



Centre de Recherche en Economie et Droit de l'Energie

**FORECASTING THE VOLATILITY  
OF CRUDE OIL FUTURE USING INTRADAY DATA**

**Benoît SEVI**

**Cahier de recherche n° 13.10.104**

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 volatility of crude oil futures using intraday data

**Benoît Sévi\***

Aix-Marseille Université (Aix-Marseille School of Economics), CNRS & EHESS

October 15, 2013

## **Abstract**

We use the information in intraday data to forecast the volatility of crude oil at a horizon of 1 to 66 days using a variety of models relying on the decomposition of realized variance in its positive or negative (semi-variances) part and its continuous or discontinuous part (jumps). We show the importance of these decompositions in predictive regressions using a number of specifications. Nevertheless, an important empirical finding comes from an out-of-sample analysis which unambiguously shows the limited interest of considering these components. Overall, our results indicates that a simple autoregressive specification mimicking long memory and using past realized variances as predictors does not perform significantly worse than more sophisticated models which include the various components of realized variance.

*JEL Classification:* C22, C53, G13 ,Q4

*Keywords:* volatility forecasting, crude oil futures, realized variance, jumps, realized semivariance.

---

\*GREQAM, Aix-Marseille Université, Château La Farge - Route des Milles, 13290 Les Milles Aix-en-Provence, FRANCE, Email: benoit.sevi@gmail.com

# 1 Introduction

Recent literature suggests that high-frequency data are useful for predicting future volatility. Moreover, the decomposition between the continuous and the jump component, on the one hand, and negative and positive intraday returns, on the other hand, may also help to obtain a better accuracy of forecasts. Finally, considering a possible leverage effect might also improve the forecast of volatility, as far as the market of interest exhibits this particular feature. In the present study, we simultaneously consider all these characteristics to provide a large-scale empirical analysis of the forecasting accuracy of various time series models derived from the innovative HAR (Heterogenous Autoregressive) specification proposed in Corsi (2009). The eleven models under consideration deliver forecasts of the cumulative volatility from 1 to 66 days (three months considering only trading days). The analysis is conducted in-sample, and more importantly, out-of-sample for the front-month WTI futures contract quoted on the NYMEX-CME over the period 1987-2010, totalizing 5732 trading days. Our results indicate that considering the various components of the realized variance only contributes to improve the in-sample fit but does not represent a significant improvement in an out-of-sample framework.

There is ample evidence in the econometric and mathematical finance literatures of an economic value in forecasting volatility using intraday data. Important contributions in this field are Fleming et al. (2003), who demonstrate the importance of intraday data for portfolio choice in a three-asset framework (see also Bandi et al. (2008)). Competitive forecasts of future volatility also have direct applications in risk management activities. Giot and Laurent (2004), and more recently Clements et al. (2008), make use of realized variance for VaR computation. More generally, better forecasts of volatility can help to ameliorate the density forecast using, for instance, the recent methodology developed in Maheu and McCurdy (2011) possibly with asymmetric features as in Hua and Zhang (2008). In addition, most of the stochastic processes used in modeling oil returns (see Geman, 2005) rely on a stochastic volatility hypothesis (see Wong and Lo, 2009) which necessitates a precise estimation of the volatility process that could be improved using tick-by-tick data. Finally, intraday data also find interesting applications in the field of option pricing. Intraday data can help to improve option pricing using the information content of these data as daily data have been used in the past to price option using GARCH models (see Duan (1995) and Heston and Nandi (2000)). For instance, the recent option pricing model in Stentoft (2008) also relies on intraday data to price derivatives with an increasing precision and Yang et al. (2008) highlight the information content of intraday data when the economic metric of option trading profits is used. Corsi et al. (2013) show that intraday data significantly improve the pricing of options using an HAR-type model.

In this paper, we analyze the forecasting accuracy of several time series models including jumps and/or semi-variance measures that are computed from intraday transaction data. Jumps have a long history in finance but have generally been estimated from daily data.<sup>1</sup> We use the asymptotic theory developed in Barndorff-Nielsen and Shephard (BNS hereafter) (2004) to detect jumps in high frequency returns. Semivariance measures, as part of the larger class of downside risk measures, also have a long history in finance and operational research. Early contributions by Choobineh and Branting (1986) and Josephy and Aczel (1993) propose optimal estimators for semivariance. Ang et al. (2006) consider semivariance in an asset pricing model to illustrate

---

<sup>1</sup>Among major contributions see Andersen et al. (2002), Pan (2002) Chernov et al. (2003), Eraker et al. (2003) and Eraker (2004). Collectively taken, these papers emphasize the importance of jumps for asset pricing.

its relevancy in pricing financial assets. Huang (2008) relies on semivariance in a portfolio choice exercise in the same vein as Markowitz (1959) (see also Markowitz (2013) for a recent update of the state-of-the-art in modern portfolio theory). Interestingly, Grootveld and Hallerbach (1999) call into question the empirical difference between considering variance or semivariance for portfolio selection. In the present paper, we consider realized semivariances that are the high-frequency counterpart of commonly-used semivariances estimated from daily data. Further note that we only compare models that use high-frequency data as their superiority is now well-established in the financial forecasting literature after Andersen et al. (2003).<sup>2</sup> Note also, that we keep the issue of comparing our results with forecasts from implied volatility for future work.<sup>3</sup> Some results on this topic are presented in Martens and Zein (2004) for the case of crude oil and Koopman et al. (2005) for other asset classes, but these two papers do not consider the various components of the realized variance as we do.

There are several motivations behind our present study. First, we assess the forecast of volatility for the most traded commodity around the world and we do so using the most traded commodity futures contract. Second, our results have implications for the management of oil contingent claims as well as for risk management activities (see Giot and Laurent (2003) and Cabedo and Moya (2003)) and portfolio selection where commodities play an increasingly important role in recent years. Third, we investigate the importance of disentangling jumps, that are present in oil futures returns, from the continuous component in line with Andersen et al. (2007) (ABD hereafter) for traditional assets and Tseng et al. (2009) for oil, as jumps are indeed present in oil price data. This is an important issue as jumps are perceived very differently ('bad volatility') from the Brownian part ('good volatility') and their forecasting power is still under intense debate (see Corsi et al. (2010)). Fourth, we will be able to compare our results for crude oil with those for individual stocks and a stock index that are developed in Patton and Sheppard (2011), thereby shedding more light on the robustness of their results.

From a methodological standpoint, we use HAR-type models along with realized semivariances and signed jumps as in Patton and Sheppard (2011). We compare these models with earlier specifications proposed in Corsi (2009) and ABD. We show that the new models only lead to small improvements for volatility forecasting that are visible in-sample but not out-of-sample. We provide a thorough analysis of the out-of-sample performance of all models (and compare their performance) for all horizons from 1 to 66 days, *i.e.* 1 day to 3 months that are relevant horizons for practitioners. The quality of out-of-sample predictions is assessed using QLIKE which is a robust measure (in the Patton's (2011) sense that will be explained later) when the variable of interest is noisy. The QLIKE measure is used along with the so-called Diebold-Mariano statistic to test for out-of-sample predictability. There are various way to test for volatility forecastability. Bellini and Figà-Talamanca (2005) improve the methodology of Christoffersen and Diebold (2000) to develop run tests. Bellini and Figà-Talamanca (2005) also discuss possible alternative using the recent realized variance concept developed in Andersen et al. (2003). We opt for the QLIKE measure that is highly relevant in our setting.

To provide a sketch of our main results, we note that there seems to be only a small or no improvement in-

---

<sup>2</sup>The superiority of high-frequency-based models is empirically demonstrated in Martens and Zein (2004) for the case of crude oil (and in very many other papers for other assets) and we do not reinvestigate the issue of comparing GARCH forecasts with forecasts from intraday data in the present paper.

<sup>3</sup>For previous contributions see Jorion (1995), Christensen and Prabhala (1998), Fleming (1998) and Day and Lewis (1992) among others. An excellent survey of this question is in Christoffersen et al. (2012).

sample in considering more sophisticated models including jumps and other components. The most important empirical finding is that, out-of-sample, there is no statistical improvement in using these models and the genuine model by Corsi (2009) provides results that are at least as good as other models. The best in-sample fit is for horizons that are between 30 and 40 trading days (*i.e.* one month and a half). While our results cannot be generalized to other time series and only hold for the given data, our out-of-sample evidence seriously calls into question the interest of considering various components of the realized variance for volatility forecasting.

The plan for the paper is as follows. In the next Section, we briefly review the relevant literature. In Section 3, we present the estimator of realized variance and some extensions that can be used to detect jumps or compute semivariances. Section 4 presents the genuine HAR model and its various extensions that will be used to understand the role of each component of the realized variance. We also report in Section 4 the results of our in-sample analysis. We discuss the out-of-sample forecasting performance of all models in Section 5 and conclude in Section 6.

## 2 Related literature

### 2.1 Forecasting the volatility of crude oil

The modeling of the crude oil futures returns volatility has been examined in a number of contributions with, most of the time, daily data and GARCH specifications. Sadorsky (2006) provides a thorough empirical analysis of the modeling and forecasting of WTI crude oil, heating oil #2, unleaded gasoline, and natural gas front-month futures contract over the period 1988-2003 (1990-2003 for natural gas). The author finds evidence of good performances of a number of GARCH-type models that are able to significantly beat the random walk model. The result is expected in light of the well-known persistent behavior of volatility. We improve upon the empirical analysis of Sadorsky (2006) as we use intraday transaction price data that convey much more information than daily (closing price) data.

Sadorsky and McKenzie (2008) extend the Sadorsky's (2006) contribution and investigate the forecast of volatility for different asset classes including WTI front-month futures using daily data along with a number of time-series models. Most promising specifications are the power autoregressive and GARCH models which perform well depending on the forecast horizon under consideration. In particular, power autoregressive models are better in predicting volatility at short horizons while GARCH-type models have superior performance for longer-horizon forecast. The intuition behind this empirical finding is the great persistence of power transformations of absolute returns well-acknowledged since Ding et al. (1993) (see also the more recent paper by Forsberg and Ghysels (2009)). Similar findings are provided in Mohammadi and Su (2010) which shows the superiority of the APARCH model compared with standard GARCH or EGARCH models.

Kang and Yoon (2013) investigates the long-memory properties of three time-series of front-month energy futures contract (the same as in Sadorsky (2006)) volatility relying on a number of long-memory models. As such, they also extend the analysis in Kang et al. (2009).<sup>4</sup> The authors find that their models based on daily

---

<sup>4</sup>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. 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.

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 pairwise comparison. Note that these different studies rely on a very noisy proxy for conditional volatility, namely the squared daily returns. Nomikos and Pouliasis (2011) follow Fong and See (2002), who show that regime shifts are clearly present in the data and dominate GARCH effects, and use Markov switching specifications to model the volatility of crude oil (and derivatives) prices. The model seems to perform quite well in-sample, but the well-known difficulties of Markov-switching models for forecasting purpose raises the issue of the potential out-of-sample performance of these specifications.

Other noteworthy contributions use option-implied information from options on WTI futures.<sup>5</sup> An early reference is Kroner et al. (1995) where daily data on futures and option prices are used to estimate implied standard deviations and GARCH models. The authors find that a combination of implied volatility and historical information outperforms predictions from a single source. Agnolucci (2009) provides an analysis of volatility forecasting also using the implied volatility as a predictor compared to GARCH forecasts and obtains similar results to Kroner et al. (1995). The main weakness of these contributions is that they extract option-implied volatility using an inversion of the well-known Black-Scholes model that may not be valid for data at hand.<sup>6</sup>

## 2.2 Forecasting volatility using realized variance

The availability of intraday (transaction or quote) data for investment purpose in recent years has led many researchers to develop an interest for these data and conduct academic research with these new tools. Among potential applications, the forecast of volatility is by far the most important domain of application.

