Centre de Recherche en Economie et Droit de l'Energie

FORECASTING THE DENSITY OF OIL FUTURES RETURNS USING MODEL-FREE IMPLIED VOLATILITY AND HIGH-FREQUENCY DATA

Florian IELPO et Benoît SEVI

Cahier de recherche n° 13.10.106

28 octobre 2013

Centre de Recherche en Economie et Droit de l'Energie

CREDEN – Equipe ART Dev – Université Montpellier 1
Faculté d'Economie, Av. Raymond Dugrand, C.S. 79606
34960 Montpellier Cedex 2, France
Tél. : +33 (0)4 34 43 25 04

Forecasting the density of oil futures returns using model-free implied volatility and high-frequency data*

Florian Ielpo†
Lombard Odier Asset Management
Paris 1 Sorbonne University

Benoît Sévi‡
Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS

September 2013

Abstract

Forecasting the density of returns is useful for many purposes in finance, such as risk management activities, portfolio choice or derivative security pricing. Existing methods to forecast the density of returns either use prices of the asset of interest or option prices on this same asset. The latter method needs to convert the risk-neutral estimate of the density into a physical measure, which is computationally cumbersome. In this paper, we take the view of a practitioner who observes the implied volatility under the form of an index, namely the recent OVX, to forecast the density of oil futures returns for horizons going from 1 to 60 days. Using the recent methodology in Maheu and McCurdy (2011) to compute density predictions, we compare the performance of time series models using implied volatility and either daily or intra-daily futures prices. Our results indicate that models based on implied volatility deliver significantly better density forecasts at all horizons, which is in line with numerous studies delivering the same evidence for volatility point forecast.

JEL Classification: C15, C32, C53, G1, Q4
Keywords: implied volatility, OVX, realized volatility, density forecasting, HAR.

*Acknowledgments to be added.
†Lombard Odier Asset Management, rue de la Corraterie 11, 1204 Geneva, Switzerland and University Paris 1 Panthéon – CERMSEM, Maison des Sciences Économiques, 106 bd de l'Hôpital, 75013 Paris, France. Email: florian.ielpo@ensae.org
‡GREQAM, Aix-Marseille Université, Château La Farge - Route des Milles, 13290 Les Milles Aix-en-Provence, FRANCE, Email: benoit.sevi@gmail.com

1 Introduction

We analyze empirically the information content of model-free implied volatility for forecasting the density of WTI futures returns at horizons going from 1 to 60 days. Our methodology follows Maheu and McCurdy (2011) who develop a bivariate time series model of volatility and returns that is relevant to forecast the density of returns at multiple horizons. We evaluate the forecast accuracy of this model using three different measures for the volatility, namely the realized volatility, the model-free implied volatility (MFIV) and the volatility index computed using options on WTI oil futures (OVX) along with density forecast from a simple EGARCH model. Our results indicate that predictions based on implied volatilities are statistically better than those based on historical data at all horizons.

While the literature considering volatility point forecast for and comparing the information content of option prices with high-frequency data and/or daily data (using GARCH or alternative parametric or nonparametric models) is vast (see below), the question of forecasting the density of returns has not received the same attention.[1] This is quite surprising considering the importance of density prediction. For instance, it allows value-at-risk computation (see Giot and Laurent (2003, 2004), Cabedo and Moya (2003) and Huang et al. (2009) for the case of crude oil). In particular, Giot and Laurent (2004) emphasize the role of intraday data to improve the estimation of the value-at-risk. Our results confirm that intraday data are valuable in improving the accuracy of density forecasts but that forward-looking information in implied volatility leads to even better density predictions that may be helpful in the context of estimating value-at-risk more accurately.

Our study deals with an energy time series for which the number of existing papers about volatility or density forecast is obviously lower than for more classical assets. However, the WTI crude oil futures contract is the most traded futures contract over the world. For instance, in some days, the front-month futures contract exhibits more than 150,000 transactions and the total open interest for the WTI futures has dramatically increased in recent years as highlighted in Tang and Xiong (2012) and Hong and Yogo (2012) among others. This is also an input of utmost importance in asset allocation through direct investment or commodity indexes. As such, an analysis of the density forecast accuracy using various volatility measures for the WTI futures should be of interest both for academics and

[1]The literature on forecasting volatility is nicely surveyed in Poon and Granger (2003) and more recent developments can be found in Andersen *et al.* (2006). These contributions consider the point forecast of volatility which is intuitively of central interest for all finance-related activities. Christoffersen *et al.* (2012) also survey the literature on volatility forecasting with an emphasize on the information extracted from option prices.

practitioners.

Comparison of volatility forecast using risk-neutral vs. physical estimates has been a very active literature in recent years. Early contributions considering implied volatility include Harvey and Whaley (1992), Day and Lewis (1992), Canina and Figlewski (1993), Jorion (1995), Christensen and Prabhala (1998), Fleming (1998) and Poteshman (2000) among others. First attempts to model and forecast volatility using intraday data are Taylor and Xu (1997), Martens (2002) and Andersen *et al.* (2003). As could be expected, literature then investigated the relative performance of models dealing with intraday data with that of models using implied volatility: Blair *et al.* (2001), Martens and Zein (2004), Koopman *et al.* (2005), Gospodinov *et al.* (2006), Becker *et al.* (2007), Giot and Laurent (2007) and Han and Park (2013) are representative examples. The superiority of implied measures is generally accepted as they often subsume the information in measures computed using high-frequency data. The intuition for this result comes, of course, from the forward-looking aspect of option implied information.

Among these studies, only Martens and Zein (2004) is interested in the comparison of volatility forecast models in the case of commodities and in particular in the case of crude oil. Another important reference is Agnolucci (2009) who compare the point forecast of the volatility of oil futures returns using implied or GARCH volatilities as possible predictors, as did Kroner et al. (1995) before him. An advantage of our approach is that we do not rely on a given parametrization in contrast with Kroner et al. (1995), Martens and Zein (2004) or Agnolucci (2009) who assumes a 'Black and Scholes' dynamics for the oil price. As such, the results in these papers are difficult to interpret as model comparison also jointly test the assumed dynamics for oil futures prices.

More classical studies make use of daily data in a GARCH framework. Kang and Yoon (2013) investigates the long-memory properties of three time-series of front-month energy futures contract (WTI, heating oil #2, and unleaded gasoline) volatility relying on a number of long-memory models. As such, they extend the analysis in Kang et al. (2009).[2] The authors find that their models based on daily data provide good in-sample fit for conditional volatility. As for the out-of-sample performance of the competing models, no model emerges as a leading forecasting model based on Diebold-Mariano pair-

[2] Kang *et al.* (2009) use different GARCH models to forecast the oil price volatility and succeed in modeling the long-memory behavior of volatility which is common to most of the financial series (power series are even more persistent). Wei et al. (2010) refine the analysis in Kang et al. (2009). Mohammadi and Su (2010) also consider the GARCH behavior of crude oil prices.

wise comparison. Note that these different studies rely on a very noisy proxy for conditional volatility, namely the squared daily returns.

We enlarge the scope of the above-cited studies in considering the issue of density forecast when a volatility index such as the OVX can be considered as a predictor.[3] The standard methodology estimates the conditional risk-neutral density from a set of option prices with various strikes (see Christoffersen *et al.* (2012)) from which a density forecast (of the risk-neutral density) can be estimated. Then, the econometrician has to make a few theoretical assumptions to recover the physical density from the risk-neutral one, which is a cumbersome step that is highly sensitive to modeling choices.[4] In this paper, we choose to rely exclusively on model-free implied volatilities. This choice has the main drawback that we disregard information about the full option-implied probability distribution. But this is precisely our aim: we want to evaluate the information content of the volatility index itself as we strongly believe that practitioners do not compute risk-neutral probability distributions from the set of available raw option prices. Therefore, we do use information from the risk-neutral world to forecast the physical world density without any mapping attempt between the two measures.[5] As such, we do not take into account the volatility risk premium, that is the difference between implied and realized volatility. It has to be noted that our choice is without consequence as the object of interest in our work is ultimately the relative forecasting performance of the models. In addition, our methodology does not rely on a given dynamics for the oil price. Our choice to use the OVX index is also motivated by the limited number of available option strikes which is low to compute robust density estimates. Indeed, out- and in-the money options are not very liquid and empirical exercices such as in Melick and Thomas (1997) are difficult to implement in a robust manner.

