Predictive Inference for Integrated Volatility

Valentina Corradi†  Walter Distaso‡
University of Warwick  Imperial College London
Norman R. Swanson§
Rutgers University

January 2010

Abstract

In recent years, numerous volatility-based derivative products have been engineered. This has led to interest in constructing conditional predictive densities and confidence intervals for integrated volatility. In this paper, we propose nonparametric estimators of the aforementioned quantities, based on model free volatility estimators. We establish consistency and asymptotic normality for the feasible estimators and study their finite sample properties through a Monte Carlo experiment. Finally, using data from the New York Stock Exchange, we provide an empirical application to volatility directional predictability.

Keywords. Diffusions, realized volatility measures, kernels, microstructure noise, jumps, prediction.

*We thank Yacine Aït-Sahalia, Torben Andersen, Giovanni Baiocchi, Tim Bollerslev, Marcelo Fernandes, Christian Gourieroux, Peter Hansen, Nour Meddahi, Antonio Mele, Michael Pitt, Mark Salmon, Olivier Scaillet, Stefan Sperlich, Victoria Zinde Walsh, and seminar participants at Universidad Carlos III-Madrid, University of Essex, University of Manchester, University of Warwick, Warwick Business School, the Conference on “Capital Markets, Corporate Finance, Money and Banking” at Cass Business School, the 2005 EC2 Conference in Instabul, and the 2nd CIDE conference in Rimini, for helpful comments on earlier drafts of this paper. Corradi and Distaso gratefully acknowledge ESRC, grant codes RES-000-23-0006, RES-062-23-0311 and RES-062-23-0790, and Swanson acknowledges financial support from a Rutgers University Research Council grant.

†University of Warwick, Department of Economics, Coventry CV4 7AL, UK, email: v.corradi@warwick.ac.uk.
‡Imperial College London, Business School, Exhibition Road, London SW7 2AZ, UK, email: w.distaso@imperial.ac.uk.
§Rutgers University, Department of Economics, 75 Hamilton Street, New Brunswick, NJ 08901, USA, email: nswanson@econ.rutgers.edu.
1 Introduction

It has long been argued that, in order to accurately assess and manage market risk, it is important to construct (and consequently evaluate) predictive conditional densities of asset prices, based on actual and historical market information (see, e.g., Diebold, Gunther and Tay, 1998). In many respects, such an approach offers various clear advantages over the often used approach of focusing on conditional second moments, as is customarily done when constructing synthetic measures of risk (see, e.g., Andersen, Bollerslev, Christoffersen and Diebold, 2006). One interesting asset class for which predictive conditional densities are relevant is volatility. Indeed, since shortly after its inception in 1993, when the VIX, an index of implied volatility, was created for the Chicago Board Options Exchange, a plethora of volatility-based derivative products has been engineered (see, e.g., Carr and Lee, 2003).

Given the development of this new class of financial instruments, it is of interest to construct conditional (predictive) volatility densities, rather than just point forecasts thereof. This is challenging, since volatility is inherently a latent variable. However, crucial steps toward the understanding of several features of financial volatility have been taken in recent years, based upon theoretical advances in the use of high frequency returns data. In particular, it is now possible to obtain precise estimators of financial volatility, under mild assumptions on the process driving the behavior of the underlying variables. Such estimators are constructed using intra day realized returns data, and therefore provide a measure of the ex post (realized) variation of asset prices. The distinct advantage of these estimators is that they exploit the often substantial amount of information contained in intra day movements of the underlying asset prices, without relying on a particular model.

The first estimator of integrated volatility is realized volatility, concurrently developed by Andersen, Bollerslev, Diebold and Labys (2001), and Barndorff-Nielsen
and Shephard (2002). Realized volatility consistently estimates the increments of quadratic variation, when the underlying asset follows a semimartingale process, a class of processes which is commonly employed in continuous time modeling. Several variants of realized volatility have subsequently been proposed. These variants are largely motivated by the need to provide consistent estimators of integrated volatility in situations which are quite common in financial markets, such as jumps in the asset price process, and market frictions leading to market microstructure noise. Leading examples include multipower variation (Barndorff-Nielsen and Shephard, 2004) and different estimators that are robust to the presence of microstructure noise (see, e.g., Zhang, 2006, Aït-Sahalia, Mykland and Zhang, 2005,2006, Zhang, Mykland and Aït-Sahalia, 2005, Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2006, 2008). Since all of the estimators discussed above are designed to measure the ex post variation of asset prices, in the remainder of the paper we will call them realized volatility measures.