Early attempts to model and forecast volatility using intraday data are Taylor and Xu (1997), Andersen and Bollerslev (1997, 1998), Andersen et al. (2001a and b, 2003), Martens (2002), Koopman et al. (2005) and Giot and Laurent (2007). Most of these papers compare volatility forecasts from GARCH models with those from high-frequency data. The literature on forecasting using realized measures is nicely surveyed in Andersen et al. (2006). The main empirical finding is that these models, even based on simple autoregressive structures such as the HAR provide much better results than GARCH-type models. The central improvement appears to come from the fact that GARCH models only use daily data while HAR models use the much larger information that is contained in intraday data.

Intraday data also have become available for futures commodity markets and a few papers have use these data for academic purpose. Wang et al. (2008) provide an early analysis of the unconditional behavior of distributions for oil and gas futures for quite a short period (1995-1999). Their analysis mainly provides evidence of long memory and near-normality of log transform of realized variance. This second feature would lead to a modeling of RV using log transform so as to obtain more normal residuals. We do not rely on the log transformation of our RV measures following Patton and Sheppard (2011) and for reasons developed below. Another use of intraday data for oil futures is in Duong and Kalev (2008) who investigate the maturity (Samuelson) effect in commodity futures markets. For the sake of forecasting, we are only aware of two papers considering

---

<sup>5</sup>As mentioned in the Introduction Section, we do not consider option-implied data in this paper but provide some references for the case of crude oil for the sake of completeness.

<sup>6</sup>In particular, the data-generating-process may include stochastic volatility and/or jumps that are not compatible with Black-Scholes assumptions.

intraday data. Martens and Zein (2004) rely on ARFIMA models and, closest to our study, Tseng et al. (2009) follow a strategy similar to ABD. The authors fit the HAR-CJ model using the realized-range as an alternative to the so-called realized variance. We extend their analysis in a number of directions. First, we use a variety of HAR-type models that have different properties from the standard HAR-CJ specification. Second, we detect jumps using two different estimators and show the importance of using an alternative to the bi-power variation as this last measure is biased in small samples. Third, and most importantly, we develop a large-scale out-of-sample analysis of the various models under consideration which is the relevant criterion for validating a forecasting model.

### 3 Volatility estimation and jump detection

#### 3.1 Estimating volatility using intraday data

In what follows, the main quantity of interest will be the daily variance of oil futures returns that will be estimated using the realized variance. For a given day  $t$ , the realized variance is computed as the sum of squared intraday returns  $r_{t,j}$  at a given sampling frequency  $1/M$ :

$$RV_{t,M} = \sum_{j=1}^M r_{t,j}^2 \quad (1)$$

The estimated realized variance heavily depends on the choice of the  $1/M$  frequency.<sup>7</sup> This aspect will be further discussed in the empirical section. As frequency tends to infinity (if intraday observations are available as often as desired), *i.e.* using *in-fill* asymptotics arguments:

$$RV_{t,M} \rightarrow \int_{t-1}^t \sigma_s^2 ds + \sum_{j=1}^{J(t)} \kappa^2(t_j) \quad \text{in probability.} \quad (2)$$

when the log-prices  $p_t$  consists of a continuous and a pure-jump component.<sup>8</sup> Equation (2) illustrates that volatility, which is by nature a latent variable, is made “observable” by the theory of quadratic variation. The next question is how to disentangle jumps from the continuous component. BNS (2004) first tackle this issue by proposing the bi-power variation (BPV) measure, which is computed as the scaled summation of the product of adjacent absolute returns. Formally, BPV is defined as follows:

$$BPV_{t,M} = \xi_1 \sum_{j=1}^{M-1} |r_{t,j}| |r_{t,j+1}| \quad (3)$$

where  $\xi_p \equiv 2^{p/2} \Gamma(\frac{1/2(p+1)}{\Gamma(1/2)}) = E(|Z|_p)$  denotes the mean of the absolute value of standard normally distributed random variable  $Z$ . The BPV is a consistent estimator of integrated volatility, and allows to decompose the realized variance into its diffusive and non-diffusive parts. As the sampling frequency increases, the presence of jumps should have no impact, because the return representing the jump is multiplied by a non-

<sup>7</sup>In what follows, we assume that  $1/M$  is the sampling frequency for all the estimators that will be used.

<sup>8</sup>We assume that  $p_\tau = \int_0^\tau \mu_s ds + \int_0^\tau \sigma_s dW_s + J_\tau$  with  $\mu$  a locally bounded drift process,  $\sigma$  a strictly positive  $c\tilde{A} d\tilde{A}$  g process and  $J$  a pure-jump process.

jump return which tends to zero asymptotically. This is true in case of rare jumps, when the probability of two consecutive jumps is negligible.

Nevertheless, the BPV has major drawbacks in empirical applications. First, in practice, the sampling frequency is not high enough to eliminate the influence of jumps. Indeed, the adjacent return does not tend to zero and thus multiplying the large (jump) intraday return results in an upward bias of the BPV. Second, the presence of zero-return that are multiplied twice (with the previous and the next intraday return) leads to a downward bias of the BPV. Some alternative jump-robust measures have been proposed in the literature to deal with the aforementioned issues (see Corsi et al. (2010) for a discussion). Among them, the median realized variance (MedRV) from Andersen *et al.* (2012) is very promising, as it has better properties in realistic settings and remains intuitive and easy to implement. The MedRV estimator reads as follows:

$$MedRV_{t,M} = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left( \frac{M}{M-2} \right) \sum_{j=2}^{M-1} \text{med}(|r_{t,j-1}|, |r_{t,j}|, |r_{t,j+1}|)^2 \quad (4)$$

With the MedRV estimator, the impact of jumps completely vanishes except in the case of two consecutive jumps (which is extremely rare at the sampling frequencies used in empirical applications). In addition, the estimator is more robust to the occurrence of zero-returns except when many zero-returns are likely to be consecutive, which is not the case in the very liquid oil futures market. We thus decide to consider the MedRV as a competitive alternative for our analysis and will provide, for the sake of comparison, all our results for both the BPV and the MedRV estimators.

### 3.2 Jump detection

The difference between  $RV$  and a jump-robust estimator such as BPV or MedRV provides, when it is statistically significant, an estimate of the sum of squared jumps  $\sum_{j=1}^{J(t)} \kappa^2(t_j)$  which have occurred during the period under investigation. Note that a small estimated value for a jump component may not actually come from jumps but be a variation due to the continuous path of the stochastic process and the presence of jumps has thus to be formally tested. BNS (2004, 2006) develop such a testing framework using asymptotic theory on realized variance and multipower variations. We use the adjusted jump ratio statistic, which has proven to have reasonable power against several empirically realistic calibrated stochastic volatility jump diffusion models and the best empirical properties in Huang and Tauchen (2005), to detect jumps in our study. The test statistic for day  $t$  reads:

$$ZJ_{BPV}(t, M) = \sqrt{M} \frac{(RV_{t,M} - BPV_{t,M})RV_{t,M}^{-1}}{((\xi_1^{-4} + 2\xi_1^{-2} - 5) \max\{1, TQ_{t,M}BPV_{t,M}^{-2}\})^{1/2}} \quad (5)$$

with  $TQ_{t,M}$  the realized tripower quarticity which writes  $TQ_{t,M} = M\xi_{4/3}^{-3} \sum_{j=1}^{M-2} |r_{t,j}|^{4/3} |r_{t,j+1}|^{4/3} |r_{t,j+2}|^{4/3}$  and converges in probability to the integrated quarticity. The  $ZJ_{BPV}$  statistic follows a standard normal distribution and allows formal testing for the presence of jumps.

A similar test may be developed with alternative jump-robust estimators of realized variance.<sup>9</sup> In particular,

<sup>9</sup>Indeed, several other estimators for identifying jumps in the series, such as QRV (Christensen *et al.*, 2010) estimator, which are shown to be



we are interested in the present paper with the extension of the adjusted jump ratio statistic to the MedRV estimator. Using limit theory in Andersen *et al.* (2012), the test may be adapted to the MedRV estimator as follows:

$$ZJ_{MedRV}(t, M) = \sqrt{M} \frac{(RV_{t,M} - MedRV_{t,M})RV_{t,M}^{-1}}{(0.96 \max\{1, MedRQ_{t,M}MedRV_{t,M}^{-2}\})^{1/2}} \quad (6)$$

with  $MedRQ_{t,M}$  an estimate of the integrated quarticity obtained using the same methodology than for MedRV. Precisely,  $MedRQ_{t,M} = \frac{3\pi}{9\pi+72-52\sqrt{3}} \left(\frac{M}{M-2}\right) \sum_{j=2}^{M-1} \text{med}(|r_{t,j-1}|, |r_{t,j}|, |r_{t,j+1}|)^4$ . Theodossiou and Zikes (2011) show by means of many simulations and empirical analysis that this test has better properties in the presence of jumps of finite or even infinite activity and zero-returns. If disentangling jumps from the continuous component helps in forecasting the volatility of oil futures returns, we are particularly interested in investigating whether the better properties of the test based on MedRV compared to BPV will translate in an improvement in volatility forecasting as well.

When the  $ZJ$  statistic is significant, then the difference between the RV and the BPV or MedRV is too large for being a realization of a Brownian motion and should be considered as a jump. Thus, the jump component for say, the BPV estimator, is defined as:

$$J_{t,\alpha}(M) = [RV_{t,M} - BPV_{t,M}] \times I[ZJ_{BPV}(t, M) > \Phi_\alpha] \quad (7)$$

where  $I[\cdot]$  is the indicator function which identifies the significance of the  $ZJ_{BPV}(t, M)$  statistic in excess of a given critical value of the Gaussian distribution  $\Phi_\alpha$ . The jump component is similarly defined using the MedRV estimator. In our analysis, we work with a significance threshold equal to 0.999 which is a well-accepted value in the literature, but we checked that jump detection thresholds equal to 0.99 or 0.95 provide quite similar estimates of jumps.

The jump component is then defined as:

$$C_{t,\alpha}(M) = BPV_{t,M} \times I[ZJ_{BPV}(t, M) > \Phi_\alpha] + RV_{t,M} \times I[ZJ_{BPV}(t, M) \leq \Phi_\alpha] \quad (8)$$

In words, in case of a jump, the squared jump component equals the difference between RV and BPV and the continuous component equals the BPV. In case of no jump, the jump component is naturally zero and the continuous component equals the RV. This formalization ensures that in both case, the sum of the squared jump component and the continuous component equals the RV.

### 3.3 Semivariances and signed jumps

Patton and Sheppard (2011) show the importance of the realized semivariance developed in Barndorff-Nielsen *et al.* (2010) for volatility forecasting. The intuition behind this estimator is simple and follows the long-

---

more robust in the presence of microstructure noise and zero-returns, may be used. Theodossiou and Zikes (2011) and Dumitru and Urga (2012) provide a very complete treatment of existing jump detection tests as well as their relative performance in case of microstructure noise and/or jumps. In light of the good properties of the MedRV-based test, we focus on this estimator as an alternative to the BPV estimator.

standing tradition of semivariance estimators beginning with Markowitz (1959). The negative realized semivariance estimator is defined as:

$$RSV_{t,M}^- = \sum_{j=1}^M r_{t,j}^2 \times I_{[r_{t,j} < 0]} \quad (9)$$

Similarly, the positive realized semivariance estimator is computed as:

$$RSV_{t,M}^+ = \sum_{j=1}^M r_{t,j}^2 \times I_{[r_{t,j} > 0]} \quad (10)$$

From both these estimators, Patton and Sheppard (2011) define signed jumps as the difference between positive and negative realized semivariances:

$$\Delta J_{t,M} = RSV_{t,M}^+ - RSV_{t,M}^- \quad (11)$$

An important aspect of this approach is that it does not consider the significance of jumps using procedures such as in Huang and Tauchen (2005) or ABD among others. Nevertheless, because the aim of the paper is in investigating forecast accuracy, this is of no importance.