We use the model from Maheu and McCurdy (2011) along with three measures of volatility. Two of them are model-free, *i.e.* they do not require any modeling assumption to derive implied volatility

[3]The volatility index OVX for the WTI futures is computed similarly as the famous VIX index for the S&P 500 index. The interested reader is referred to Christoffersen *et al.* (2012) for a very detailed presentation of the VIX methodology and theoretical and/or empirical work based on this index. Additional references are provided in the Data section.

[4]Day and Lewis (1993) is an early attempt to estimate the risk-neutral density of returns for the WTI futures market. More recently, HÃ¸g and Tsiaras (2011) rely on the approach in Fackler and King (1990), *i.e.* they avoid any assumption on the form of the utility function of the representative agent (as in Bliss and Panigirtzoglou (2004) among others). Moreover, "As reviewed in Ziegler (2007), however, empirical comparisons between risk-neutral and real-world densities of stock indices reveal that such simple transformations are generally problematic since the empirical pricing kernels imply that investors are risk-seekers for certain regions of wealth. Moreover, when option claims are not written on the market index but some other underlying, such as crude-oil futures in our case, extra assumptions are needed regarding the relationship between crude-oil prices and aggregate consumption levels so that further complications would arise." (p. 734) Nevertheless, statistical criterion in Fackler and King (1990) do not alleviate the error-in-variable issue which is a sensible question in the market of options on oil futures.

[5]Christoffersen *et al.* (2012) present methodologies dedicated to the extraction of risk-neutral densities (section 4) and its mapping into real-world densities (section 5) which is an object of interest for researchers and practitioners.

estimate. The first measure is the handily-back-calculated volatility index for the WTI futures market (OVX index) that provides a model-free estimate of the implied volatility for 22-day maturity options on WTI oil futures. As a second model-free measure, we use the MFIV computed following Andersen and Bondarenko (2010) that builds on the corridor variance concept (more on that below). The MFIV is highly correlated with the OVX but offers an alternative to the volatility index by considering all option prices. Then, we compare the predictions from the model in Maheu and McCurdy (2011) using the two model-free measures with the forecast from the same model using the realized volatility as an input and the prediction from an EGARCH model.

Our results provide evidence of the superiority of implied volatility to deliver competitive forecasts of the density of oil futures returns. The results are statistically significant at high thresholds. The sole exception is for the EGARCH model which is not significantly dominated by the model based on implied volatilities at very short horizons (2-3 days).

Our paper is organized as follows. The next Section describes the econometric approach for modelling simultaneously returns and volatility and how density forecasts are derived from this model. Section 4 provides all our empirical findings and evaluates their robustness. Section 5 provides concluding remarks.

2 Quantitative approach

2.1 Extracting implied volatility

We use two different measures of implied volatility that are model-free. We abstract from giving all details for the computation of both implied volatilities and refer the reader to the original contributions.[6]

Martens and Zein (2004) also use option prices to compute implied volatilities. Their methodology follows the approximation derived in Barone-Adesi et Whaley (1987) to calculate the implied volatility of an American style option, which is the case of WTI options on futures. Next, Martens and Zein (2004) consider the "weighted average of the two closest-to-the-money calls and two closest-to-the-

[6]Both methodologies are very long to be presented here in details and such a presentation is not likely to help the reader to better understand what is done in the paper. Moreover, the methodologies are rather standard in financial economics and can be found in a very detailed manner in numerous applications.

money puts where the weights are chosen so that the average strike equals the underlying futures prices."(p. 1013) Such a measure shares some similarities with model-free implied volatility that we use.

The first measure is similar to the official volatility index OVX for WTI futures crude oil that is quoted on the Chicago Board Options Exchange (CBOE). The full characteristics of the volatility index can be found at the webpage "`http://www.cboe.com/micro/oilvix/`". The methodology follows Carr and Wu (2009) whose main idea is to compute the variance risk premium using variance swaps on futures contracts. Because variance swaps on energy commodities are not liquid enough and are not publicly quoted, we proceed as in Trolle and Schwartz (2010) who use theoretical developments in a series of contributions by Britten-Jones and Neuberger (2000), Jiang and Tian (2005) and Carr and Wu (2009) to derive a synthetic variance swap using a cross-section of publicly traded options in futures contracts. The methodology allows to recover the OVX when it did not exist thereby permitting to assess the forecasting value of this measure for predicting the density of futures returns over a longer period.

The second measure follows Andersen and Bondarenko (2010) and is called MFIV. The MFIV is the model-free implied volatility that is derived using the concept of barrier variance contract. Because the barrier variance contract differs from the variance swap, the two measures only coincide when the barrier is infinite which makes the MFIV slightly different from the OVX. For robustness, we perform the empirical analysis using both measures and show that the MFIV has better performance for very short horizons while the OVX outperforms the MFIV for horizons beyond one month and a half.

The problem in estimating implied volatility is that the number of available and liquid strikes can be low thereby limiting the possibility to extract implied volatility in a model-free manner as in Jiang and Tian (2005) or Andersen and Bondarenko (2007). In addition, the derived implied-volatilities are not free from measurement errors (see Covrig and Low, 2003). As such, the use of most of available strikes is useful in measuring the implied volatility. Nevertheless, deep out-of-the-money or in-the-money options may be less liquid and more prone to the error-in-variable problem.[7]

As noted in Jorion (1995) and further recalled in Martens and Zein (2004), using options on futures, what we do, has the clear advantage to limit measurement errors as options and futures are often

[7] Covrig and Low (2003) thus rely on quoted implied volatility instruments. Those instruments deliver better volatility forecasts than those from historical models even for shorter horizons. These results may come from the noisy proxy for volatility against which the volatility prediction performance is evaluated.

traded at the same location. Moreover, futures do not pay dividends which is a clear advantage in option quotation.

2.2 Modeling both returns and volatility

Making multi-step ahead forecast requires modeling both returns and volatility. We follow the methodology developed in Maheu and McCurdy (2011) who develop such a bivariate model so as to be able to derive a forecast of the density of returns several periods ahead by means of simulations. Beyond model-free implied volatilities, we use the model in Maheu and McCurdy (2011) along with the realized volatility estimate, which makes use of intraday data. This original measure developed in Andersen and Bollerslev (1998) follows the early contributions by French et al. (1987) and Hsieh (1991). The estimated realized variance is known to converge to the integrated variance as the sampling frequency tends to infinity (Barndorff-Nielsen and Shephard, 2002).

The two most sensitive assumptions in the Maheu and McCurdy's setting are (1) the cross-equation which permits to link returns with the volatility measure and (2) the assumed distribution of volatility. Following Barndorff-Nielsen and Shephard (2002) and Andersen *et al.* (2003), we retain their first cross-equation where it is assumed that the conditional variance of returns equals the conditional expectation of quadratic variation which is best proxied by the conditional expectation of realized volatility when the sampling interval tends to zero, or:

$$\sigma_t^2 = E_{t-1}(QV_t) = E_{t-1}(RV_t)$$

for a given day t. Voluntarily ignoring the volatility risk premium for reasons given above, we also assume that the investor can use a model-free option-implied volatility (IV) directly provided by the OVX or the MFIV to form her expectation of the conditional volatility. In this case, we have:

$$\sigma_t^2 = E_{t-1}(IV_t)$$

With these three volatility measures (VM) in hand, and assuming log-normality, we then have:

$$E_{t-1}(VM_t) = \exp\left(E_{t-1}\log(VM_t) + \frac{1}{2}\text{Var}_{t-1}(\log(VM_t))\right) \tag{1}$$

Otherwise stated, a forecast of (implied or realized) volatility including all information up to $t-1$ permits to estimate the conditional variance of returns σ_t^2 thereby allowing to model returns simultaneously with volatility.