In this paper, we develop a method for constructing conditional (predictive) densities and associated conditional (predictive) confidence intervals for daily volatility, given observed market information. We show that the proposed estimators are consistent and asymptotically normally distributed, under mild assumptions on the underlying diffusion process. Our results require no parametric assumption on either the functional form of the estimated density, or on the specification of the diffusion process driving the asset price.

The intuition for the approach taken in the paper is the following. Since integrated volatility is unobservable, we use the realized (volatility) measures discussed above as a key ingredient in the construction of kernel estimators. In other words, we construct feasible estimators of conditional densities. However, this introduces a technical difficulty, because each realized measure can be decomposed into integrated
volatility and an estimation error term. Formally,

\[ RM_{t,M} = IV_t + N_{t,M}, \]

where \( RM_{t,M} \) and \( N_{t,M} \) denote a particular realized volatility measure and its corresponding estimation error, respectively. \( IV_t \) denotes integrated volatility, and the subscripts \( t \) and \( M \) refer to a given day, \( t \), and the number of intraday observations, \( M \), used in the construction of the realized measure. Our estimators are therefore based on a variable which is subject to measurement error. We provide sufficient conditions under which conditional density (and confidence interval) estimators based on (the unobservable) integrated volatility and ones based on realized measures are asymptotically equivalent, so that measurement error is asymptotically negligible.

The idea of using a realized measure as a basis for predicting integrated volatility has been also adopted in other papers. Andersen, Bollerslev, Diebold and Labys (2001, 2003), Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev and Meddahi (2004, 2005) deal with the problem of pointwise prediction of integrated volatility, using ARMA models based on realized volatility. The latter authors also investigate the important issue of evaluating the loss of efficiency associated with the use of realized volatility measures, relative to optimal (unfeasible) forecasts (based on the entire volatility path). Andersen, Bollerslev and Meddahi (2006), Aît-Sahalia and Mancini (2008), and Ghysels and Sinko (2006) address the issue of forecasting volatility in the presence of microstructure effects.

The papers cited above deal with pointwise prediction of integrated volatility. To the best of our knowledge, Corradi, Distaso and Swanson (2009) was the first paper to focus on estimation of the conditional density of integrated volatility, by establishing uniform rates of convergence for kernel estimators based on realized measures. However, with regard to notions such as hedging derivatives based on volatility, the crucial question becomes how to assess the interval within which future daily volatility will
fall, with a given level of confidence. In this respect, the uniform convergence result of Corradi, Distaso and Swanson (2009) is not sufficient. This paper provides an answer to this sort of questions by establishing asymptotic normality for estimators of conditional confidence intervals. This is a substantially more challenging task, as the realized measures and hence the measurement error are arguments of the uniform kernel, which is non-differentiable, and thus standard mean value expansion tools are no longer usable. Indeed, the case of conditional densities can be treated essentially as a corollary of the conditional interval case. Moreover, the current paper deals with the general class of cadlag (right continuous with left limit) volatility processes. This makes the computation of the moment structure of the measurement error much more complicated.

In order to assess the finite sample behavior of our statistics, we carry out a Monte Carlo experiment in which pseudo true predictive intervals are used in conjunction with intervals based on various realized measures.

An empirical application to volatility directional predictability, based on New York Stock Exchange data, highlights the potential of our method and reveals the informational content of different volatility estimators.

The rest of the paper is organized as follows. Section 2 defines and establishes the asymptotic properties of the conditional density and confidence interval estimators. Section 3 studies the applicability of the established asymptotic results to various well known realized measures. In Section 4, the results of a Monte-Carlo experiment designed to assess the finite sample accuracy of our asymptotic results are discussed. Section 5 contains an empirical illustration based upon data from the New York Stock Exchange. All proofs are contained in the Appendix.
2 Setup and main results

Denote the log-price of a financial asset at a continuous time $t$ as $Y_t$, and let

$$dY_t = \mu_t dt + \sigma_t dW_t + J_t,$$

(2)

where the drift $\mu_t$ is a predictable process, the diffusion term $\sigma_t$ is a \textit{cadlag} process, $W_t$ is a standard Brownian motion and $J_t$ denotes a finite activity jump process. This specification is very general, and (for example) allows for jump activity in volatility, stochastic volatility and leverage effects.