### 3.4 Data and preliminary analysis

This paper uses tick-by-tick (transaction) price data from the NYMEX-CME for the front-month futures contract. Alternative crude oil benchmarks may be considered as in Sadosky (2006), Kang et al. (2009), Kang and Yoon (2003) and Mohammadi and Su (2010), among others, where the authors rely on alternative U.S. markets and/or international crude oil markets. However, the WTI contract strongly remains the most traded commodity futures throughout the world and most of oil-based derivatives are priced with respect to this contract. In addition, the WTI is the reference for most of investors involved in energy commodities and enters in the GSCI and other commodity indices that are widely traded worldwide. The full sample period is from January 1987 to December 2010. After removing days with a shortened trading session or too few transactions, we obtain 5732 daily observations.<sup>10</sup>

Before computing realized variance, three choices have to be made. First, to deal with the so-called microstructure noise issue (see Hansen and Lunde (2006)), kernel-based methods, pre-averaging or sub-sampling may be used. We reproduced all our estimations with the two-times scale estimator in Zhang et al. (2005) and results were found to be similar qualitatively and quantitatively, possibly due to the high liquidity of the oil market for which the microstructure noise is a quite minor problem, so we only report estimates using the genuine realized variance estimator in Eq. (1) in the present paper.

Second, calendar sampling may be replaced with business-time sampling, i.e. sampling every  $n$  transactions rather than every  $s$  seconds. We choose calendar sampling because (i) it is much more used in the literature and (ii) the asymptotic properties of business sampling schemes are clearly less well-known.

---

<sup>10</sup>Our cleaning of the data follows ABD.

Third, a sampling frequency has to be selected. Theory suggests that returns should be computed at the highest possible frequency, so that estimators converge asymptotically towards the true conditional volatility following *in-fill* asymptotics arguments. However, it is well-known since Andersen and Bollerslev (1997, 1998) and Taylor and Xu (1997) that microstructure noise (due to price discreteness, bid-ask spread, non-synchronous trading, etc.)<sup>11</sup> may impact the realized variance estimator at high frequency. To deal with this issue while making our results comparable with the rest of the literature, we follow the 5 minutes 'rule-of-thumb' which is a quite conservative choice in light of the evidence in Chevallerier and SÃ©vi (2012) who show that higher frequencies may be chosen for the highly liquid WTI front-month futures contract.<sup>12</sup> Additionally, recent research by Liu et al. (2012) shows that the standard 5-minute realized variance measures is difficult to beat in forecasting exercises.

The nominal oil price, daily returns and realized variance computed from 5-minute intraday returns are plotted in Figure 2. We observe evidence of heteroscedasticity with clusters of volatility in 1990 (Iraq war), 1996 and the boom and bust of 2008-2009 (note that the volatility is much higher during the decreasing phase during than the increasing one). Sample autocorrelations for the realized variance of crude oil and S&P 500 futures, which is reported for the sake of comparison, are plotted in Figure 3. Interestingly, autocorrelations appear to decay hyperbolically as it is expected for long-memory processes. We observe that long memory is more evident in the case of the stock index. In Figure 4, we plot the autocorrelations for both positive and negative realized semivariances. It is evident that negative realized semivariance is much more serially correlated than the positive one. In summary, we may expect good performances in forecasting the realized variance and a different behavior from the negative and positive realized semivariances.

Signed jumps are plotted in Figure 5, from which we do not observe any particular pattern. Importantly, signed jumps do not appear to cluster in time. In addition, the average level of signed jumps magnitude seems to be rather constant over time, which is a significant feature for such a long period of time.

## 4 Modeling conditional volatility using Corsi's HAR and some HAR's extensions

### 4.1 The econometric specifications

Our empirical approach adopts the HAR model developed in Corsi (2009). This model has been successfully applied in a number of recent contributions including ABD, Liu and Maheu (2009), Corsi et al. (2010) and Maheu and McCurdy (2011), among many others. The main feature of this model is that it can be estimated using standard ordinary least squares (OLS) but succeeds in reproducing long memory, an important empirical characteristic of volatility processes. In fact, the HAR model is a superposition of short-memory processes that is equivalent to a pseudo long-memory process.

Bollerslev and Wright (2001) emphasize the long-memory behavior of the volatility series and suggest a method-

<sup>11</sup>See, again, Hansen and Lunde (2006) for a thorough discussion of this issue and Andersen *et al.* (2011) for a theoretical and empirical analysis of the impact of microstructure noise on the forecast of realized variance. To deal with this issue, we use staggered versions of BPV and MedRV as advocated in Huang and Tauchen (2005).

<sup>12</sup>To keep the paper self-contained, we report in Figure 1 the volatility signature plot which provides the average of our three realized estimators at various sampling frequencies.

ology which consider this stylized fact along with daily data and allows to produce volatility forecasts that significantly outperform standard volatility predictions from time-series models. Martens and Zein (2004) use ARFIMA model and intraday data to deal with the long memory issue in the volatility time series. Investigations of the long memory feature of commodity volatility series also is in Baillie et al. (2007), Chen et al. (2006), Elder and Serletis (2008) and Wang et al. (2011) where the authors rely on GARCH model extensions to deal with the long memory modeling issue. The authors show empirically that GARCH-class models are only able to replicate long memory that is present in financial time-series under restricted circumstances. A recent paper of Liu and Wan (2012) also highlights the long-range dependence in volatility measures for Shanghai fuel oil futures returns. Arouri et al. (2012) show that long-memory and structural breaks are important features of crude oil markets. Martens et al. (2009) also show the critical importance of long-memory for modeling the volatility of the S&P 500 futures.

To facilitate the presentation of the models, we now drop the sampling frequency from the notation of realized estimators. Let us first consider the following notation for the average realized variance over the period  $[t + 1, t + h]$ :

$$RV_{t+1,t+h} = h^{-1}[RV_{t+1} + RV_{t+2} + \dots + RV_{t+h}]$$

Using this notation, and as a basis for our analysis, we consider the HAR-RV model which is derived from Corsi's (2009) original specification:

$$RV_{t+1,t+h} = \beta_0 + \beta_1 RV_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (12)$$

where  $\varepsilon_t$  is an unpredictable error term. Our HAR-RV model differs from the HAR specification in Corsi (2009, Eq. (8)) in two dimensions. First, our model relies on the realized variance when Corsi (2009) uses the realized volatility (the square root of the realized variance). This choice is motivated by the fact that Andersen et al. (2007) find similar results when using the realized variance, the realized volatility or the logarithm of realized volatility when estimating the HAR model parameters. Consequently, recent literature, including Patton and Sheppard (2011) among others, often disregards the nonlinear transformations of the realized variance which does not provide evident added value. Second, the lag structure in our model is slightly different from the one in Corsi's (2009) HAR model as we avoid overlapping terms in explanatory variables. This reparametrization is drawn from Patton and Sheppard (2011) and allows for direct interpretation of the effect of each term through its estimate, which is not the case with overlapping terms. In the rest of the paper, we refer to the "original Corsi's model" for the specification in Eq. (12) despite we use realized variance and a different lag structure.

In words, the HAR model considers the modeling of the realized variance over the period  $[t + 1, t + h]$  using as predictors the one-day, one-week and one-month lagged averaged realized variances. The model follows the heterogenous structure of  $\tilde{M}\tilde{A}_{\frac{1}{4}}$  Iler et al. (1997). Importantly, with this model, we are interested in the forecast of the cumulative realized variance over a given period which is a quantity of interest for most investors and derivatives practitioners. In particular, recent variance swaps are indexed on the realized variance over a period of interest which corresponds to a cumulative quantity in terms of realized variance.

An relevant issue in specifying an HAR model is the lags that should be considered. In particular, is the one-day, one-week and one-month lag originally advocated in Corsi (2009) optimal? Craioveanu and Hillebrand (2008) investigate this issue. The authors systematically analyze a number of alternative lag structure and conclude in favor of the original structure proposed in the genuine Corsi's (2009) HAR model. While the results in Craioveanu and Hillebrand (2008) do not use oil price data as an input, we adopt the one-day, one-week, one-month lag structure in the empirical part of the paper.<sup>13</sup>

Simple and intuitive extensions of the HAR model are suggested in ABD where nonlinear forms of the original structure are estimated. In particular, the authors consider the logarithmic and square-root analog of the HAR. The motivation for these nonlinear forms comes from the near-log-normality of realized variance. We follow Patton and Sheppard (2011), who recommend to avoid the use of power or log transform for Mincer-Zarnowitz-type regressions as they could lead to an asymptotic distortion of the test, and only estimate linear forms of the HAR model. Also, as in Patton and Sheppard (2011), we use Weighted Least Squares (WLS) to limit the role of periods of high variance. To do this, we start with the estimation of the model using OLS and then consider the fitted value from that model as the weights in the WLS estimation. From now on, we only use WLS for both in and out-of-sample estimation.

We now present the eleven time series models that will be estimated and compared in and out-of-sample.<sup>14</sup> For each model, we also provide basic comments so as to situate the model in the literature and the empirical results related to the model:

**Model 1 – HAR-RV** The HAR-RV is the original specification in Corsi (2009) (see Eq(12)).

**Model 2 – HAR-RV-J** The HAR-RV-J is the first model in ABD (Section 4, Eq. 11), where the simple HAR model is completed with a jump component using the one-day lagged squared jump. Using this model, the main finding in ABD is that  $\beta_{SQJ}$  is negative and significant. Recently, Corsi et al. (2010) have shown that a more accurate jump detection test allows to find significant predictability for the jump component but with a positive sign.

$$RV_{t+1,t+h} = \beta_0 + \beta_1 RV_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \beta_{SQJ} J_t + \varepsilon_t \quad (13)$$

**Model 3 – HAR-CJ** The HAR-CJ model is the ultimate specification in ABD (Section 6, Eq. 26) where continuous and squared jumps components are separated at different horizons. This is also the model used in Tseng et al. (2009) using oil data. Recall that in our work we use the convention in Patton and Sheppard (2011) so that there is no overlap between lags for the different component of the HAR model. As such, our model is slightly different from the model in ABD.

$$RV_{t+1,t+h} = \beta_0 + \beta_{C1} C_t + \beta_{SQJ1} J_t + \beta_{C5} C_{t-1,t-4} + \beta_{SQJ5} J_{t-1,t-4} + \beta_{C22} C_{t-5,t-21} + \beta_{SQJ22} J_{t-5,t-21} + \varepsilon_t \quad (14)$$

**Model 4 – PS** The PS model is the basic specification in Patton and Sheppard (2011) (Eq. (17)) which decomposes the one-day lagged realized variance into a positive and negative component using realized semi-variances. The main result in Patton and Sheppard (2011) is the high significance of negative realized

<sup>13</sup>Further work may consider alternative structure as robustness check but we doubt that our results could be significantly modified.