We thus have first to forecast the point volatility. The HAR specification of Corsi (2009) (following Muller *et al.* (1997)) has been widely used to model realized variance or volatility (see Andersen *et al.* (2007), Liu and Maheu (2009), Maheu and McCurdy (2011), among many others) because it is able to model the strong persistence of the conditional variance. We also rely on this model in the present paper as it can be consistently estimated by OLS and deliver good forecasts of conditional volatility.

We also use the HAR to model the OVX or MFIV dynamics following Fernandes *et al.* (2011) who show that the HAR model also deliver a very competitive representation for the VIX index. In particular, the authors show the ability of the pseudo-long-memory model Ã la Corsi to deal with the persistence in the VIX series. We empirically show that the HAR is also relevant to model the OVX and the MFIV. In particular, we show that estimates are significant for the HAR-OVX and the HAR-MFIV but that these estimates represent different volatility cycles than for the HAR model using realized volatility.

Let us now detail the general form of the bivariate model under consideration.[8] For a given volatility measure VM_t, which could be either the realized volatility or one of the implied volatility, the bivariate model can be written as follows:

$$\begin{aligned} r_t =& \mu + \epsilon_t,\ \epsilon_t = \sigma_t u_t \quad u_t \sim NID(0,1) \\ log(VM_t) =& \omega + \phi_1 log(VM_{t-1}) + \phi_2 log(VM_{t-5,5}) + \phi_3 log(VM_{t-22,22}) \\ &+ \gamma u_{t-1} + \eta v_t \quad v_t \sim NID(0,1) \end{aligned} \tag{2}$$

with r_t the daily return, u_t and v_t unpredictable error terms that are NID(0,1). The estimation of this system plus the cross-equation restriction is done by maximizing the conditional log-likelihood. As in Maheu and McCurdy (2011), we allow for an asymmetric response of volatility to past negative

[8]Further details can be found in Maheu and McCurdy (2011) from which we do not reproduce the integrality of mathematical developments to save space.

returns through γ. The forecasts obtained using the bivariate model are compared with those from the following standard EGARCH model:

$$\begin{aligned} r_t &= \mu + \sigma_t \epsilon_t \\ \log \sigma_t^2 &= \omega + \alpha \epsilon_{t-1} + \theta |\epsilon_{t-1}| + \beta \log \sigma_{t-1}^2 \end{aligned} \tag{3}$$

where $\epsilon_t \sim N(0,1)$. Note that the normality assumption for the error terms has no impact on our results. Using alternative distributions such as the GED, the Student or even a mixture of normals, which is known to accommodate various combinations of skewness and kurtosis, our empirical results only provide a slightly better in-sample fit which does not translate in a better out-of-sample density forecast. Consequently, we only present the results for the Gaussian case.

2.3 Evaluating density forecasts

The methodology developed in Maheu and McCurdy (2011) provides multiperiod density forecasts as “they provide more information to discern among models” (p. 72-73). Indeed, different models may have different performance at different horizons and an analysis of the relative performance of models at various horizons may be of interest to the econometrician as well as for derivatives pricing and risk management activities. In the present paper, multiperiod forecasts are essential because they confirm that models based on implied volatilities dominate at all horizons.

The assess the performance of the four models under consideration, we use the standard DMW test of equal predictive accuracy developed in Diebold and Mariano (1995) and West (1996) as our models are non-nested.[9] The DMW requires a loss function and, as in Maheu and McCurdy (2011) we rely on the predictive likelihood developed in Amisano and Giacomini (2007). For a given predictive density $f_{\mathcal{M},k}(. \mid \Phi_t, \theta)$ of a model $\mathcal{M}$ at a horizon k using information Φ_t up to time t and θ the vector of estimated parameters, the average predictive likelihood will be computed as:

$$\mathcal{L}_{\mathcal{M},k} = \frac{1}{T - \tau - k_{max} + 1} \sum_{t=\tau+k_{max}-k}^{T-k} \log f_{\mathcal{M},k}(r_{t+k} \mid \Phi_t, \theta) \tag{4}$$

[9] When models are nested, Clark and West (2007) suggest to modify the DMW statistic whose asymptotic distribution becomes highly non-standard. In the case of non-nested models, no adjustment is necessary to the standard statistic.

with $T = 1992$ the total number of daily observations in our sample, τ the length of the window for estimation, k_{max} the maximum forecast horizon and r_{t+k} the realized return at the horizon k. For comparability of our results with the existing literature, we select the estimation window τ to be 1200 days and predict up to 60 days. Because $\mathcal{L}_{\mathcal{M},k}$ can be computed for all k from 1 to 60 days, we obtain a term structure of average predictive likelihood which will be helpful to discuss the forecasting performance of the different models investigated in the present study across horizons. The DMW statistic using the predictive likelihood is then given by:

$$t^k_{\mathcal{M},\mathcal{N}} = (\mathcal{L}_{\mathcal{M},k} - \mathcal{L}_{\mathcal{N},k}) \times \left(\frac{(\hat{\sigma}_{\mathcal{M},\mathcal{N},k})}{\sqrt{T - \tau - k_{max} + 1}}\right)^{-1}. \tag{5}$$

Following the literature on economic forecasting, $\hat{\sigma}_{M,N,k}$ is estimated through the heteroscedasticity- and autocorrelation-consistent estimate of the long run variance developed in Newey and West (1987). Amisano and Giacomini (2007) note that their test displays good properties for large enough sample ($n \geq 150$) which makes the test best suited for financial applications in contrast with applications in macroeconomics. This is the case of our empirical application. The main theoretical drawback of the Diebold and Mariano methodology is that it does not consider explicitly the parameter estimation error. Nevertheless, under realistic assumptions, the parameter estimation error vanishes asymptotically.[10]

3 Data

We purchased option data from CBR Reuters and tick (transaction) data from TickData for the period going from October 8, 2001 to October 29, 2009. As such, this paper represents the most extensive study using model-free implied volatility for the WTI futures market. Dealing with intraday data requires appropriate filtering (cleaning) of the data. As is common in this literature (see Andersen et al. (2003)), we remove days with (1) transactions outside the official trading period, (2) transactions with a variation of more than 2% in absolute value compared to the previous transaction and (3) transactions not reported in chronological order. We also remove days with insufficient trading activity. In particular, we do not further consider days with a shortened trading period and days with more than

[10] Corradi and Swanson (2006) survey the evaluation of predictive density and provide interesting insights about the relevant methodology under various assumptions.

ten zero-return. We end with trading days where liquidity is sufficient to ensure that our estimate of realized variance is consistently estimated. The total number of days where we have a sufficient number of intraday returns along with option data is 1992, which will be the length of our sample.

Quotation of the OVX index officially begins July 14, 2008 and it is back-calculated by the CBOE and publicly available from May 10, 2007. Recent studies such as Aboura and Chevallier (2012) use a time period beginning in May 2007 for their empirical work. In our paper, we further back-calculate the OVX index as well as the MFIV following the lines of Andersen and Bondarenko (2010) to increase the number of observations.[11] As a consequence, our estimates will be statistically more robust. This is an important feature of our study. The sample Pearson coefficient of correlation for the two measures over the full sample is around 90% indicating a strong convergence of the two measures. However, we will show in the empirical Section that the statistical performance of each measure is statistically different for some horizons, thereby motivating our use of both measures.