Following the standard practice in the literature, we introduce market frictions assuming that transaction data are contaminated by measurement error, so that the observed log-price process is given by $X = Y + \epsilon$. In other words, we decompose the observed transaction price into the “true” (frictionless) one and a “noise” term which captures market microstructure effects. Finally, we assume that there are a total of $MT$ observations from the process $X$, consisting of $M$ intradaily observations for $T$ days, viz:

$$X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}, \quad t = 0, \ldots, T \text{ and } j = 1, \ldots, M.$$

(3)

Daily integrated volatility, is defined as $IV_t = \int _{t-1} ^t \sigma_s^2 ds, \quad t = 1, \ldots, T$. Since $IV_t$ is not observable, different realized measures, based on the sample $X_{t+j/M}$, are used as proxies for $IV_t$. Each realized measure, $RM_{t,M}$, will have an associated estimation error, as in (1).

Our objective is to construct a nonparametric estimator of the density and confidence intervals of integrated volatility at time $T+1$, conditional on actual information. We analyze the properties of both kernel based and local polynomial estimators. We start from Nadaraya-Watson estimators for conditional confidence intervals:

$$ \widehat{CI}_{T,M}(u_1, u_2 | RMT,M) = \widehat{F}_{RMT+1,M|RMT,M}(u_2 | RMT,M) - \widehat{F}_{RMT+1,M|RMT,M}(u_1 | RMT,M) $$
\[
\frac{1}{T} \sum_{t=0}^{T-1} 1_{\{u_1 \leq RM_{t+1,M} \leq u_2\}} K \left( \frac{RM_{t,M} - RM_{T,M}}{\xi} \right) ;
\]

and for conditional densities:
\[
\hat{f}_{T,M}(x|RM_{T,M}) = \hat{f}_{RM_{T+1,M|RM_{T,M}}}(x|RM_{T,M}) = \frac{1}{1} \sum_{t=0}^{T-1} K \left( \frac{RM_{t,M} - RM_{T,M}}{\xi_1} \right) K \left( \frac{RM_{t,M} - x}{\xi_2} \right).
\]

Here, \( K \) is a kernel function, and \( \xi, \xi_1 \) and \( \xi_2 \) are bandwidth parameters.

We need the following assumptions.

**Assumption A1:** \( IV_t \) is a strictly stationary \( \alpha \)-mixing process with mixing coefficients satisfying
\[
\sum_{j=1}^{\infty} j^\lambda \alpha_j^{1 - 2/\delta} < \infty, \text{ with } \lambda > 1 - 2/\delta \text{ and } \delta > 2 .
\]

**Assumption A2:** (i) The kernel \( K \) is a symmetric, nonnegative, continuous function with bounded support \([-\Delta, \Delta]\), at least twice differentiable on the interior of its support, satisfying \( \int K(s)ds = 1, \int sK(s)ds = 0 \). (ii) Let \( K^{(j)} \) be the \( j \)-th derivative of the kernel. Then, \( K^{(j)}(-\Delta) = K^{(j)}(\Delta) = 0 \), for \( j = 1, \ldots, J, J \geq 1 \).

**Assumption A3:** (i) \( f(\cdot) \) and, for any fixed \( x, f(x|\cdot) \) are absolutely continuous with respect to the Lebesgue measure in \( \mathbb{R}_+ \), and at least twice continuously differentiable. (ii) For any fixed \( x, u \) and \( z, f(z) > 0, f(x|z) > 0, \) and \( 0 < F(u|z) < 1 \).