<sup>14</sup>For each model,  $\varepsilon_t$  is assumed to be an unpredictable error term.

semivariance and the insignificance of positive one.

$$RV_{t+1,t+h} = \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (15)$$

**Model 5 – PSlev** The PSlev model is similar to the PS model except that it includes a possible inverse leverage effect. The leverage effect can be captured using the specification in Martens et al. (2009) or Patton and Sheppard (2011) which are quite easy to use. The PSlev model is specified in Eq. (18) in Patton and Sheppard (2011) and mainly checks whether the superior significance of negative realized semivariance does not come from a leverage effect.

$$RV_{t+1,t+h} = \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \gamma RV_t I_{[r_t < 0]} + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (16)$$

Extensions of this model to consider a possible leverage effect with a 5 or 22 days horizons have also been estimated but do not provide interesting empirical results.

**Model 6 – HAR-RSV** The model HAR-RSV uses the heterogenous structure of the genuine HAR over negative and positive RSV (this is the model in Eq. (19) in Patton and Sheppard (2011)). The model assumes that positive and negative realized semivariances can have different forecasting power at different lags.

$$RV_{t+1,t+h} = \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \beta_5^- RSV_{t-1,t-4}^- + \beta_5^+ RSV_{t-1,t-4}^+ + \beta_{22}^- RSV_{t-5,t-21}^- + \beta_{22}^+ RSV_{t-5,t-21}^+ + \varepsilon_t \quad (17)$$

**Model 7 – CG** As a benchmark for their study based on nonparametric estimators, Chen and Ghysels (2011, Eq. (7)) propose a version of the HAR-RSV model that includes the one-day lagged squared jump component in the spirit of ABD. This is the model CG in our study.

$$RV_{t+1,t+h} = \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \beta_5^- RSV_{t-1,t-4}^- + \beta_5^+ RSV_{t-1,t-4}^+ + \beta_{22}^- RSV_{t-5,t-21}^- + \beta_{22}^+ RSV_{t-5,t-21}^+ + \beta_{SQJ} J_t + \varepsilon_t \quad (18)$$

**Model 8 – HAR-RV-SJ** The HAR-RV-SJ model introduces the notion of signed jumps as defined in our Eq. (11). The model is similar to the HAR-CJ model of ABD except that the squared jump component is replaced with the signed jump as advocated in Patton and Sheppard (2011, Eq. (21)).

$$RV_{t+1,t+h} = \beta_0 + \beta_J \Delta J_t + \beta_C C_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (19)$$

**Model 9 – HAR-CSJ** Model HAR-CSJ is a new specification where we consider signed jumps over a longer interval than one day.<sup>15</sup> Our methodology, following the definition of signed jumps developed in Patton and Sheppard (2011), allows to consider jumps over short period of time and in addition, takes into account the *signs of the jumps*.

$$RV_{t+1,t+h} = \beta_0 + \beta_{1J} \Delta J_t + \beta_{1C} C_t + \beta_{5J} \Delta J_{t-1,t-4} + \beta_{5C} C_{t-1,t-4} + \beta_{J22} \Delta J_{t-5,t-21} + \beta_{22C} C_{t-5,t-21} + \varepsilon_t \quad (20)$$

**Model 10– HAR-RV-SJd** The HAR-RV-SJd model is Eq.(22) in Patton and Sheppard (2011). The model discrim-

<sup>15</sup>Tauchen and Zhou (2011) also develop an empirical model based on the characteristics of jumps in financial time series. But the authors consider the jumps quite differently. First, they are interested in estimating jump distribution characteristics (mean, volatility, intensity). Second, they are interested in significant jumps in the sense of Huang and Tauchen and, as such, need a very large sample (at least two years) to estimate the distribution characteristics.

inates between positive and negative signed jumps:

$$RV_{t+1,t+h} = \beta_0 + \beta_J^- \Delta J_t I_{[\Delta J_t < 0]} + \beta_J^+ \Delta J_t I_{[\Delta J_t > 0]} + \beta_C C_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (21)$$

**Model 11– HAR-CSJd** Model HAR-CSJd is our second new specification where we discriminate between positive and negative signed jumps at various horizons:

$$\begin{aligned} RV_{t+1,t+h} = & \beta_0 + \beta_{1J}^- \Delta J_t I_{[\Delta J_t < 0]} + \beta_{1J}^+ \Delta J_t I_{[\Delta J_t > 0]} + \beta_{1C} C_t \\ & + \beta_{5J}^- \Delta J_{t-1,t-4} I_{[\Delta J_{t-1,t-4} < 0]} + \beta_{5J}^+ \Delta J_{t-1,t-4} I_{[\Delta J_{t-1,t-4} > 0]} + \beta_{5C} C_{t-1,t-4} \\ & + \beta_{22J}^- \Delta J_{t-5,t-21} I_{[\Delta J_{t-5,t-21} < 0]} + \beta_{22J}^+ \Delta J_{t-5,t-21} I_{[\Delta J_{t-5,t-21} > 0]} + \beta_{22C} C_{t-5,t-21} + \varepsilon_t \end{aligned} \quad (22)$$

Note that in this last specification,  $\Delta J_{t-1,t-4}$  is not computed as the difference between positive and negative RSV over the 4 days of interest but as the sum of signed jump over either days where they are positive or days where they are negative. As such, we avoid to hide the importance of signed jumps during a given period. In this case, aggregation may be detrimental.

We may expect that the forecast of volatility may be improved using additional predictors such as the day-of-the-week (see Martens et al. (2009) for an application to the S&P 500 futures) or maturity effects as we deal with futures contracts. We tried to consider the day-of-the-week effect, and in particular the 'Wednesday effect' that might be present in WTI crude oil market due to the release of oil stocks level by the Department of Energy (DoE) each Wednesday.<sup>16</sup> We found this effect insignificant and so do not report results including this variable here. We also considered the maturity effect or Samuelson effect, the widespread view that the volatility of futures returns should be higher near maturity. This also is insignificant as may be expected from the recent contribution by Duong and Kalev (2008) who showed the absence of maturity effect for WTI crude oil futures using intraday data.

Table 1 provides a typology of the eleven models that are analyzed in the present study. Recall that all models including jump and continuous components will be estimated with both BPV and MedRV for the sake of comparison. With all these models in hand, we are now able to compare their empirical forecasting performance first in an in-sample predictive regression and then in an out-of-sample exercise.

## 4.2 In-sample fit analysis

In contrast with ABD, we estimate the model for all  $h = 1$  to 66 when the authors only estimate the models for  $h = 1$ ,  $h = 5$  and  $h = 22$ . This, of course, provides a more precise picture of the predictive power of the various components under consideration. Additionally, it allows to improve our understanding of whether a given component does add information and, of so, when it does.

We first address the issue of comparing models in-sample, *i.e.* the relative performance of models specified in Eq. (12) to (22). An overall performance comparison is visible from the Figure 6 where all  $R^2$  are plotted for forecast horizons varying from 1 to 66 days. Interestingly, from this plot we observe that slightly different

<sup>16</sup>The release of oil stocks sometimes takes place on other days of the week which makes the day-of-the-week effect a special-day effect as for instance in the analysis of Kim (2013).

**Table 1**  
Typology of the empirical specifications used in the paper

Model number	Model name	Reference	Eq. number in this paper
1	HAR-RV	Corsi (2009) Eq. 8	Eq. 12
2	HAR-RV-J	Andersen et al. (2007) Eq. 11	Eq. 13
3	HAR-CJ	Andersen et al. (2007) Eq. 26	Eq. 14
4	PS	Patton and Sheppard (2011) Eq. 17	Eq. 15
5	PSlev	Patton and Sheppard (2011) Eq. 18	Eq. 16
6	HAR-RSV	Patton and Sheppard (2011) Eq. 19	Eq. 17
7	CG	Chen and Ghysels (2011) Eq. 7	Eq. 18
8	HAR-RV-SJ	Patton and Sheppard (2011) Eq. 21	Eq. 19
9	HAR-CSJ	new specification	Eq. 20
10	HAR-RV-SJd	Patton and Sheppard (2011) Eq. 22	Eq. 21
11	HAR-CSJd	new specification	Eq. 22

models deliver very different performances with  $R^2$  going from a low 10-15% to a high 45-50% for the best specifications.<sup>17</sup> In addition, we note that the ranking of the various models is preserved for different forecast horizons, *i.e.* we do not observe significant crossing lines in this plot. Finally, an important feature of the Figure 6 is that  $R^2$  are first increasing and then decreasing for longer forecast horizons. The turning point is not the same for all models. Some of them exhibit their best  $R^2$  for  $h = 20$  when it is  $h = 50$  for other models. We note that most of the models generally have quite stable  $R^2$  for horizons going from 20 to 60 days.

Reporting all estimates for the different models at various horizons is not feasible and would have only a very limited interest. However, to get a better picture of the significance of parameters, we report in Tables 2 and 3 the estimated parameters for all models with a forecast horizon  $h = 10$ . This horizon is illustrative of the main features of the various models. These Tables allow to draw finer conclusions about the relative importance of the various coefficients of each model. In particular, we will be able to understand the additional information content of the jump and signed components such as signed jumps and signed semivariances. We will also be able to conclude about the interest of considering the MedRV estimator of realized variance in place of the BPV estimator. In addition, the various parameters for four important specifications are plotted in Figures 1 to 4 of the supplemental appendix to give a more precise view of the value of parameters for various forecast horizons. For each of these four models, we plot the estimates as well as their *t-stat* to gauge the level and significance of the coefficients at different horizons.

From Table 2, we observe that all the coefficients in the basic HAR-RV model are highly significant. From Figure 1 of the supplemental appendix, we note that this feature is verified for all horizons. This is a common finding in empirical literature using the HAR model due to the well-known (and discussed in Section 2 and the introductory part of Section 4) long memory behavior of volatility. The explanatory power for the HAR-RV specification is roughly in line with the one in ABD. The HAR-RV-J model (columns 2 and 3 of Table 2) leads to a significant increase in the explanatory power (around 10%). This is due to the inclusion of the squared jump component as an additional predictor. This new variable is highly significant at all horizons (see Figure 2 of

<sup>17</sup>Overall, the explanatory power in our study is in line with the one in Andersen et al. (2007). More precisely, it is similar to the explanatory power the foreign exchange spot market (DM/\$), and higher than for the interest rate futures market (30-year U.S. Treasury yield) but lower than for the equity futures market (U.S. S&P 500 index).



the supplemental appendix for a plot of the parameters using the MedRV to detect jumps) and the estimate is unambiguously negative. This is an important result which is in line with the findings in ABD and contrasts with the recent results in Corsi et al. (2010) who find a positive and significant effect of the jump component when measuring the squared jump using their new threshold estimator.<sup>18</sup> Here, we show that the squared jump component is highly significant with both the BPV, which is known to be biased in small samples but also with the MedRV which does not exhibit this weakness. We can thus conclude that in the oil futures market, jumps significantly reduce the impact of lagged volatility that is a highly persistent component and this effect exists for horizons going from 1 day to three months.

The results for the HAR-CJ model are given in the columns 4 and 5. The explanatory power of the regressions are higher than for the HAR-RV-J model from a rough 8-10%. As in ABD, the lagged squared jump components are not significant, except for very short horizons in the case of one-day and one-week lagged jump components and for longer horizons in the case of one-month lagged jump component. This can be observed from Figure 3 of the supplemental appendix where the estimates for the HAR-CJ using the MedRV estimator to estimate the jump component is used. In unreported estimations, we find a similar profile for the squared jump components estimated from the BPV. As in the literature examining other asset markets, the continuous component represents most of the relevant information for predicting future volatility and the squared jump components have no additional value. These results do not contradict our findings for the HAR-RV-J model as in this latter we do not use the continuous component but the realized variance as a predictor along with the one-day-lagged squared jump component. When we disentangle the jump from the continuous component, there is no more additional information in the jump component as far as predicting future volatility is concerned.

