 (0.03)	0.03 (0.03)	-	0 (0.03)	0 (0.03)	0.03 (0.03)	0.07 (0.07)	0.33 (0.37)
CIV2	0 (0.03)	0.03 (0.03)	0.37 (0.43)	-	0.03 (0.03)	0.03 (0.03)	0.07 (0.1)	0.4 (0.43)
CIV3	0 (0.03)	0.03 (0.03)	0.43 (0.43)	0.3 (0.47)	-	0 (0.03)	0.03 (0.13)	0.4 (0.47)
CIV4	0.07 (0.13)	0.17 (0.23)	0.33 (0.47)	0.23 (0.4)	0.17 (0.3)	-	0.1 (0.2)	0.5 (0.63)
DVOL	0.03 (0.1)	0.03 (0.1)	0.13 (0.2)	0.1 (0.2)	0.03 (0.13)	0.03 (0.1)	-	0.23 (0.3)
UVOL	0.03 (0.03)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.03)	0.03 (0.03)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.03 (0.1)	0.13 (0.23)	0.1 (0.17)	0.03 (0.1)	0 (0)	0.1 (0.17)	0.37 (0.47)
BSIV	0.03 (0.13)	-	0.23 (0.37)	0.17 (0.23)	0.1 (0.17)	0.03 (0.07)	0.07 (0.2)	0.33 (0.57)
CIV1	0.07 (0.07)	0.07 (0.07)	-	0 (0.03)	0.03 (0.07)	0.07 (0.07)	0.1 (0.1)	0.3 (0.33)
CIV2	0.07 (0.07)	0.03 (0.07)	0.3 (0.4)	-	0.07 (0.07)	0.07 (0.07)	0.1 (0.17)	0.3 (0.4)
CIV3	0.03 (0.07)	0.03 (0.03)	0.37 (0.4)	0.23 (0.33)	-	0.03 (0.07)	0.13 (0.17)	0.3 (0.4)
CIV4	0.1 (0.2)	0.1 (0.13)	0.17 (0.33)	0.13 (0.23)	0.07 (0.13)	-	0.1 (0.2)	0.33 (0.57)
DVOL	0.03 (0.1)	0.03 (0.07)	0.03 (0.1)	0.03 (0.07)	0.03 (0.07)	0.03 (0.07)	-	0.17 (0.27)
UVOL	0.03 (0.03)	0 (0)	0 (0)	0 (0)	0 (0)	0.03 (0.03)	0.03 (0.03)	-

NOTE: Proportion of firms for which the null hypothesis of equal predictive ability between model i (presented in rows) and model j (presented in columns) was rejected in favor of the row model. The realized losses are based on the QLIKE criterion, while the relevant p-values are computed using the two-tailed test of Diebold and Mariano (1995). The test is conducted at the 5% (10%) level. The asymptotic variance of the statistic is computed using the Newey-West estimator with 10 lags. The parameters of each model are estimated by rolling regressions using 4 years of historical data.

TABLE B3
Diebold-Mariano Tests, Daily Horizon, Volatility, MSE

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.03 (0.07)	0.33 (0.47)	0.2 (0.3)	0.13 (0.13)	0.03 (0.07)	0.5 (0.6)	0.6 (0.73)
BSIV	0 (0.03)	-	0.67 (0.87)	0.43 (0.67)	0.33 (0.43)	0 (0)	0.4 (0.43)	0.7 (0.77)
CIV1	0 (0)	0 (0)	-	0 (0)	0 (0)	0 (0)	0.03 (0.13)	0.2 (0.33)
CIV2	0 (0)	0 (0)	0.8 (0.83)	-	0 (0)	0 (0)	0.13 (0.2)	0.4 (0.47)
CIV3	0 (0)	0 (0)	0.73 (0.83)	0.5 (0.63)	-	0 (0)	0.17 (0.33)	0.5 (0.6)
CIV4	0.03 (0.1)	0 (0.03)	0.43 (0.53)	0.3 (0.37)	0.17 (0.23)	-	0.53 (0.67)	0.67 (0.73)
DVOL	0 (0)	0 (0)	0.03 (0.07)	0 (0)	0 (0)	0 (0)	-	0.1 (0.13)
UVOL	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0.03 (0.07)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0 (0)	0.2 (0.43)	0.13 (0.23)	0.03 (0.1)	0 (0)	0.43 (0.57)	0.47 (0.7)
BSIV	0 (0)	-	0.5 (0.53)	0.37 (0.43)	0.23 (0.33)	0 (0)	0.33 (0.43)	0.6 (0.73)
CIV1	0 (0)	0 (0)	-	0 (0)	0 (0)	0 (0)	0 (0.07)	0.2 (0.3)
CIV2	0 (0)	0 (0)	0.57 (0.8)	-	0 (0)	0 (0)	0.07 (0.17)	0.4 (0.43)
CIV3	0 (0)	0 (0)	0.6 (0.7)	0.4 (0.53)	-	0 (0)	0.17 (0.3)	0.47 (0.5)
CIV4	0.03 (0.07)	0 (0.03)	0.27 (0.43)	0.13 (0.3)	0.03 (0.1)	-	0.43 (0.57)	0.47 (0.67)
DVOL	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	-	0.1 (0.1)
UVOL	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	-

NOTE: Proportion of firms for which the null hypothesis of equal predictive ability between model i (presented in rows) and model j (presented in columns) was rejected in favor of the row model. The realized losses are based on the MSE criterion, while the relevant p-values are computed using the two-tailed test of Diebold and Mariano (1995). The test is conducted at the 5% (10%) level. The asymptotic variance of the statistic is computed using the Newey-West estimator with 10 lags. The parameters of each model are estimated by rolling regressions using 1 year of historical data.