**Assumption A4:** There exists a sequence \( b_M \), with \( b_M \to \infty \), as \( M \to \infty \), such that
\[
E \left( |N_{t,M}|^k \right) = O \left( b_M^{-k/2} \right) , \text{ for some } k \geq 2 .
\]

Assumption A1 requires the daily volatility process to be strong mixing. Of note is that the mixing coefficients of the integrated and of the instantaneous volatility process are of the same order of magnitude. In fact, let \( B_{\sigma^2,t_1} \) and \( B_{IV,t_1} \) be the sigma-fields generated by \( (\sigma_s^2, t_1 \leq s \leq t_2) \) and by \( \left( \int_{s-1}^{s} \sigma_u^2 du, t_1 \leq s \leq t_2 \right) \), respectively, and also define:

\[
\alpha_{\sigma^2}(m) = \sup_n \sup_{A_1 \in B_{\sigma^2,t_1}, A_2 \in B_{\sigma^2,n+m}} | \Pr (A_1 \cap A_2) - \Pr (A_1) \Pr (A_2) | ,
\]
\[
\alpha_{IV}(m) = \sup_n \sup_{A_1 \in B_{IV,t_1}, A_2 \in B_{IV,n+m}} | \Pr (A_1 \cap A_2) - \Pr (A_1) \Pr (A_2) | .
\]
Then, it follows that $\alpha_{IV}(m) \leq C\alpha_{\sigma^2}(m - 1)$, for some constant $C$. Sufficient conditions for A1 are provided by Meyn and Tweedie (1993, p.536), for the continuous semimartingale case, and by Masuda (2004, Section 3) for the case with jumps.

A2 and A3 are standard assumptions in the literature on nonparametric density estimation. We require a kernel function with a bounded support in order to avoid boundary bias problems at zero, which are dealt with using a boundary corrected kernel function (see, e.g., Müller, 1991).

Assumption A4 requires that the $k$–th moment of the measurement error decays to zero at a fast enough rate, in order to ensure that the feasible density estimators (based on realized estimators) are asymptotically equivalent to the unfeasible ones (based on the latent volatility). In Section 3 we shall provide primitive conditions under which A4 is satisfied by the most commonly used realized measures.

For conditional confidence intervals, we have the following result:

**Theorem 1.** Let A1-A4 hold. If $\xi \to 0$, $T\xi \to \infty$, $T\xi^5 \to 0$, and $T^{k+3}b_M^{-1}\xi \to 0$, then:

$$\sqrt{T\xi} \left( \widehat{CI}_{T,M}(u_1, u_2|RM_{T,M}) - CI(u_1, u_2|RM_{T,M}) \right) \overset{d}{\to} N(0, V(u_1, u_2)),$$

where

$$V(u_1, u_2) = \int K^2(u)d\mu \left( CI(u_1, u_2|RM_{T,M}) (1 - CI(u_1, u_2|RM_{T,M})) \right).$$

A consistent estimator for $V(u_1, u_2)$ can be obtained by replacing $f(RM_{T,M})$ and $CI(u_1, u_2|RM_{T,M})$ with $\hat{f}_{T,M}(RM_{T,M})$ and $\widehat{CI}_{T,M}(u_1, u_2|RM_{T,M})$, respectively.

Besides standard regularity conditions relating the sample size $T$ to the bandwidth parameter $\xi$, it is interesting to notice that Theorem 1 imposes an extra condition ($T^{k+3}b_M^{-1}\xi \to 0$) on the relative rate at which both $M$ and $T$ tend to infinity.
The key point in the proof of this theorem is to show the asymptotic equivalence between the estimator based on realized measures and that based on integrated volatility, that is to show that:

$$\frac{1}{T\xi} \sum_{t=0}^{T-1} \left(1_{\left\{ u_1 \leq R_{t+1,M} \leq u_2 \right\}} K \left( \frac{R_{t,M} - R_{T,M}}{\xi} \right) - 1_{\left\{ IV_t \leq u_2 \right\}} K \left( \frac{IV_t - R_{T,M}}{\xi} \right) \right) = o_p \left( \frac{1}{\sqrt{T\xi}} \right).$$

One difficulty arises because the measurement error enters in the indicator function, so that standard mean value expansions do not apply. As shown in detail in the Appendix, we proceed by conditioning on a subset on which $\sup_t |N_{t,M}|$ approaches zero at an appropriate rate, and show that the probability measure of this subset approaches one at rate $\sqrt{T\xi}$.

Turning now to our predictive density estimator, we have the following result.