The column 6 of Table 2 reports the results for the PS model of Patton and Sheppard (2011). Importantly, the  $R^2$  is the same as for the standard HAR-RV model thereby indicating that the decomposition between positive and negative semivariances does not contribute to improve the fit of the predictive regression. This is a particular characteristic of the WTI market as Patton and Sheppard (2011) find a better  $R^2$  for their model than for the HAR-RV when studying the Spyder, an exchange-traded-fund that replicates the S&P 500 stock index. An additional, and important, feature of the oil market is that the positive realized semivariance is highly significant for all horizons up to two months (see top panel in Figure 4 of the supplemental appendix). This is at odds with the results in Patton and Sheppard (2011) for the Spyder as the authors find an insignificant positive realized semivariance in similar regression to ours. Our intuition behind this result is the particular characteristic of the commodity markets where a crisis situation is more often related to high prices (high positive returns) rather than to low prices (high negative prices) which explains the “inverse leverage effect” generally acknowledged in these markets (see Geman, 2005).<sup>19</sup> We check in column 7 with the PSlev model that a possible leverage (or inverse leverage) effect does not affect our estimates for realized semivariances. This is not the case in our empirical application.

The last three columns of Table 2 report results for the HAR-RSV model and an extension of this model suggested in Chen and Ghysels (2011) which includes the one-day-lagged squared jump component. As for the

<sup>18</sup>The intuition behind the empirical result in Corsi et al. (2010) is that jumps are generally associated with an increase rather than a drop in volatility. The specific feature and the sign of jumps that may be more often negative in commodity markets may explain our findings.

<sup>19</sup>Nomikos and Andriosopoulos (2012) evaluate a number of stochastic models for energy spot prices and provide empirical evidence of a “leverage effect” for WTI while some other energy futures markets exhibit an “inverse leverage effect” as advocated in Geman (2005).

HAR-RSV model, we observe that all the coefficients are highly significant. This is true for lagged negative realized semivariances but also for positive ones. If we compare this model with the genuine HAR-RV model, we note a higher explanatory power due to the decomposition of the variance into its two signed parts. Then, adding the one-day-lagged squared jump component to the analysis leads to a remarkable increase in the explanatory power as it was the case when switching from the HAR-RV model to the HAR-RV-J model. Again, we observe the importance of the the jump component either using BPV or MedRV to detect and measure this component, at least when the continuous component is not considered as a predictor.

Table 3 reports results for models 8 to 11 (see Table 1), *i.e.* models including signed jumps components as advocated in Patton and Sheppard (2011). The results for the HAR-RV-SJ model are reported in the first two columns of the Table 3. We observe that the estimate for the signed jump component is significant at the 5% level and that the explanatory power of the model is similar to the HAR-RV-J's one. This indicates that there is no specific gain in-sample in considering signed jumps. In addition, from columns 5 and 6 of the Table 3, we also conclude that there is no additional increase in explanatory power when signed jumps are considered separately depending on their sign. Nevertheless, we observe that positive signed jumps are highly significant when negative ones are insignificant which is in line with our previous remark on the importance of considering the commodity markets differently from other financial assets. Here, a positive jump in price is associated with future lower volatility while there is no correlation between negative signed jump and future volatility.

In the HAR-RV-SJ model, we decompose the continuous component from the signed jumps at daily, weekly and monthly lags. In fact, we adopt a heterogenous structure as is the case for RV in the genuine HAR model. Our results for this model where the continuous component is estimated either with BPV (column 3) or MedRV (column 4) show that the one-day and one-week lagged signed jumps are insignificant. Conversely, the one-month-lagged signed jump component along with the three continuous components are significant at the 1% level and lead to the highest  $R^2$  of our study with 32% at a 10-day horizon and around 45% for horizons higher than one month and a half. This result highlights the information content of signed jumps when the latter are measured over a sufficient period of time, as evidenced in Tauchen and Zhou (2011). Daily and weekly measures of signed jumps appear to be too noisy to have some predictive power in our setting where future realized variance is the object of interest. Finally, the estimates for the HAR-CSJd model are provided in columns 7 and 8 of Table 3. We note that the signed jump components are insignificant, in contrast with the continuous components, except for the one-month lag where the positive signed jumps are highly significant. We obtain a  $R^2$  that is in line with that of HAR-CSJ and HAR-RV-J models as can be observed from Figure 6.

As a summary of our results from the in-sample analysis, we conclude that the additional explanatory power of signed jumps as advocated in Patton and Sheppard (2011) is limited in the WTI oil futures market. In particular, the HAR-RV-J model developed in the seminal study by ABD delivers results that are as good as more sophisticated models in our predictive regression and significantly better than (i) models that ignore the jump component and (ii) models that decompose the realized variance into its positive and negative part. In the next section, we will compare our models in an out-of-sample analysis so as to validate or invalidate in-sample conclusions.

## 5 Forecasting conditional volatility and forecasts evaluation

In this Section, we analyze the out-of-sample performances of the eleven models. As such, we investigate the real forecasting performance of the models whose in-sample fit as been tested in the previous Section. Out-of-sample analysis is the only way to gauge the forecasting performance in this particular “volatility horse race” (Giot and Laurent (2007)) and to conclude about the additional predictive performance of the various components of the realized variance.

As is common in the forecasting literature, we rely on the Diebold-Mariano-White (DMW) statistic developed in Diebold and Mariano (1995) and West (1996) (see Patton and Sheppard (2009) for an excellent survey on this topic). The use of this statistic, which compares the forecast ability of two (possibly nested) competing models, requires a loss function that is a measure of the difference between the realized value and the forecast in a pseudo out-of-sample forecasting environment. Because we do not observe volatility but rather use a proxy that is imperfect in nature, we need to consider a robust loss function in the sense of Patton (2011), that is, a loss function which ranks two competing forecasts similarly whether the ranking is done using the true conditional variance or some conditionally unbiased volatility proxy. Patton (2011) demonstrates the robustness of the *Q-LIKE* loss function that is defined as follows:

$$\mathcal{L}(\hat{\sigma}^2, f) = \log f + \frac{\hat{\sigma}^2}{f}$$

with  $f$  the volatility point forecast and  $\hat{\sigma}^2$  is the proxy for the conditional volatility.

Let us consider two forecast models  $A$  and  $B$ . At each point in time  $t$ , the difference in loss functions for the two models is computed as:

$$d_{t,\{A,B\}} = \mathcal{L}(\hat{\sigma}_t^2, f_t^A) - \mathcal{L}(\hat{\sigma}_t^2, f_t^B)$$

Then, a DMW test of equal predictive accuracy is a simple Wald test that the expected value of this difference is zero. The number of daily observations in our sample is  $T = 5732$  and we use a rolling window framework<sup>20</sup> with estimation windows of  $w = 1000$  observations. Thus, with  $h$ -steps forecasts from 1 to 66 days, we are left with  $\tau = T - w - 66 + 1 = 4667$  predictions for each horizon. The loss can be computed for each day and is averaged to yield the mean loss for each model at horizon  $h = 1, \dots, 66$ . The losses are also used to compute the HAC long run variance following Newey-West (1987) procedure with a lag chosen to be  $h_{\max} * 0.15$  due to the overlapping nature of our exercise. The DMW statistic is then given by:

$$DM - QLIKE_{\tau,\{A,B\}} = \frac{\bar{d}_{t=1,\dots,\tau,\{A,B\}}}{\hat{\Sigma}_\tau / \sqrt{\tau}}$$

with  $\hat{\Sigma}_\tau$  an estimator of the asymptotic standard deviation of  $\Sigma_\tau = \sqrt{\text{var}[\sqrt{\tau} \bar{d}_{t=1,\dots,\tau,\{A,B\}}]}$ . Martens et al. (2009), Patton and Sheppard (2011), among many others, use this measure to evaluate their forecasts.<sup>21</sup> The

<sup>20</sup> See West (2006) who motivates the choice of a rolling, recursive or fixed window scheme.

<sup>21</sup> Martens et al. (2009) also use a series of other measures like the Model Confidence Set that are better suited when the number of models is quite large.

statistic is standard normal and permits an easy comparison of pairs of models at each horizon.

We report DMW statistics for various forecasting horizons in Tables 4 to 9. When models are nested (for instance the HAR-RV and the HAR-RV-J models), we use the adjustment to the DMW statistic suggested in Clark and West (2007). The interpretation of the figures in the Tables is as follows. The reported statistic compares the model whose name is given in the headline with the model whose name is given in the head-column. A positive statistic indicates that the latter outperforms the former. Significant statistics are highlighted in bold fonts. For instance, in Table 4, the reported 3.3026 in the first column indicates that the HAR-RV-J model significantly outperforms the HAR-RV model for the one-day forecast horizon. Another example is the -2.2020 reported in the third column, which indicates the superiority of model HAR-RV-J over the PS model.

Table 4, which reports results for  $h = 1$ , indicates that forecasts from the HAR-RV models are significantly dominated by most of the other models. As could be expected from the in-sample analysis, the forecasting performance of the HAR-RV-J model is very good as it outperforms most of other models, often significantly, and is never significantly outperformed. The HAR-CJ model also has good relative forecast accuracy, but not as good as the HAR-RV-J model. Other more sophisticated models including signed variances or signed jumps do not exhibit interesting performances at this one-day horizon.

Considering the one-week horizon, we observe from Table 5 that no model is statistically superior to others. Except an anecdotal significant statistic indicating the better performance of the PS model over the HAR-RV-J model, we can conclude that at this horizon, all models deliver comparable forecasting performances. This was not expected from the in-sample analysis where  $R^2$  were between a low 15% and a high 27%.

Conclusions for the two-week forecast horizon are quite different as the HAR-RV and the PS models seem to deliver forecasts that are significantly better than other models. We observe in Table 6 that these two models are never dominated by others and often significantly dominate other models. A comparison between this pair of models leads to an insignificant 0.5172. For longer horizons, one model exhibits very interesting forecast performance. This is the genuine HAR-RV model which is never beaten at the one, two and three-month horizons. This result is surprising for two reasons. First, the HAR-RV has modest in-sample fit when compared to more sophisticated models such as the HAR-CJ, the CG, the HAR-CSJ or the HAR-CSJd models. This points to the well-known in-sample over-fitting which does not translate into good out-of-sample forecasts. Second, the HAR-RV model is a very innovative model but it remains relatively simple when compared to its successors initiated in ABD and many other recent contributions. Despite the fact that more complicated specifications have good out-of-sample performance, they never outperform the HAR-RV which is our preferred model for all horizons except the one-day ahead. These results are corroborated in further empirical analysis, using alternative loss functions that are also discussed in Patton (2011), whose results are reported in a supplemental appendix (available on-line). Collectively taken, our results indicate that the HAR-RV is a very attractive model for all loss functions, thereby showing that our empirical findings are not linked to the use of a particular loss function.

To summarize, our findings show that, while in-sample performance can be dramatically improved using the decomposition between jumps and the continuous component and negative and positive realized semivariances, these decompositions does not lead to significant improvements out-of-sample in oil markets. This

important empirical fact is not tested in previous papers where no out-of-sample analysis is conducted. It may be the case that for other assets that have been studied previously, such as FX, S&P 500 futures, T-Bonds futures, etc., an artificial importance has been given to the jump component for forecasting purpose. Such an analysis is, of course, far beyond the scope of the present paper.