We compute the realized variance as the sum of the squared 5-minute intraday returns using the last tick method (Wasserfallen and Zimmermann, 1985). We ignore overnight returns which are known to follow a very different dynamics and only use intraday returns that cover the trading session. The 5-minute sampling interval is shown to be relevant for the analysis of high-frequency WTI crude oil futures prices in Chevallier and SÃ©vi (2012). We also experimented with alternative sampling intervals (1, 2, 10 and 20 minutes) and alternative computation schemes (Zhang *et al.*, 2005) allowing to deal with the microstructure noise detrimental impact (see Hansen and Lunde (2006) for a discussion of this issue and Andersen et al. (2011) for an empirical study of the impact of microstructure noise on volatility forecasts). All these alternative considerations lead to qualitatively and quantitatively similar results.

In Table 1, we report descriptive statistics for daily returns and the three measures of volatility that we use in the paper as well as their log counterparts. We first note the average level of implied volatility which is about 40%. This is line with existing studies on oil volatility. Consistent with the available empirical literature, we find that returns and volatility have a positive excess kurtosis. The returns' skewness is slightly below zero, whereas it is strongly positive in the cases of both realized volatility

[11] We refer the interested reader to the papers of Andersen and Bondarenko (2007) and Andersen *et al.* (2010) for a short introduction to the computation of the VIX along with other model-free implied volatility measures. The excellent survey by Christoffersen *et al.* (2012) gives in its Table 2 a list of volatility indices computed throughout the world. Most of them follow the VIX methodology. The official web site of the CBOE also provides some details about the VIX calculation (http://www.cboe.com/micro/vix/vixwhite.pdf).

and the implied volatilities. As demonstrated in Andersen *et al.* (2007), the logarithm transform of the selected volatility measures is much closer to a Gaussian distribution in light of the values of the third and fourth moments. On top of that, we observe that the OVX and the annualized square root of the realized volatility have a very similar standard deviation, despite the difference in their average. The OVX is on average higher than the realized volatility, expressing the net market demand for hedging. See Chernov (2007) for a formal approach of this variance risk premium.

All four series are plotted in Figure 1. Daily returns exhibits strong heteroscedasticity in accordance with the significant excess kurtosis reported in Table 1. Large daily changes mainly occur in the end of 2001, following the 9/11 event, and in the end of 2008-beginning of 2009 period following the large drop in crude oil prices from a high $147.27 dollars per baril to a low $33.85 dollars per barrel. The OVX and MFIV series are transformed into daily volatility measures using the 252 trading days convention. As for the realized volatility, recall that we exclude the overnight returns but that density forecasts computed using overnight returns have similar performance in the out-of-sample exercise.[12] The same events than those mentioned for returns can explain the high increase in volatility for the three measures while a large jump in volatility is also visible in the early 2003 due to a large drop in oil price. Overall, model-free implied volatilities are less erratic than realized volatility is.

4 Empirical results

4.1 Estimation results

As mentioned above, the parameters of the EGARCH and HAR models are estimated using a rolling window scheme. West (2006) suggest using rolling window "[...] when one wishes to guard against moment or parameter drift that is difficult to model explicitly."(p. 107) The window size is chosen to be 1260 as in Maheu and McCurdy (2011) making the first predictive density estimated to be available for November, 22th of 2006.

The number of estimation windows is given by $T - \tau - k_{max} + 1 = 678$. We report in Table 2 the sample statistics for the estimated parameters of the EGARCH model. Estimates are broadly in line with the literature, as volatility is perceived as persistent – with β being slightly lower than 1 – and subject to

[12] Results are available upon request and not reported to save space.

leverage effects, as α is negative and different from zero (its 95% quantile is equal to -0.037).

Table 3 reports similar sample statistics for the HAR-RV model. Again, given the empirical quantiles obtained throughout the rolling estimation, every parameter driving the volatility process is statistically different from zero. We observe that the parameters for the one-day, one-week and one-month average log realized volatility are positive and highly significant. Also, a leverage effect seems to be present. As in the EGARCH case, γ is negative, indicating the existence of a leverage effect. When investigating the contributions of each volatility component in the HAR model, the five days average component clearly dominates the others. The 22 days average comes second and the most recent data have the weakest influence. These results are in line with those presented in Andersen *et al.* (2007) or Corsi (2009) among others, but qualitatively different as the most recent data often have a considerable influence for stock indexes or exchange rates in existing empirical studies.

In the case of the OVX index – as presented in Table 4 – it appears that the leverage effect is weaker than it is for the case of realized volatility. Also, the shortest component of the HAR model has the highest contribution in the dynamics of the VIX. The 5 days average has a weaker influence and the 22 days one seems to have a very limited impact on the future of volatility. These results confirm the empirical findings in Fernandes *et al.* (2011) and support our decision to consider the HAR as a time-series model for the OVX dynamics. Finally, from Table 5, we note that empirical estimates for the HAR model using the MFIV are very similar to the OVX' estimates except that the weight of the one-day lagged volatility is even larger in modeling the current level of MFIV. Importantly, there is no evidence of a leverage effect for implied measures of volatility.

In summary, volatility cycles are very different for the realized volatility and the implied volatility measures as can be seen for the estimates of the HAR components. Moreover, a different leverage effect is diagnosed for the EGARCH and the realized volatility dynamics, which even disappears when implied volatilities are used. All in all, our results point to the fact that the two implied volatility measures can be adequately modeled in a HAR framework, and have quite similar dynamics, thanks to the HAR components analysis.

4.2 Comparing the accuracy of density forecasts

The density forecasts are computed using the rolling estimation scheme that is similar to Maheu and McCurdy (2011). The forecast of the density of returns is provided for all horizons going from 1 to 60 days. We plot the average predictive likelihood computed from Eq. (4) for each measure of volatility and for all horizons in Figure 2. As expected, the predictive likelihood decreases on average when the forecast horizon is longer. As such, our results are conform to the bulk of the literature (see Christoffersen and Diebold (2000) among others) as the density forecast accuracy is better for shorter horizons. From this first graphical representation of our results, it appears that density predictions using option-implied volatilities as an input are better than those using information from historical series of returns, either daily or intradaily.

The graphical comparison of average predictive likelihoods for all four models is not sufficient to deliver statistical evidence of the superiority of one model over another. Hence, for each pair of models and each horizon we compute the DMW statistic given in Eq. (5) to gauge the relative forecasting performance of both models. The DMW for two sets of pairs of models are plotted in Figure 3 and 4.

Figure 3 reports the DMW statistic when then EGARCH model estimated using daily data is involved. We observe that the EGARCH model never delivers statistically superior density forecast except at the anecdotal 9-day horizon when the EGARCH is compared to the RV model. After 24 days (one trading month), the EGARCH model is statistically inferior to the RV model. As for the models using implied volatilities as volatility measures, Figure 3 provides evidence of the their good performance with respect to the EGARCH model, except in the very short-term (1-5 days) where the EGARCH model delivers density forecasts that are as good as for the implied volatility models in a statistical sense. Overall, evidence from Figure 3 highlights the low performance of the EGARCH model relatively to the RV or implied volatility models. Our conclusion for the EGARCH-RV pair confirms the findings in Maheu and McCurdy (2011) that intraday data may be valuable in forecasting the future density of returns. However, because models based on implied volatilities also dominate the EGARCH specification, we are left with the interesting question of the relative performance of the RV model with respect to the OVX or MFIV model. In particular, is the good performance of non-EGARCH model coming from the use of high-frequency data or from a measure of volatility that include forward-looking information?