**Theorem 2.** Let $A1-A4$ hold. If $\xi_1, \xi_2 \to 0$, $T\xi_1 \xi_2 \to \infty$, $T\xi_1^5 \xi_2 \to 0$, $T\xi_1 \xi_5 \to 0$, and $T^{1+1} b_M^{-1} \xi_1 \xi_2 \to 0$, then:

$$\sqrt{T\xi_1 \xi_2} \left( \hat{f}_{T,M}(x|R_{T,M}) - f(x|R_{T,M}) \right) \xrightarrow{d} N \left( 0, \frac{f(x|R_{T,M})}{f(R_{T,M})} \left( \int K^2(u)du \right)^2 \right).$$

A viable alternative to kernel based estimators is to use local linear estimators. One advantage of such estimators is that they do not suffer from the boundary problem. Local linear estimator of conditional confidence intervals are obtained from the following optimization problem:

$$\hat{\alpha}_{T,M}(u_1, u_2, R_{T,M}) = \arg \min_{\alpha} Z_{T,M}(\alpha; u_1, u_2, R_{T,M}),$$

where

$$Z_{T,M}(\alpha; u_1, u_2, R_{T,M}) = \frac{1}{T\xi} \sum_{t=0}^{T} \left(1_{\{u_1 \leq R_{t+1,M} \leq u_2\}} - \alpha_0 - \alpha_1 \left( R_{t,M} - R_{T,M} \right) \right)^2 K \left( \frac{R_{t,M} - R_{T,M}}{\xi} \right)$$

and $\alpha = (\alpha_0, \alpha_1)'$. The local linear estimator of the conditional confidence interval is given by $\hat{\alpha}_{0,T,M}(u_1, u_2, R_{T,M})$. These estimators for conditional distributions have
been recently used by Aït-Sahalia, Fan and Peng (2009) for testing the correct specification of diffusion models. Similarly, local linear conditional density estimation (see Fan, Yao and Tong, 1996), are derived from:

$$\hat{\beta}_{T,M}(x, RM_{T,M}) = \arg \min_{\beta} S_{T,M}(\beta; x, RM_{T,M}),$$

where

$$S_{T,M}(\beta; x, RM_{T,M}) = \frac{1}{T \xi_{1} \xi_{2}} \sum_{t=0}^{T} \left( K \left( \frac{RM_{t+1,M} - x}{\xi_{2}} \right) - \beta_{0} - \beta_{1} (RM_{t,M} - RM_{T,M}) \right)^{2} K \left( \frac{RM_{t,M} - RM_{T,M}}{\xi_{1}} \right),$$

and $\beta = (\beta_{0}, \beta_{1})'$. The conditional density is given by the estimator of the constant in the least square minimization above, $\hat{\beta}_{0,T,M}(x, RM_{T,M})$. We have the following result.

**Theorem 3.** Let A1-A4 hold.

(i) If $\xi \to 0$, $T \xi \to \infty$, $T \xi_{5} \to 0$, and $T^{\frac{k+3}{k-1}} b^{-1}_{M} \xi_{1} \to 0$, then:

$$\sqrt{T \xi} (\hat{\alpha}_{0,T,M}(u_{1}, u_{2}, RM_{T,M}) - CI(u_{1}, u_{2} | RM_{T,M})) \overset{d}{\to} N(0, V(u_{1}, u_{2})).$$

(ii) If $\xi_{1}, \xi_{2} \to 0$, $T \xi_{1} \xi_{2} \to \infty$, $T \xi_{1}^{5} \xi_{2} \to 0$, $T \xi_{1} \xi_{2}^{5} \to 0$, and $T^{\frac{k+3}{k-1}} b^{-1}_{M} \xi_{1} \xi_{2} \to 0$, then:

$$\sqrt{T \xi_{1} \xi_{2}} \left( \hat{\beta}_{0,T,M}(x, RM_{T,M}) - f(x | RM_{T,M}) \right) \overset{d}{\to} N \left( 0, \left( \frac{f(x | RM_{T,M})}{f(RM_{T,M})} \left( \int K^{2}(u) du \right)^{2} \right) \right).$$

The theorem shows that kernel and local linear estimators are asymptotically equivalent.

### 3 Applications to specific volatility estimators

We now provide primitive conditions on the moments of the drift, variance and noise, which ensure that Assumption A4 is satisfied by some commonly used realized measures, namely: Realized Volatility ($RV_{t,M}$, Andersen, Bollerslev, Diebold and Labys, 2001, and Barndorff-Nielsen and Shephard, 2002), Normalized Bipower and Tripower Variation ($BV_{t,M}$ and $TPV_{t,M}$, Barndorff-Nielsen and Shephard, 2004), Two Scale
Let $\mathcal{V}_{t,M}$ be defined in Lemma 1; see Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008).