## 6 Conclusion

This paper presents results from an empirical analysis of eleven predictive HAR-type time-series models whose aim is to forecast realized variance. It is shown that considering independently the squared jump component, the continuous component, signed jumps and realized semivariances of both signs significantly help to improve the fit of the predictive regressions. However, when confronted to the challenge of forecasting the realized variance in a pseudo out-of-sample experiment, the simple HAR model of Corsi (2009) provides results that are at least as good as more sophisticated models and often significantly better.

One limit of our analysis is that we do not consider periodicity and this latter feature may have a significant impact on the detection of jumps. The issue of periodicity has been investigated in the early contribution by Andersen and Bollerslev (1997). The recent methodology developed in Boudt et al. (2011) following the early work in Taylor and Xu (1997) may be used to detect jumps and the forecasting properties of the estimated jump component may be compared with more classical estimates based on standard tests developed in Barndorff-Nielsen and Shephard (2004, 2006).

A first possible extension of the present paper may be to consider the semi-parametric approach in Becker *et al.* (2011). Their method consists in finding periods of relatively similar volatility dynamics to weight past volatilities in an optimal way. The authors show that the forecasting accuracy of their model is better than that of more standard approaches but at the cost of estimating semi-parametric specifications. Another possible extension would be a comparison of implied-volatility vs. realized measures in line with Giot and Laurent (2007) as a new volatility index (OVX) has been introduced on the CME platform to deal with volatility risk in U.S. oil futures markets. Finally, the HAR model may also be used along with a model for the conditional correlation (see for instance Christodoulakis and Satchell (2002) or Christodoulakis (2007)) to develop a mixed-frequency model in the spirit of Halbleib and Voev (2012). The idea is to use intraday data along with a time-series model for the dynamics of univariate volatilities and a model of conditional correlation that could be the DCC model of Engle (2002) for the modeling of the dynamics of correlation(s) using this time daily data.

	HAR-RV	HAR-RV-J(BPV)	HAR-RV-J(MED)	HAR-CJ(BPV)	HAR-CJ(MED)	PS	PSlev	HAR-RSV	CG(BPV)	CG(MED)
$\beta_0$	6.038E-05*** (5.808)	5.698E-05*** (5.577)	5.171E-05*** (5.312)	4.202E-05*** (5.168)	3.936E-05*** (5.002)	5.997E-05*** (5.792)	5.990E-05*** (5.824)	6.526E-05*** (7.556)	5.693E-05*** (6.966)	5.622E-05*** (6.831)
$\beta_1$	0.138*** (6.972)	0.332*** (6.974)	0.370*** (8.726)							
$\beta_5$	0.168*** (6.316)	0.104*** (4.535)	0.110*** (4.934)			0.167*** (6.298)	0.165*** (6.328)			
$\beta_{22}$	0.568*** (10.037)	0.491*** (7.807)	0.475*** (7.281)			0.567*** (10.028)	0.564*** (10.041)			
$\beta_{SQJ1}$		-0.354*** (-6.499)	-0.366*** (-8.343)	-0.008 (-0.886)	0.019 (1.663)				-0.395*** (-11.289)	-0.376 (-10.398)
$\beta_{C1}$				0.291*** (8.731)	0.290*** (8.470)					
$\beta_{SQJ5}$				-0.029 (-1.446)	-0.029 (-0.521)					
$\beta_{C5}$				0.238*** (5.127)	0.247*** (5.008)					
$\beta_{SQJ22}$				0.004 (0.090)	0.001 (0.026)					
$\beta_{C22}$				0.551*** (8.358)	0.570*** (8.686)					
$\beta_1^-$						0.160*** (5.571)	0.115*** (3.106)	0.144*** (5.581)	0.392*** (11.848)	0.396*** (11.221)
$\beta_1^+$						0.127*** (3.827)	0.147*** (4.149)	0.134*** (4.084)	0.345*** (10.589)	0.360*** (10.288)
$\beta_5^-$								0.202*** (5.920)	0.105*** (4.204)	0.106*** (4.123)
$\beta_5^+$								0.149*** (3.684)	0.125*** (3.424)	0.123*** (3.321)
$\beta_{22}^-$								0.905*** (8.578)	0.765*** (7.235)	0.769*** (7.120)
$\beta_{22}^+$								0.154** (2.084)	0.118* (1.814)	0.109* (1.684)
$\gamma$							0.018 (1.419)			
$R^2$	0.147	0.236	0.237	0.307	0.311	0.149	0.146	0.213	0.284	0.284

**Table 2**

Predictive regression between average realized variance over 10 days and predictors which are components of realized variance with various lags following models 1 to 7 listed in Table 1. Estimation is by WLS with fitted values of an OLS regression for weights. Newey-West adjusted t-statistics are given in parenthesis. Asterisks indicate statistical significance (at least) at the 10%(\*), 5%(\*\*) and 1%(\*\*\*) level.

	HAR-RV-SJ(BPV)	HAR-RV-SJ(MED)	HAR-CSJ(BPV)	HAR-CSJ(MED)	HAR-RV-SJd(BPV)	HAR-RV-SJd(MED)	HAR-CSJd(BPV)	HAR-CSJd(BPV)
$\beta_0$	5.153E-05*** (5.325)	5.094E-05*** (5.235)	4.465E-05*** (5.742)	4.174E-05*** (5.509)	5.093E-05*** (5.284)	5.116E-05*** (5.260)	4.379E-05*** (5.766)	4.136E-05*** (5.508)
$\beta_5$	0.110*** (4.938)	0.110*** (4.958)			0.122*** (5.454)	0.110*** (4.984)		
$\beta_{22}$	0.480*** (7.762)	0.478*** (7.308)			0.495*** (7.631)	0.476*** (7.268)		
$\Delta J1$	-0.024** (-2.122)	-0.019** (-2.300)	-0.009 (-0.867)	-0.017 (-1.637)				
$\Delta J1^-$					0.009 (0.921)	-0.014 (-1.214)	0.001 (0.067)	-0.028* (-1.663)
$\Delta J1^+$					-0.062*** (-8.126)	-0.025*** (-3.603)	-0.024 (-1.556)	0.003 (0.171)
$\beta_{C1}$	0.359*** (9.052)	0.372*** (8.755)	0.282*** (9.153)	0.282*** (8.855)	0.341*** (7.212)	0.375*** (8.751)	0.289*** (9.093)	0.279*** (8.569)
$\Delta J5$			0.008 (0.578)	0.009 (0.831)				
$\Delta J5^-$							0.033* (1.910)	0.019 (1.203)
$\Delta J5^+$							-0.061* (-1.889)	-0.036 (-1.097)
$\beta_{C5}$			0.223*** (5.143)	0.247*** (5.148)			0.252*** (5.424)	0.262*** (5.310)
$\Delta J22$			-0.148*** (-4.313)	-0.125*** (-4.119)				
$\Delta J22^-$							-0.122 (-1.431)	-0.139 (-1.653)
$\Delta J22^+$							-0.162*** (-5.003)	-0.126*** (-4.301)
$\beta_{C22}$			0.552*** (9.149)	0.564*** (8.876)			0.549*** (8.567)	0.555*** (8.612)
$R^2$	0.233	0.238	0.317	0.320	0.235	0.238	0.316	0.317

**Table 3**

Predictive regression between average realized variance over 10 days and predictors which are components of realized variance with various lags following models 8 to 11 listed in Table 1. Estimation is by WLS with fitted values of an OLS regression for weights. Newey-West adjusted t-statistics are given in parenthesis. Asterisks indicate statistical significance (at least) at the 10%(\*), 5%(\*\*) and 1%(\*\*\*) level.

	HAR-RV-J	HAR-CJ	PS	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
HAR-RV	<b>3.3026</b>	<b>3.0720</b>	1.7115	<b>2.5774</b>	<b>3.0550</b>	<b>2.8610</b>	<b>3.5116</b>	<b>2.7388</b>	1.5825
HAR-RV-J		-1.0327	<b>-2.2020</b>	<b>-3.0868</b>	-0.8972	<b>-1.9903</b>	-0.6124	<b>-1.9607</b>	0.8811
HAR-CJ			-1.0638	<b>-2.3628</b>	0.1045	-0.3423	0.6159	-0.3972	1.2261
PS				-0.9906	1.3642	0.8511	1.5479	0.8597	1.3178
HAR-RSV					<b>2.3428</b>	<b>2.4138</b>	<b>2.8952</b>	<b>2.2590</b>	1.4520
CG						-0.4052	0.5229	-0.4727	1.0778
HAR-RV-SJ							1.0174	-0.1889	1.1556
HAR-CSJ								-1.0105	0.9854
HAR-RV-SJd									1.1690

**Table 4**

Diebold-Mariano (1995) statistic computed with the QLIKE loss function  $\mathcal{L}(\hat{\sigma}^2, h) = \log h + \frac{\hat{\sigma}^2}{h}$  with a forecast horizon  $h = 1$ . A positive test statistic indicates that the model whose name is given in the headline outperforms the model whose name is provided in the head-column. The statistic is computed using the Newey-West (1987) HAC which is a consistent estimate of the asymptotic variance. The number of lags for the computation of the HAC is chosen to be  $\lfloor 1.2 \times T^{1/3} \rfloor$ . Significant statistics are indicated in bold fonts.

	HAR-RV-J	HAR-CJ	PS	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
HAR-RV	-1.4196	1.1869	0.0207	1.4771	0.9292	-1.5107	0.3534	-1.1906	1.4782
HAR-RV-J		1.5266	<b>2.4853</b>	1.6850	1.2043	-0.8849	0.8932	-0.6066	1.6093
HAR-CJ			-1.1524	1.4654	0.2589	-1.3974	-0.8296	-1.2881	1.2740
PS				1.4758	0.9139	-1.4762	0.3353	-1.1696	1.4660
HAR-RSV					-1.1913	-1.5686	-1.4235	-1.5043	0.7432
CG						-1.2665	-1.0869	-1.2145	1.6240
HAR-RV-SJ							1.2792	0.7561	1.6825
HAR-CSJ								-1.2274	1.7054
HAR-RV-SJd									1.6844

**Table 5**

DM statistic for the QLIKE loss function with a forecast horizon  $h = 5$ . Other comments idem to Table 4

	HAR-RV-J	HAR-CJ	PS	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
HAR-RV	-1.8055	<b>-2.7934</b>	0.5172	-0.1677	0.9924	<b>-4.7788</b>	<b>-3.4251</b>	<b>-3.3135</b>	0.6503
HAR-RV-J		-1.5828	<b>3.1309</b>	1.1503	1.3081	-1.2069	-1.4642	-0.7478	1.0961
HAR-CJ			<b>2.7745</b>	<b>2.9210</b>	1.7611	0.6447	0.5871	1.3315	1.6496
PS				-0.4694	0.8831	<b>-2.8032</b>	<b>-3.0679</b>	<b>-3.1993</b>	0.5105
HAR-RSV					1.0720	<b>-2.6683</b>	<b>-3.2218</b>	-1.8515	0.6604
CG						-1.6053	-1.6551	-1.4366	-0.5760
HAR-RV-SJ							-0.3530	1.2740	1.4880
HAR-CSJ								1.2262	1.5208
HAR-RV-SJd									1.2638

**Table 6**

DM statistic for the QLIKE loss function with a forecast horizon  $h = 10$ . Other comments idem to Table 4