Figure 4 provides a clear response to this question. Indeed, at all horizons, the models based on implied volatility measures outperform the model based on intraday data. As Maheu and McCurdy (2011) do not consider the implied volatility in their analysis, we cannot compare our empirical finding for WTI crude oil futures with an existing result for stock or bond markets. However, the result is particularly interesting as the forward-looking aspect of the OVX or MFIV translates into a significant improvement in density forecasting. Moreover, we can think of our results as a lower bound to the potential improvement coming from implied measures as we only consider the volatility index and not the full information set available in the numerous option quoted prices. Finally, note also from Figure 4 that predictions from the OVX dominate those from the MFIV for horizons longer than 35 days (one month and a half). Because the MFIV makes use of the full set of available options and that options that are far from moneyness are more prone to measurement errors, we might think that the better performance of the OVX is related to its use of more liquid options only.[13]

For completeness, some numerical results are reported in Table 6 where we further observe the superiority of models based on implied volatility. Importantly, the significance level of our results is high with DMW statistics often higher than 4 in absolute value. This indicates that our results are not likely to be driven by small sample effects.

To further put our results in perspective, recall that Blair *et al.* (2001) investigate the contribution of high-frequency data for forecasting the volatility of the S&P 100 stock index and show that in the short term, tick-by-tick data help in forecasting the volatility beyond the information already incorporated in the volatility index but that overall; "The evidence for incremental forecasting information in intraday returns is insignificant."(p. 5) Our results corroborates this finding in the case of density forecast. Also for the S&P 100 stock index, Koopman *et al.* (2005) show that realized volatility has very good forecast performance in comparison with standard GARCH or models based on implied volatility at horizons shorter than one week. We provide evidence that in the case of crude oil, models based on implied volatility outperforms models based on realized volatility for all horizons including the very short term.

[13] The MFIV is an aggregated index as is the OVX. The fact that the MFIV does use the full set of available option prices should not be confounded with the estimation of a risk-neutral density as described in the Introduction Section.

5 Conclusion

This paper makes use of implied volatility indexes to compute density predictions for the WTI crude oil futures and compares these predictions to density forecasts based on realized volatility and the EGARCH model. Models using implied volatility under the form of an index significantly outperform their counterpart based on historical data. This finding highlights the information content of option implied volatility that is forward-looking by nature.

As for the implications of our analysis, Christoffersen and Diebold (2000) note that, "if volatility fluctuates in a forecastable way, volatility forecasts are useful for risk management" (p. 12). The same conclusion applies to the forecastability of the density of returns. This feature might then translate into an improvement of quantile estimation and improve, for instance, value-at-risk estimates and more generally risk management. Such an economic analysis of the potential gain associated with better prediction of the density of returns may be of interest for the energy finance community and is left for future work.

Tables

WTI Oil futures	Returns	RV(5min)	OilVIX	MFV	Log RV(5min)	Log OilVIX	Log MFV
Mean	0.155	30.184	41.339	40.332	3.338	3.680	3.655
Standard Dev.	0.397	13.207	13.456	13.315	0.354	0.275	0.275
Skewness	-0.173	3.347	1.924	2.060	0.777	1.082	1.169
Excess Kurtosis	3.142	28.457	4.082	4.767	1.221	1.065	1.381
5% quantile	-0.039	17.172	27.977	27.480	2.843	3.331	3.314
95% quantile	0,039	55.938	72.057	71.297	4.024	4.277	4.267
Observations	1992						

Table 1
Summary Statistics

EGARCH	ω	α	θ	β
Average	-0.528	-0.067	0.097	0.942
Standard Dev.	0.302	0.026	0.014	0.039
Skewness	-0.350	-0.703	1.140	-0.365
Kurtosis	-0.724	-0.447	13.056	-0.697
5% quantile	-0.918	-0.119	0.073	0.891
95% quantile	-0.132	-0.037	0.113	0.993

Table 2
Sample statistics for the estimates of the EGARCH model over all windows for out-of-sample forecast.

HAR-RV-Gaussian	ω	ϕ_1	ϕ_2	ϕ_3	γ	μ	η
Average	-1.316	0.090	0.446	0.308	-0.037	0.001	0.454
Standard Dev.	0.625	0.012	0.057	0.024	0.003	0.000	0.013
Skewness	0.741	-0.887	0.159	0.506	0.016	-0.312	0.099
Kurtosis	-1.150	0.415	-1.219	-1.099	-0.953	-0.403	-1.478
5% quantile	-1.969	0.063	0.360	0.279	-0.042	0.000	0.436
95% quantile	-0.303	0.104	0.528	0.348	-0.033	0.001	0.474

Table 3
Sample statistics for the parameters of the estimated HAR-RV over all 678 windows for out-of-sample forecast.

HAR-RV-Gaussian OilVIX	ω	ϕ_1	ϕ_2	ϕ_3	γ	μ	η
Average	-0.135	0.854	0.095	0.033	-0.002	0.001	0.084
Standard Dev.	0.083	0.027	0.017	0.017	0.001	0.000	0.002
Skewness	0.341	0.057	-0.202	0.133	0.401	-0.311	0.475
Kurtosis	-0.635	-1.305	-0.830	-1.609	-0.817	-0.407	-1.305
5% quantile	-0.241	0.816	0.067	0.012	-0.003	0.000	0.082
95% quantile	0.003	0.891	0.123	0.058	0.000	0.001	0.087

Table 4
Sample statistics for the parameters of the estimated HAR-OVX over all windows for out-of-sample forecast.

HAR-RV-Gaussian MFIV	ω	ϕ_1	ϕ_2	ϕ_3	γ	μ	η
Average	-0.128	0.873	0.080	0.030	0.002	0.001	0.083
Standard Dev.	0.080	0.019	0.020	0.020	0.001	0.000	0.003
Skewness	0.501	-0.236	-0.102	0.161	0.567	-0.300	0.086
Kurtosis	-0.709	-0.756	-0.537	-1.685	0.260	-0.440	-1.342
5% quantile	-0.217	0.841	0.044	0.005	0.001	0.000	0.079
95% quantile	0.012	0.903	0.110	0.058	0.004	0.001	0.087

Table 5
Sample statistics for the parameters of the estimated HAR-MFIV over all windows for out-of-sample forecast.

Models	RV(5min)	OilVIX	MFIV
horizon = 5 days			
EGARCH	1.526	-2.023	-2.737
RV(5min)		-2.531	-3.218
OilVIX			-2.210
horizon = 10 days			
EGARCH	0.824	-3.243	-3.814
RV(5min)		-3.172	-3.703
OilVIX			-1.420
horizon = 30 days			
EGARCH	-3.693	-4.665	-4.999
RV(5min)		-3.779	-4.099
OilVIX			0.640
horizon = 60 days			
EGARCH	-4.204	-4.862	-4.932
RV(5min)		-4.737	-4.886
OilVIX			3.374

Table 6
Average Diebold-Mariano statistics using the predictive likelihood as in Amisano and Giacomini (2007). The pairwise statistic goes from 1 day to 60 days. A positive statistic indicates that the first model is superior to the second model. The asymptotic distribution of the statistic is standard normal.

Figures

Figure 1
The figure plots the daily log-returns (top-left panel), the daily realized volatility (square-root of the daily realized variance) (top-right panel), the back-calculated OVX (bottom-left panel) and the model-free implied volatility (bottom-right panel) for the period October 8, 2001 to October 29, 2009 (1992 observations). All volatility measures are in annualized terms.