**Lemma 1.** Let $Y_t$ follow (2) and $\epsilon$ be defined by (3). If $E\left((\sigma_t^2)^{2(k+\delta)}\right) < \infty$ and $E\left((\mu_t)^{2(k+\delta)}\right) < \infty$, with $\delta > 2$, then there is a sequence $b_M$, where $b_M \to \infty$ as $M \to \infty$, such that:

(i) If $J_t \equiv 0$ for all $t$ (no jumps) and $\epsilon \equiv 0$ (no microstructure noise), then $E\left(|\mathcal{V}_{t,M} - IV_t|^k\right) = O(b_M^{-k/2})$, with $b_M = M$.

(ii) If $\epsilon \equiv 0$, then $E\left(|\mathcal{V}_{t,M} - IV_t|^k\right) = O(b_M^{-k/2})$, $E\left(|TP\mathcal{V}_{t,M} - IV_t|^k\right) = O(b_M^{-k/2})$, with $b_M = M$.

(iii) If $J_t \equiv 0$ for all $t$, $\epsilon_t \sim i.i.d. (0, \sigma_t^2)$, $E(\epsilon_t^{2k}) < \infty$, $E(\epsilon_tY_t) = 0$, and $l/M^{1/3} = O(1)$, then $E\left(\left|\mathcal{V}_{t,M} - IV_t\right|^k\right) = O(b_M^{-k/2})$, with $b_M = M^{1/3}$.

(iv) If $J_t \equiv 0$ for all $t$, $\epsilon_t \sim i.i.d. (0, \sigma_t^2)$, $E(\epsilon_t^{2k}) < \infty$, $E(\epsilon_tY_t) = 0$, and $e/M^{1/2} = O(1)$, then $E\left(\left|\mathcal{V}_{t,e,M} - IV_t\right|^k\right) = O(b_M^{-k/2})$, with $b_M = M^{1/2}$.

(v) If $J_t \equiv 0$ for all $t$, $\epsilon_t \sim i.i.d. (0, \sigma_t^2)$, $E(\epsilon_t^{2k}) < \infty$, $E(\epsilon_tY_t) = 0$, $\kappa(0) = 1$, $\kappa(1) = \kappa(1)(0) = \kappa(1)(1) = 0$, and $H/M^{1/2} = O(1)$, then $E\left(\left|\mathcal{V}_{t,H,M} - IV_t\right|^k\right) = O(b_M^{-k/2})$, with $b_M = M^{1/2}$.

(vi) If $J_t \equiv 0$ for all $t$, $\epsilon_t$ is strictly stationary, $E(\epsilon_t^{4k}) < \infty$, $E(\epsilon_tY_t) = 0$, $\kappa(0) = 1$, $\kappa(1) = \kappa(1)(0) = \kappa(2)(0) = \kappa(1)(1) = 0$, $\epsilon_{t+j/M}$ is geometrically mixing in the sense that, for any $j$, there exists a constant $|\rho| < 1$ such that $E\left(\left|\epsilon_{t+j/M} \epsilon_{t+(j-s)/M}, \ldots \right|^k\right) \approx \rho^* \epsilon_{t+(j-s)/M}$ and $H/M^{1/2} = O(1)$, then $E\left(\left|\mathcal{V}_{t,H,M} - IV_t\right|^k\right) = O(b_M^{-k/2})$, with $b_M = M^{1/2}$. 

10
Lemma 1 shows that \( b_M \) grows with \( M \) at different rates across different realized volatility measures. Hence, the different rates of convergence of the volatility estimators are reflected in different regularity conditions (see also Corradi and Distaso, 2006). Moreover, part (vi) of the Lemma allows for some serial dependence in the microstructure noise, requiring an additional condition on the kernel function, i.e. \( \kappa^{(2)}(0) = 0 \). Such a condition is satisfied, for example, by fifth or higher order kernels. Some form of correlation between noise and price can be allowed, following the approach of Kalnina and Linton (2008) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008).