	HAR-RV-J	HAR-CJ	PS	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
HAR-RV	<b>-4.3015</b>	<b>-3.3385</b>	<b>-2.5126</b>	0.1061	<b>-2.5506</b>	<b>-4.4310</b>	<b>-3.1077</b>	<b>-4.4355</b>	<b>-3.2045</b>
HAR-RV-J		<b>-2.0448</b>	<b>4.0416</b>	<b>2.0337</b>	-0.2033	-1.5529	-1.4875	-1.4790	-1.7761
HAR-CJ			<b>3.1583</b>	<b>3.2458</b>	<b>2.2253</b>	1.8258	1.0221	1.8103	0.3657
PS				0.6038	<b>-2.2552</b>	<b>-4.4014</b>	<b>-2.8944</b>	<b>-4.4511</b>	<b>-2.9896</b>
HAR-RSV					<b>-2.5535</b>	<b>-2.2103</b>	<b>-2.9805</b>	<b>-2.2212</b>	<b>-2.9878</b>
CG						-0.2654	-1.7268	-0.2956	-1.8923
HAR-RV-SJ							-1.2260	-0.2344	-1.5883
HAR-CSJ								1.2281	-0.7027
HAR-RV-SJd									-1.5623

**Table 7**

DM statistic for the QLIKE loss function with a forecast horizon  $h = 22$ . Other comments idem to Table 4



	HAR-RV-J	HAR-CJ	PS	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
HAR-RV	<b>-3.6959</b>	<b>-2.9948</b>	<b>-1.9654</b>	-0.1150	<b>-2.0308</b>	<b>-4.0044</b>	0.4463	<b>-3.7539</b>	<b>-2.1786</b>
HAR-RV-J		<b>-1.9709</b>	<b>3.5712</b>	<b>2.0339</b>	-0.1192	-1.7675	0.7110	-1.1798	-1.0553
HAR-CJ			<b>2.8404</b>	<b>3.2334</b>	<b>1.9912</b>	1.7296	1.0592	1.8692	<b>2.3864</b>
PS				0.3721	-1.8236	<b>-4.0908</b>	0.5067	<b>-3.7630</b>	<b>-1.9846</b>
HAR-RSV					<b>-5.3804</b>	<b>-2.3496</b>	0.4637	<b>-2.1872</b>	<b>-2.2612</b>
CG						-0.4196	0.7225	-0.2183	-0.9959
HAR-RV-SJ							0.7674	<b>2.2461</b>	-0.7623
HAR-CSJ								-0.7465	-0.9348
HAR-RV-SJd									-0.8968

**Table 8**

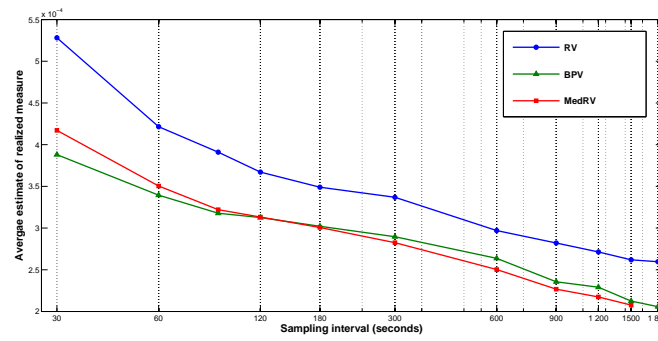
DM statistic for the QLIKE loss function with a forecast horizon  $h = 44$ . Other comments idem to Table 4

	HAR-RV-J	HAR-CJ	PS	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
HAR-RV	<b>-2.3815</b>	-1.8599	-1.8730	0.2078	-1.6898	<b>-2.9999</b>	<b>-2.7389</b>	<b>-2.8046</b>	<b>-2.0640</b>
HAR-RV-J		-1.1395	<b>2.0690</b>	1.4349	-0.3986	<b>-2.2326</b>	-1.8409	-1.1402	-1.2268
HAR-CJ			1.6630	<b>2.8295</b>	1.3000	0.8600	-0.7621	0.9658	0.1910
PS				0.6556	-1.4238	<b>-2.9839</b>	<b>-2.5498</b>	<b>-2.5971</b>	-1.8658
HAR-RSV					<b>-4.7409</b>	-1.8779	<b>-4.7023</b>	-1.6315	<b>-2.9995</b>
CG						-0.1462	<b>-2.6861</b>	0.0866	-1.3719
HAR-RV-SJ							-1.5188	1.6047	-0.9202
HAR-CSJ								1.6354	0.9844
HAR-RV-SJd									-1.0431

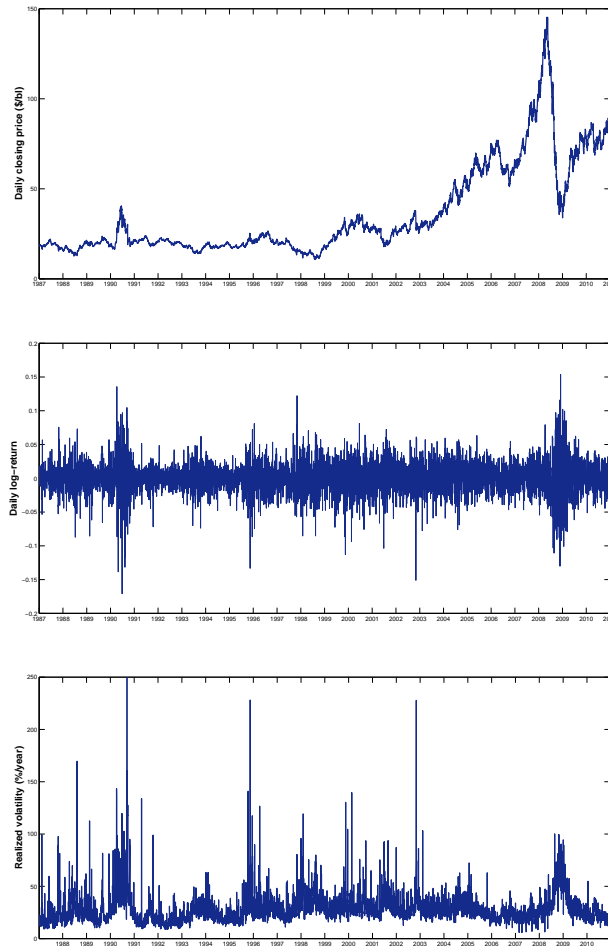
**Table 9**

DM statistic for the QLIKE loss function with a forecast horizon  $h = 66$ . Other comments idem to Table 4

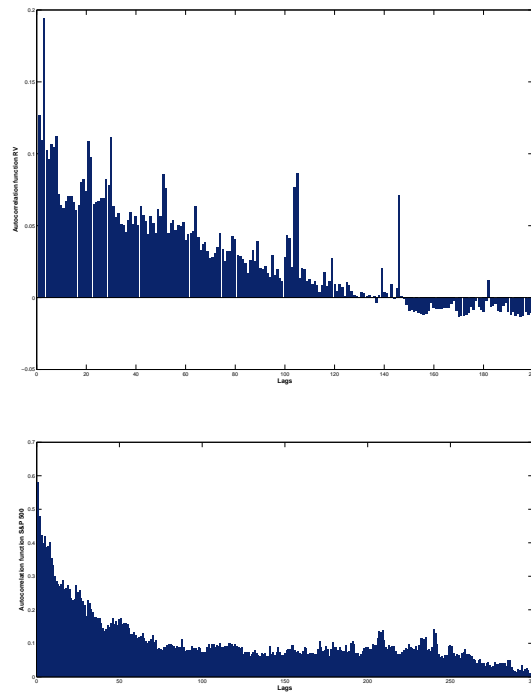
# Figures



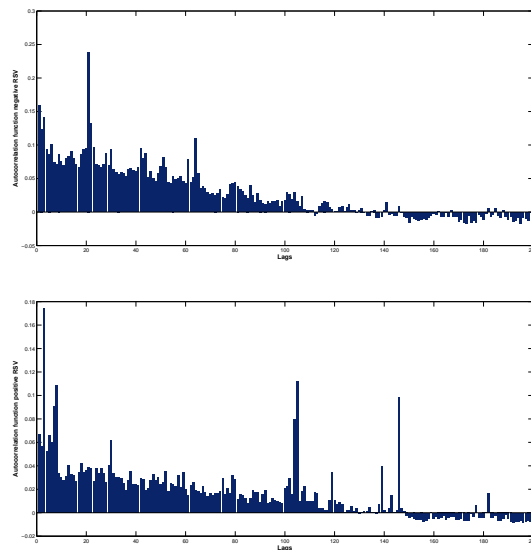
**Figure 1**  
Volatility signature plot for the WTI crude oil front-month futures realized volatility, bipower variation and median realized estimators (January 1987 - December 2010).



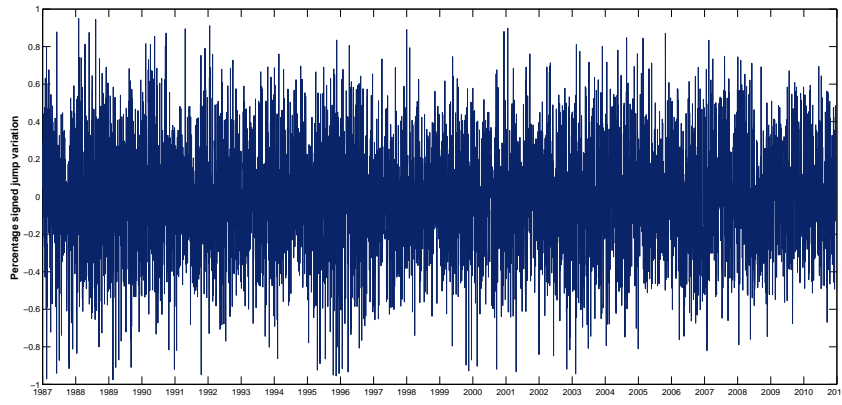
**Figure 2**  
Raw time-series (top panel), log-returns (middle panel) and realized variance (bottom panel) in annualized terms for the WTI NYMEX front-month futures contract over the period .



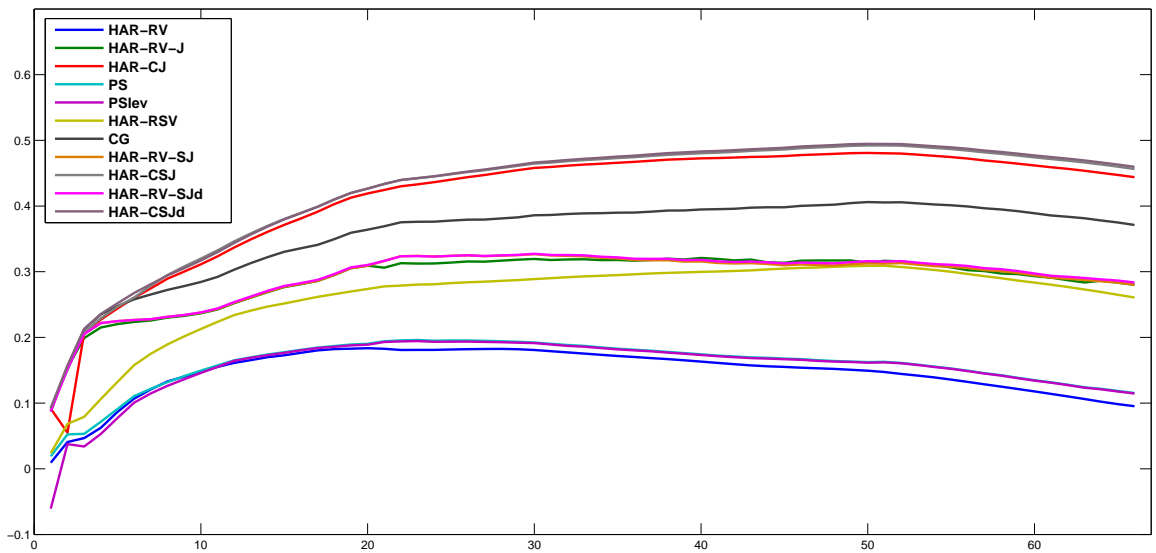
**Figure 3**  
 Sample autocorrelation functions for crude oil futures RV (top) and S&P 500 futures RV (bottom) contracts.