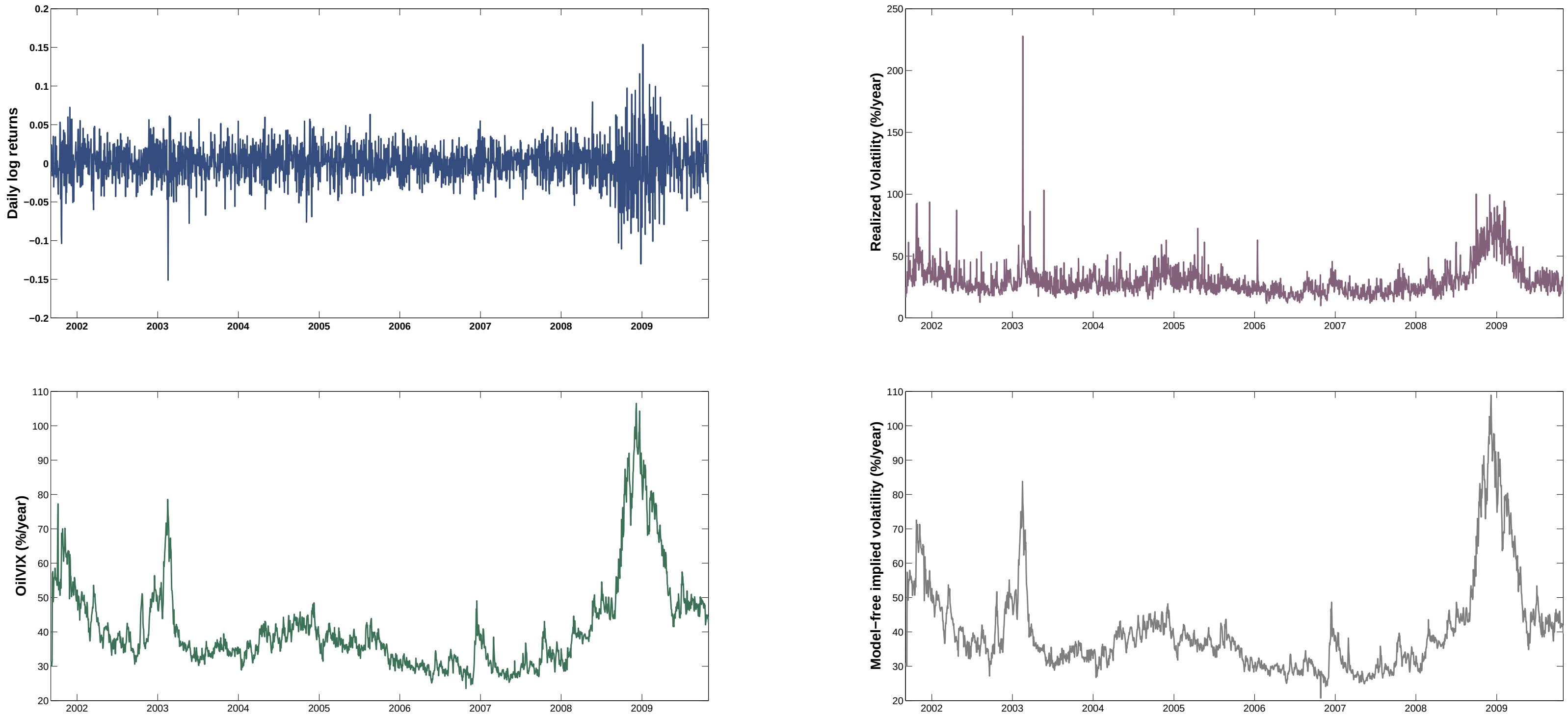


Figure 2
Average predictive likelihoods for horizons going from 1 to 60 days.

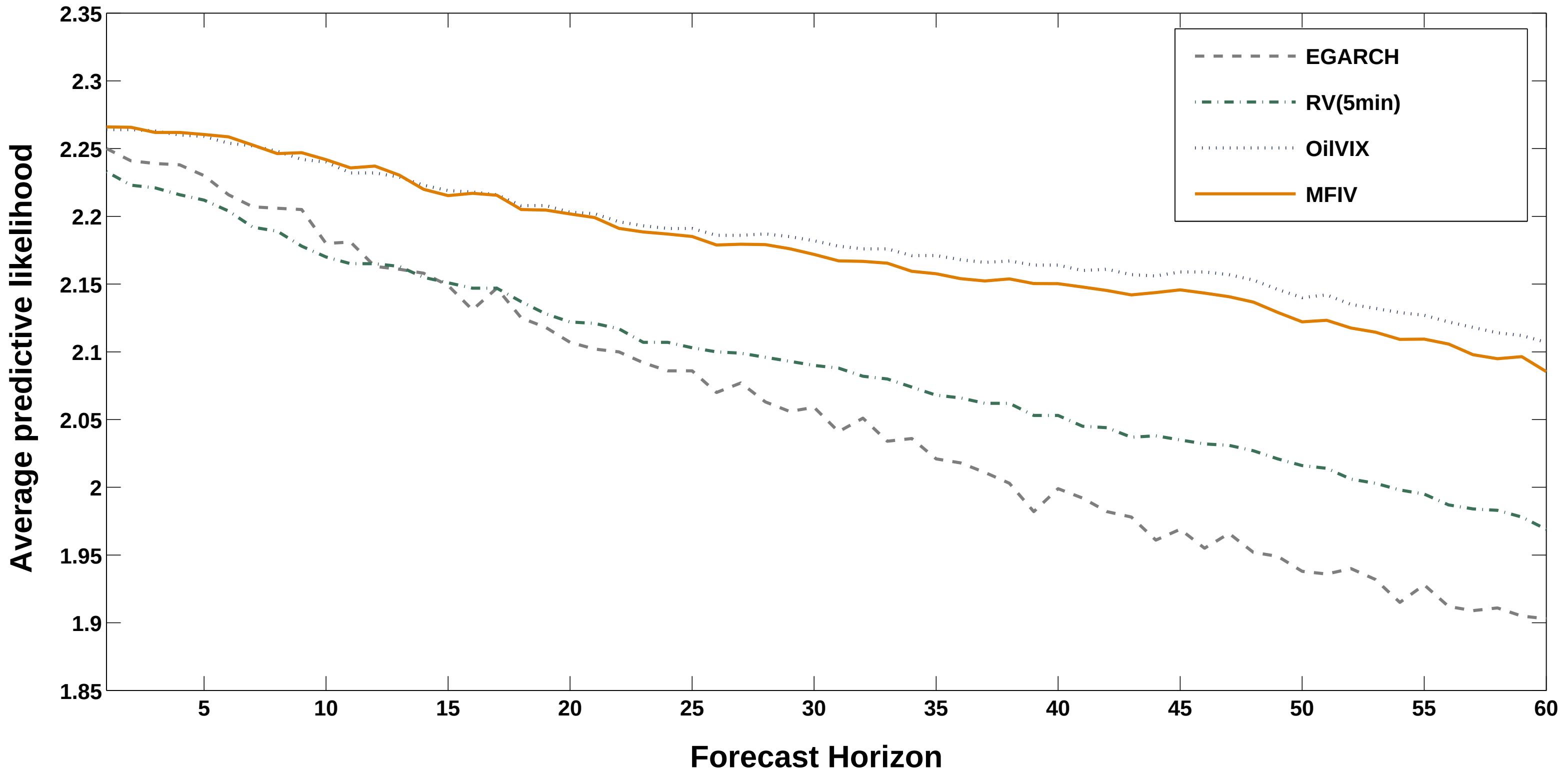


Figure 3
Diebold-Mariano test statistics of the difference between average predictive likelihoods between EGARCH and alternative models