TABLE B4

Diebold-Mariano Tests, Daily Horizon, Volatility, QLIKE

Panel A. Univariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.1 (0.13)	0.43 (0.6)	0.37 (0.43)	0.13 (0.23)	0.03 (0.07)	0.83 (0.9)	0.73 (0.8)
BSIV	0 (0.07)	-	0.8 (0.83)	0.7 (0.8)	0.5 (0.67)	0.03 (0.03)	0.43 (0.6)	0.73 (0.77)
CIV1	0 (0)	0 (0)	-	0 (0)	0 (0)	0 (0)	0.07 (0.17)	0.17 (0.3)
CIV2	0 (0)	0 (0)	0.8 (0.9)	-	0 (0)	0 (0)	0.1 (0.27)	0.37 (0.47)
CIV3	0 (0.03)	0 (0)	0.83 (0.9)	0.73 (0.83)	-	0 (0.03)	0.23 (0.33)	0.57 (0.67)
CIV4	0.03 (0.1)	0.1 (0.2)	0.57 (0.77)	0.43 (0.57)	0.37 (0.4)	-	0.73 (0.83)	0.7 (0.8)
DVOL	0 (0)	0 (0)	0 (0.07)	0 (0)	0 (0)	0 (0)	-	0.1 (0.2)
UVOL	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.07)	-

Panel B. Bivariate Models: $RV_{t,t+H} = \alpha_i + \beta_i IV_{i,t} + \beta_{i,RV} RV_{t-H,t} + \epsilon_{i,t+H}$

	MFIV	BSIV	CIV1	CIV2	CIV3	CIV4	DVOL	UVOL
MFIV	-	0.03 (0.07)	0.3 (0.43)	0.13 (0.27)	0.07 (0.13)	0 (0.03)	0.5 (0.77)	0.63 (0.73)
BSIV	0.03 (0.03)	-	0.67 (0.8)	0.47 (0.7)	0.3 (0.47)	0 (0.03)	0.3 (0.43)	0.63 (0.73)
CIV1	0 (0)	0 (0)	-	0 (0)	0 (0)	0 (0)	0.07 (0.07)	0.17 (0.17)
CIV2	0 (0)	0 (0)	0.67 (0.77)	-	0 (0)	0 (0)	0.1 (0.1)	0.2 (0.43)
CIV3	0.03 (0.07)	0 (0)	0.73 (0.87)	0.6 (0.77)	-	0 (0.03)	0.1 (0.23)	0.43 (0.6)
CIV4	0.07 (0.1)	0.03 (0.03)	0.4 (0.53)	0.33 (0.43)	0.1 (0.27)	-	0.47 (0.73)	0.67 (0.73)
DVOL	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	-	0.07 (0.17)
UVOL	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	-

NOTE: Proportion of firms for which the null hypothesis of equal predictive ability between model i (presented in rows) and model j (presented in columns) was rejected in favor of the row model. The realized losses are based on the QLIKE criterion, while the relevant p-values are computed using the two-tailed test of Diebold and Mariano (1995). The test is conducted at the 5% (10%) level. The asymptotic variance of the statistic is computed using the Newey-West estimator with 10 lags. The parameters of each model are estimated by rolling regressions using 1 year of historical data.

Research Papers 2010

2010-20: Charlotte Christiansen: Intertemporal Risk-Return Trade-off in

2010-21: Marco Aiolfi, Carlos Capistrán and Allan Timmermann : Forecast Combinations

2010-22: Ivan Nourdin, Giovanni Peccati and Mark Podolskij: Quantitative Breuer-Major Theorems

2010-23: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Bootstrapping Density-Weighted Average Derivatives

2010-24: Søren Johansen and Morten Ørregaard Nielsen: Likelihood inference for a fractionally cointegrated vector autoregressive model

2010-25: Tom Engsted and Bent Nielsen: Testing for rational bubbles in a co-explosive vector autoregression

2010-26: Robinson Kruse: On European monetary integration and the persistence of real effective exchange rates

2010-27: Sanne Hiller and Robinson Kruse: Milestones of European Integration: Which matters most for Export Openness?

2010-28: Robinson Kruse: Forecasting autoregressive time series under changing persistence

2010-29: Nikolaus Hautsch and Mark Podolskij: Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence

2010-30: Martin M. Andreasen: Non-linear DSGE Models and The Central Difference Kalman Filter

2010-31: Morten Ørregaard Nielsen and Per Frederiksen: Fully Modified Narrow-Band Least Squares Estimation of Weak Fractional Cointegration

2010-32: Mogens Bladt and Michael Sørensen: Simple simulation of diffusion bridges with application to likelihood inference for diffusions

2010-33: Fernando Baltazar-Larios and Michael Sørensen: Maximum likelihood estimation for integrated diffusion processes

2010-34: Leonidas Tsiaras: The Forecast Performance of Competing Implied Volatility Measures: The Case of Individual Stocks