**Figure 4**  
 Sample autocorrelation functions for crude oil futures  $RSV^-$  (top panel) and  $RSV^+$  (bottom) contracts.



**Figure 5**  
Signed jumps.



**Figure 6**  
The reported  $R^2$  are for the eleven predictive regressions described in Table 1 and are computed using WLS parameter estimates along with the original data.

## References

- 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.
- ANDERSEN, T.G., BENZONI, L., LUND, J., 2002. An empirical investigation of continuous-time equity return models. *Journal of Finance* 57, 1239-1284.
- ANDERSEN, T.G., BOLLERSLEV, T., 1997. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115-158.
- 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., 2001. The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42-55.
- ANDERSEN, T.G., BOLLERSLEV, T., DIEBOLD, F.X., LABYS, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579-625.
- ANDERSEN, T.G., DOBREV, D., SCHAUMBURG, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169, 75-93.
- ANG, A., CHEN, J., XING, Y., 2006. Downside risk. *Review of Financial Studies* 19, 1191-1239.
- AROURI, M.E.H., LAHIANI, A., LÄCVY, A., NGUYEN, D. K., 2012. Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. *Energy Economics* 34, 283-293.
- BAILLIE, R.T., HAN, Y.-W., MYERS, R.J., SONG, J., 2007. Long memory models for daily and high frequency commodity futures returns. *Journal of Futures Markets* 27, 643-668.
- BANDI, F., RUSSELL, J.R., ZHU, Y., 2008. Using high-frequency data in dynamic portfolio choice. *Econometric Reviews* 27, 163-198.
- BARNDORFF-NIELSEN, O., KINNEBROCK, S., SHEPHARD, N., 2010. Measuring downside risk – Realised semivariance. In: T. Bollerslev, J. Russell and M. Watson (eds) *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford University Press.
- BARNDORFF-NIELSEN, O., SHEPHARD, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1-37.
- BARNDORFF-NIELSEN, O., SHEPHARD, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1-30.
- BECKER, R., CLEMENTS, A.E., HURN, S., 2011. Semi-parametric forecasting of realized volatility. *Studies in Nonlinear Dynamics & Econometrics* 15, Article 1.
- BELLINI, F., FIGÀ-TALAMANCA, G., 2005. Runs tests for assessing volatility forecastability in financial time series. *European Journal of Operational Research* 163, 102-114.
- BOLLERSLEV, T., WRIGHT, J.H., 2001. High-frequency data, frequency domain inference, and volatility forecasting. *Review of Economics and Statistics* 83, 596-602.
- CABEDO, J.D., MOYA, I., 2003. Estimating oil price 'value at risk' using the historical simulation approach. *Energy Economics* 25, 239-253.
- CHEN, Z., DAIGLER, R.T., PARHIZGARI, A.M., 2006. Persistence of volatility in futures markets. *Journal of Futures Markets* 26, 571-594.
- CHEN, X., GHYSELS, E., 2011. News – good or bad – and its impact on volatility predictions over multiple horizons. *Review of Financial Studies* 24, 46-81.
- CHERNOV, M., GALLANT, A.R., GHYSELS, E., TAUCHEN, G., 2003. Alternative models for stock price dynamics. *Journal of Econometrics* 116, 225-257.
- CHEVALLIER, J., SÁCVI, B., 2012. On the volatility-volume relationship in energy futures markets using intraday data. *Energy Economics* 34, 1896-1909.
- CHOOBINEH, F., BRANTING, D., 1986. A simple approximation for semivariance. *European Journal of Operational Research* 27, 364-370.
- CHRISTODOULAKIS, G.A., 2007. Common volatility and correlation clustering in asset returns. *European Journal of Operational Research* 182, 1263-1284.
- CHRISTODOULAKIS, G.A., SATCHELL, S.E., 2002. Correlated ARCH (CorrARCH): Modelling the time-varying conditional correlation between financial asset returns *European Journal of Operational Research* 139, 351-370.
- 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., 2013. Forecasting with option implied information In: G. Elliot and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, vol 2A, Chap 10, p. 581-656, Elsevier.
- CLARK, T., WEST, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138, 291-311.
- CLEMENTS, M.P., GALVAO, A.B., KIM, J.H., 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance* 15, 729-750.
- CORSI, F., 2009. A simple approximate long memory model of realized volatility. *Journal of Financial Econometrics* 7, 174-196.
- CORSI, F., FUSARI, N., LA VECCHIA, D., 2013. Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics* 107, 284-304.
- CORSI, F., PIRINO, D., RENÒ, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159, 276-288.
- CRAIOVEANU, M., HILLEBRAND, E., 2008. Models for daily realized stock volatility time series. Unpublished manuscript.
- 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.
- DING, Z., GRANGER, C.W.J., ENGLE, R.F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.

- DUMITRU, A.-M., URGA, G., 2012. Identifying jumps in financial assets: A comparison between nonparametric jump tests. *Journal of Business and Economic Statistics* 30, 242-255.
- DUAN, J.C., 1995. The GARCH option pricing model. *Mathematical Finance* 5, 13-32.
- DUONG, H.N., KALEV, P.S., 2008. The Samuelson Hypothesis in futures markets: an analysis using intraday data. *Journal of Banking and Finance* 32, 489-500.
- ELDER, J., SERLETIS, A., 2008. Long memory in energy futures prices. *Review of Financial Economics* 17, 146-155.
- ENGLE, R.F., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics* 20, 339-350.
- ERAKER, B., 2004. Do stock prices and volatility jump?: Reconciling evidence from spot and option prices. *Journal of Finance* 59, 1367-1403.
- ERAKER, B., JOHANNES, M.S., POLSON, N.G., 2003. The impact of jumps in returns and volatility. *Journal of Finance* 53, 1269-1300.
- FONG, W.M., SEE, K.H., 2002. A Markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics* 24, 71-95.
- GEMAN, H., 2005. *Commodity and Commodity Derivatives: Modeling and Pricing for Agriculturals, Metals and Energy*. John Wiley and Sons, Ltd. Edition.
- 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.
- GROOTVELD, H., HALLERBACH, W., 1999. Variance vs downside risk: Is there really that much difference? *European Journal of Operational Research* 114, 304-319.
- HANSEN, P.R., LUNDE, A., 2006. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics* 24, 127-218.
- HESTON, S., NANDI, S., 2000. A closed-form GARCH option valuation model. *Review of Financial Studies* 13, 585-625.
- HUA, Z., ZHANG, B., 2008. Improving density forecast by modeling asymmetric features: An application to S&P500 returns. *European Journal of Operational Research* 185, 716-725.
- HUANG, X., 2008. Portfolio selection with a new definition of risk. *European Journal of Operational Research* 186, 351-357.
- HUANG, X., TAUCHEN, G., 2005. The relative price contribution of jumps to total price variance. *Journal of Financial Econometrics* 3, 456-499.
- JOSEPHY, N.H., ACZEL, A.D., 1993. A statistically optimal estimator of semivariance. *European Journal of Operational Research* 67, 267-271.
- KALEV, P.S., DUONG, H.N., 2008. A test of the Samuelson hypothesis using realized range. *Journal of Futures Markets* 28, 680-696.
- 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.
- KIM, M.S., 2013. Modeling special-day effects for forecasting intraday electricity demand. *European Journal of Operational Research* 230, 170-180.
- 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.
- LIU, L., PATTON, A.J., SHEPPARD, K., 2012. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Unpublished manuscript.
- LIU, L., WAN, J., 2012. A study of Shanghai fuel oil futures price volatility based on high frequency data: Long-range dependence, modeling and forecasting. *Economic Modelling* 29, 2245-2253.
- MAHEU, J.M., MCCURDY, T.H., 2011. Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics* 160, 69-76.
- MARKOWITZ, H.M., 1959. *Portfolio Selection: Efficient Diversification of Investments*, 2nd Ed. Wiley, Yale University Press, Basil Blackwell.
- MARKOWITZ, H.M., 2013. Mean-variance approximations to expected utility. *European Journal of Operational Research*, forthcoming.
- MARTENS, M., 2002. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets* 22, 497-518.
- MARTENS, M., VAN DIJK, D., DE POOTER, M., 2009. Forecasting S&P 500 volatility: long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting* 25, 282-303.
- MARTENS, M., ZEIN, J., 2004. Predicting financial volatility: high-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets* 24, 1005-1028.
- MOHAMMADI, H., SU, L., 2010. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics* 32, 1001-1008.
- MÁLLER, U.A., DACOROGNA, M.M., DAVÁ©, 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.
- NOMIKOS, N., ANDRIOSOPOULOS, K., 2012. Modelling energy spot prices: Empirical evidence from NYMEX. *Energy Economics* 34, 1153-1169.
- NOMIKOS, N.K., POULIASIS, P.K., 2011. Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics* 33, 321-337.
- PAN, J., 2000. The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3-50.
- PATTON, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246-256.
- PATTON, A.J., SHEPPARD, K., 2009. Evaluating volatility forecasts. In: T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (Eds) *Handbook of Financial Time Series*. Springer.
- PATTON, A.J., SHEPPARD, K., 2011. Good volatility, bad volatility: signed jumps and the persistence of volatility. Unpublished Manuscript.

- SADORSKY, P., 2006. Modeling and forecasting petroleum futures volatility. *Energy Economics* 28, 467-488.
- SADORSKY, P., MCKENZIE, M.D., 2008. Power transformation models and volatility forecasting. *Journal of Forecasting* 27, 587-606.
- STENTOF, L., 2008. Option pricing using realized volatility. *CREATES Research Paper 2008-13*.
- TAUCHEN, G., ZHOU, H., 2011. Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics* 160, 102-118.
- TAYLOR, S.J., XU, X., 1997. The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317-340.
- THEODOSIOU, M., ŽIKEŠ, P., 2011. A comprehensive comparison of alternative tests for jumps in asset prices. Working Paper, Central Bank of Cyprus, no. 2.
- TSENG, T.-C., CHUNG, H., HUANG, C.-S., 2009. Modeling jump and continuous components in the volatility of oil futures. *Studies in Nonlinear Dynamics & Econometrics* 13, Issue 3, Article 5.
- WANG, Y., WU, C., WEI, Y., 2011. Can GARCH-class models capture long memory in WTI crude oil markets? *Economic Modelling* 28, 921-927.
- WANG, T., WU, J., YANG, J., 2008. Realized volatility and correlation in energy futures markets. *Journal of Futures Markets* 28, 993-1011.
- 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. Elliot, C. Granger and A. Timmermann (Eds), , *Handbook of Economic Forecasting* Volume 1, 99-134. Amsterdam: North-Holland.
- WONG, H.Y., LO, Y.W., 2009. Option pricing with mean reversion and stochastic volatility. *European Journal of Operational Research* 197, 179-187.
- YANG, C., BANDI, F.M., RUSSELL, J.R., 2008. Realized volatility forecasting and option pricing. *Journal of Econometrics* 147, 34-46.
- ZHANG, L., MYKLAND, P.A., AÅT-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.