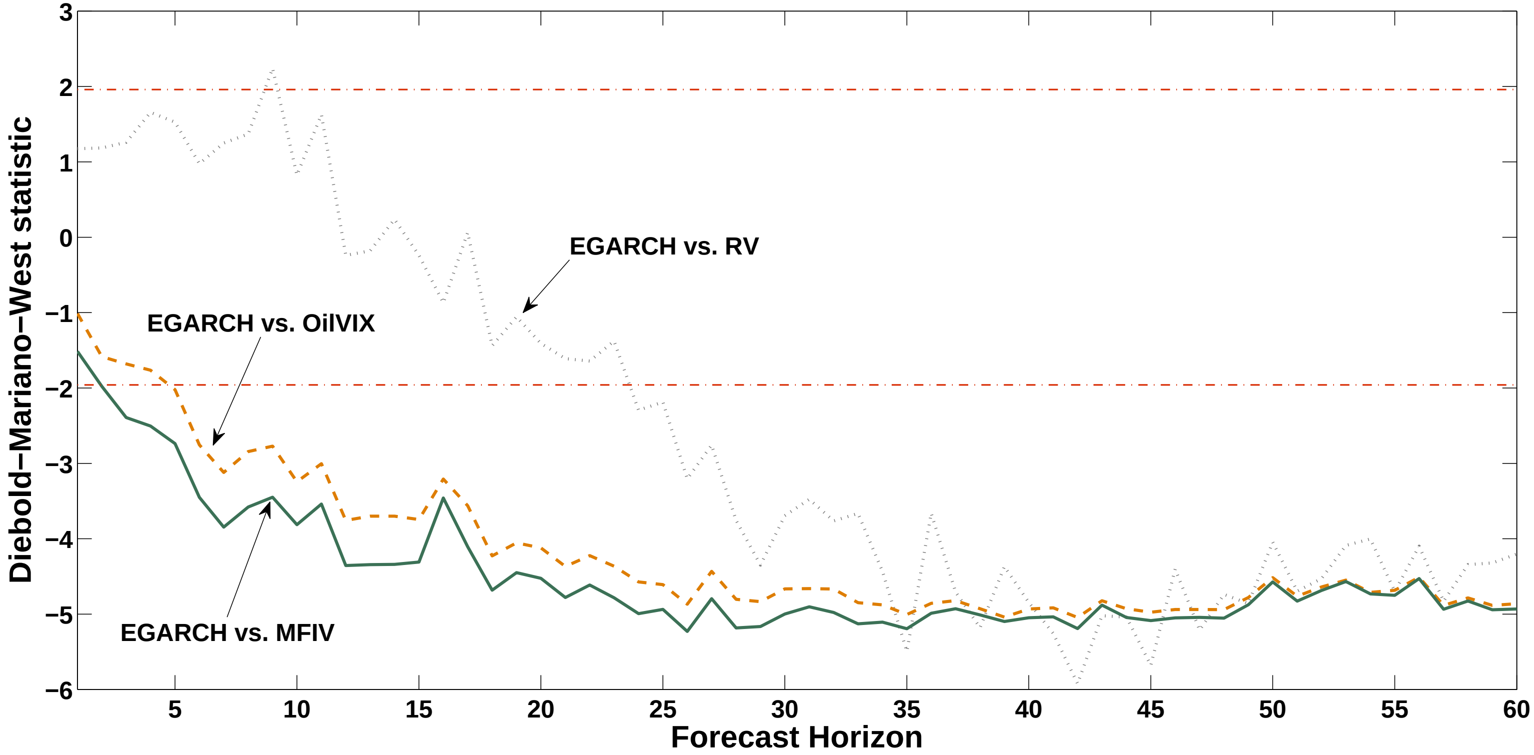
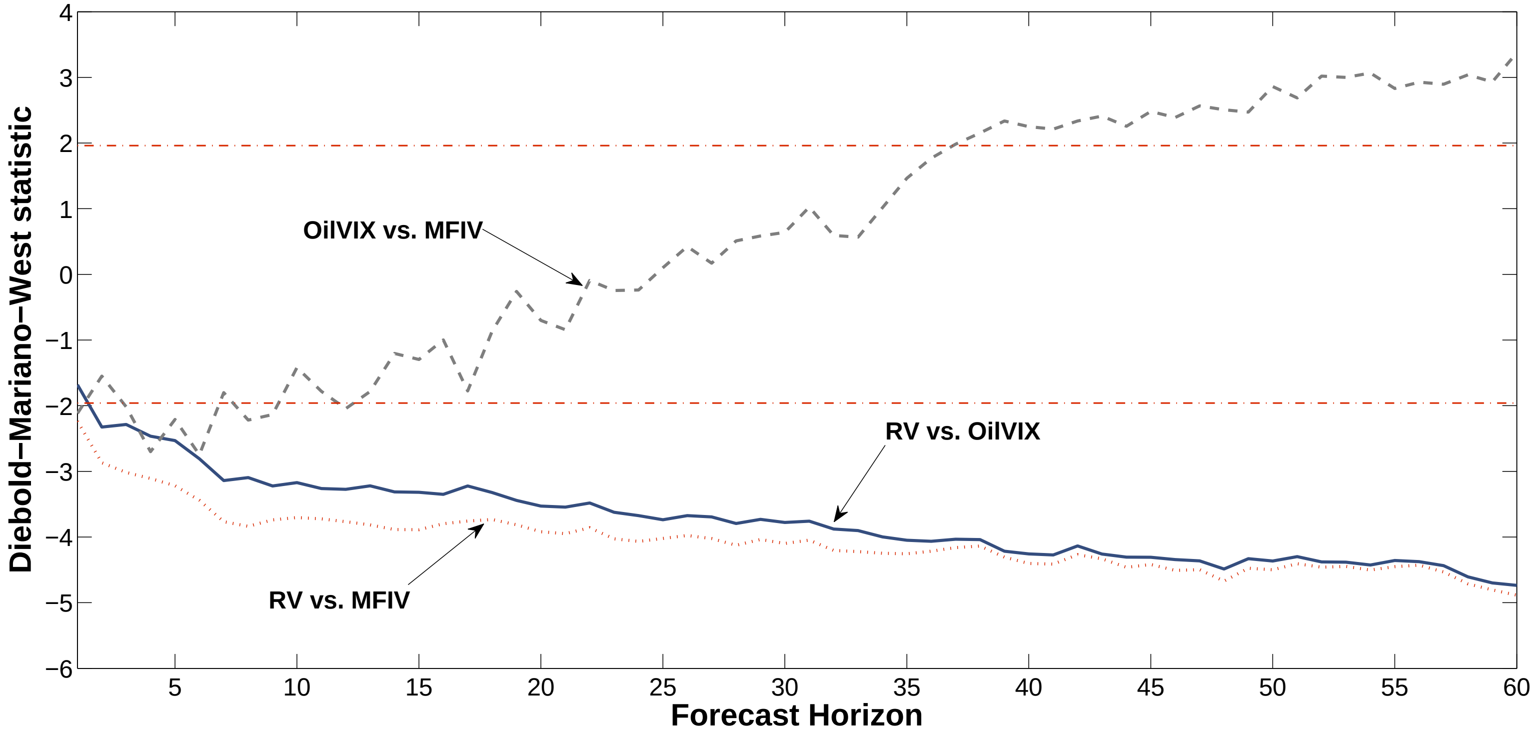


Figure 4
Diebold-Mariano test statistics of the difference between average predictive likelihoods between pairs of models



References

ABOURA, S., CHEVALLIER, J., 2013. Leverage vs. feedback: Which effect drives the oil market? *Finance Research Letters* 10, 131-141.

AGNOLUCCI, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics* 31, 316-321.

AMISANO, G., GIACOMINI, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics* 25, 177-190.

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

ANDERSEN, T.G., BOLLERSLEV, T., CHRISTOFFERSEN, P.F., DIEBOLD, F.X., 2006. Volatility and correlation forecasting. In: G. Elliott, C.W.J. Granger and A. Timmermann, *Handbook of Economic Forecasting* (Chap. 15), vol. 2, Elsevier.

ANDERSEN, T.G., BOLLERSLEV, T., DIEBOLD, F.X., 2007. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* 89, 701-720.

ANDERSEN, T.G., BOLLERSLEV, T., DIEBOLD, F.X., LABYS, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579-625.

ANDERSEN, T. G., BOLLERSLEV, T., MEDDAHI, N., 2011. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* 160, 220-234.

ANDERSEN, T.G., BONDARENKO, O., 2007. Construction and interpretation of model-free implied volatility. NBER Working Paper no. 13449, Cambridge, MA.

ANDERSEN, T.G., BONDARENKO, O., 2010. Dissecting the pricing of equity index volatility. Unpublished manuscript, Northwestern University and University of Illinois at Chicago.

BARNDORFF-NIELSEN, O., SHEPHARD, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253-280.

BARONE-ADESI, G., WHALEY, R.E., 1987 . Efficient analytical approximation of American option values. *Journal of Finance* 42, 301-320.

BECKER, R., CLEMENTS, A., WHITE, S., 2007. Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking and Finance* 31, 2535-2549.

BLAIR, B.J., POON, S., 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.

BLISS, R., PANIGIRTZOGLOU, N., 2004. Option-implied risk aversion estimates. *Journal of Finance* 59, 407-446.

BRITTEN-JONES, M., NEUBERGER, A., 2000. Option prices, implied price processes, and stochastic volatility. *Journal of Finance* 55, 839-866.

CABEDO, J.D., MOYA, I., 2003. Estimating oil price 'value at risk' using the historical simulation approach. *Energy Economics* 25, 239-253.

CANINA, L., FIGLEWSKI, S., 1993. The informational content of implied volatility. *Review of Financial Studies* 6, 659-681.

CARR, P., WU, L., 2009. Variance risk premiums. *Review of Financial Studies* 22, 1311-1341.

CHEVALLIER, J., SÉVI, B., 2012. On the volatility-volume relationship in energy futures markets using intraday data. *Energy Economics* 34, 1896-1909.

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

CHRISTOFFERSEN, P.F., DIEBOLD, F.X., 2000. How relevant is volatility forecasting for financial risk management? *Review of Economics and Statistics* 82, 12-22.

CHRISTOFFERSEN, P.F., JACOBS, K., CHANG, B.Y., 2012. Forecasting with option implied information In: G. Elliot and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Elsevier, forthcoming.

CLARK, T., WEST, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138, 291-311.

CORRADI, V., SWANSON, N.R., 2006. Predictive density evaluation. In: G. Elliot, C.W.J. Granger and A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Chap. 5), vol. 1, Elsevier.

COVRIG, V., LOW, B.S., 2003. The quality of volatility traded on the over-the-counter currency market: a multiple horizons study. *Journal of Futures Markets* 23, 261-285.

CORSI, F., 2009. A simple approximate long memory model of realized volatility. *Journal of Financial Econometrics* 7, 174-196.

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

DAY, T.E., LEWIS, C.M., 1993. Forecasting futures market volatility. *Journal of Derivatives* 1, 33-50.

DIEBOLD, F.X., MARIANO, R.S., 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13, 253-263.

FACKLER, P.L., KING, R.P., 1990. Calibration of option-based probability assessments in agricultural commodity markets. *American Journal of Agricultural Economics* 72, 73-83.

FERNANDES, M., MEDEIROS, M.C., SCHARTH, M., 2011. Modeling and predicting the CBOE market volatility index. Unpublished manuscript.

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

FRENCH, K.R., SCHWERT, G.W., STAMBAUGH, R.F., 1987. Expected stock return and volatility. *Journal of Financial Economics* 19, 3-29.

GIOT, P., LAURENT, S., 2003. Market risk in commodity markets: a VaR approach. *Energy Economics* 25, 435-457.

GIOT, P., LAURENT, S., 2004. Modelling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11, 379-398.

GIOT, P., LAURENT, S., 2007. The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets* 27, 337-359.

HAN, H., PARK, M.D., 2013. Comparison of realized measure and implied volatility in forecasting volatility. *Journal of Forecasting*, forthcoming.

HANSEN, P.R., LUNDE, A., 2006. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics* 24, 127-218.

HARVEY, C.R., WHALEY, R.E., 1992. Market volatility prediction and the efficiency of the S&P 100 index option market. *Journal of Financial Economics* 31, 43-73.

HØG, E., TSIARAS, L., 2011. Density forecasts of crude oil prices using option-implied and ARCH-type models. *Journal of Futures Markets* 31, 727-754.

HONG, H., YOGO, M., 2012. What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics* 105, 473-490.

HSIEH, D.A., 1991. Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance* 46, 1839-1877.

HUANG, D., YU, B., FABOZZI, F.J., FUKUSHIMA, M., 2009. CAViaR-based forecast for oil price risk. *Energy Economics* 31, 511-518.

JIANG, G.J., TIAN, Y.S., 2005. The model-free implied volatility and its information content. *Review of Financial Studies* 18, 1305-1342.

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

KANG, S.H., KANG, S.-M., YOON, S.-M., 2009. Forecasting volatility of crude oil markets. *Energy Economics* 31, 119-125.

KANG, S.-H., YOON, S.-M., 2013. Modelling and forecasting the volatility of petroleum futures prices. *Energy Economics* 36, 354-362.

KOOPMAN, S. J., JUNGBACKER, B., HOL, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445-475.

KRONER, K.F., KNEAFSEY, K.P., CLAESSENS, S., 1995. Forecasting volatility in commodity markets. *Journal of Forecasting* 14, 77-95.

LIU, C, MAHEU, J.M., 2009. Forecasting realized volatility: A Bayesian model-averaging approach. *Journal of Applied Econometrics* 24, 709-733.

MAHEU, J.M., MCCURDY, T.H., 2011. Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics* 160, 69-76.

MARTENS, M., ZEIN, J., 2004. Predicting financial volatility: high-frequency time-series forecasts vis-Ã -vis implied volatility. *Journal of Futures Markets* 24, 1005-1028.

MELICK, W.R., THOMAS, C.P., 1997. Recovering an asset's implied pdf from option prices: an application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis* 32, 91-115.

MOHAMMADI, H., SU, L., 2010. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics* 32, 1001-1008.

MÜLLER, U.A., DACOROGNA, M.M., DAVÃ©, R.D., OLSEN, R.B., PICTET, O.V., 1997. Volatilities of different time resolutions – Analyzing the dynamics of market components. *Journal of Empirical Finance* 4, 213-239.

POON, S.-H., GRANGER, C.W.J., 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41, 478-539.

POTESHMAN, A.M., 2000. Forecasting future volatility from option prices. Working Paper, University of Illinois at Urbana-Champaign.

SÉVI., B., 2013. Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, forthcoming.

TANG, K., XIONG, W., 2012. Index investment and financialization of commodities. *Financial Analysts Journal* 68, 54-74.

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

TROLLE, A., SCHWARTZ, E.S., 2010. Variance Risk Premia in Energy Commodities. *Journal of Derivatives* 18, 1-18.

WASSERFALLEN, W., ZIMMERMANN, H., 1985. The behavior of intraday exchange rates. *Journal of Banking and Finance* 9, 55-72.

WEI, Y., WANG, Y., HUANG, D., 2010. Forecasting crude oil market volatility: further evidence using GARCH-class models. *Energy Economics* 32, 1477-1484.

WEST, K.D., 1996. Asymptotic inference about predictive ability. *Econometrica* 64, 1067-1084.

WEST, K.D., 2006. Forecast Evaluation. In: G. Elliott, C.W.J. Granger and A. Timmermann, *Handbook of Economic Forecasting* (Chap. 3), vol. 1, Elsevier.

ZHANG, L., MYKLAND, P.A., AÃ¯T-SAHALIA, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association* 100, 1394-1411.

ZIEGLER, A., 2007. Why does implied risk aversion smile? *Review of Financial Studies* 20, 859-904.