**Remark 1.** From a practical point of view, the asymptotic normality results stated in the Theorems are useful, as they facilitate the construction of confidence bands around estimated conditional densities and confidence intervals. The sort of empirical problem for which these results may be useful is the following. Suppose that we want to predict the probability that integrated volatility will take a value between \( IV_l \) and \( IV_u \), say, given actual information. Then, asymptotically, \( \Pr \left( (IV_l \leq IV_{T+1} \leq IV_u) | IV_T = RM_{T,M} \right) \) will fall in the interval \( \left( \hat{F}_{T,M}(IV_u|R M_{T,M}) - \hat{F}_{T,M}(IV_l|R M_{T,M}) \right) \pm \hat{V}^{1/2}(IV_l, IV_u) z_{\alpha/2} / \sqrt{T} \), with probability \( 1 - \alpha \), where \( \hat{V}(IV_l, IV_u) \) is defined in Theorem 1 and \( z_{\alpha/2} \) denotes the \( \alpha/2 \) quantile of a standard normal.

**Remark 2.** In empirical work, volatility is often modelled and predicted with ARMA models that are constructed using logs of realized volatility. For example, Andersen, Bollerslev, Diebold and Labys (2001, 2003) use the log of realized volatility for modelling and predicting stock returns and exchange rate volatility. According to these authors, one reason for using logs is that while the distribution of realized volatility is highly skewed to the right, the distribution of logged realized volatility is much closer to normal. It is immediate to see that a Taylor expansion of \( \log(RM_{t,M}) \) around \( IV_t \) ensures that \( \mathbb{E} \left( |\log(RM_{t,M}) - \log(IV_t)|^k \right) = O(b_{M}^{-k/2}) \). Therefore, the statements in
the theorems above hold in the case where we are interested in predictive densities and confidence intervals for the log of integrated volatility.

**Remark 3.** Meddahi (2003) has shown that integrated volatility is not Markovian. Hence, by including a conditioning set containing also past information, we may improve the accuracy of the predictive densities. For $RM_{t,M}^{(d)} = (RM_{t,M}, \ldots, RM_{t-(d-1),M})$, our conditional confidence interval estimators would be given by

$$
\hat{CI}_{T,M}(u_1, u_2|RM_{T,M}^{(d)}) = \frac{1}{T \xi^d} \sum_{t=d}^{T-1} \mathbb{1}_{\{u_1 \leq RM_{t+1,M} \leq u_2\}} K\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right),
$$

where $K$ is a $d-$dimensional kernel function. Extension of the results of the theorems to cover this general case is straightforward (and available upon request), but is not reported for notational simplicity.

To alleviate the curse of dimensionality, one may use (weighted) averages of past realized measures. This is the route followed in the empirical section.

## 4 Monte Carlo Results

In this section, we carry out a Monte Carlo experiment to assess the finite sample behavior of the conditional confidence interval estimator defined in (4) and studied in Theorem 1. In particular, for different experimental designs we use the six realized measures discussed in Section 3 to construct:

$$
G_{T,M}(u_1, u_2) = \hat{V}^{-1/2}(u_1, u_2) \sqrt{T \xi} \left(\hat{CI}_{T,M}(u_1, u_2|RM_{T,M}) - CI(u_1, u_2|RM_{T,M})\right).
$$

Our objective is to assess the empirical level properties of $G_{T,M}(u_1, u_2)$. In our experiments, data are generated according to the following data generating process (DGP):

$$
\begin{align*}
\text{d}Y_t &= (m - \sigma_t^2/2) \text{d}t + \text{d}z_t + \sigma_t \text{d}W_{1,t}, \\
\text{d} \sigma_t^2 &= \psi (v - \sigma_t^2) \text{d}t + \eta \sigma_t \text{d}W_{2,t},
\end{align*}
$$

(5)
where $W_{1,t}$ and $W_{2,t}$ are two correlated Brownian motions, with $\text{corr}(W_{1,t}, W_{2,t}) = \rho$. Following Aït-Sahalia and Mancini (2008), we set $m = 0.05, \psi = 5, \upsilon = 0.04, \eta = 0.5,$ and $\rho = -0.5$.

As we do not have a closed form expression for the distribution of integrated volatility implied by the DGP in (5), we need to rely on a simulation based approach. We begin by simulating $S$ paths of length 2 (given stationarity) from (5), using a Milstein scheme with a discrete interval $1/N$, keeping the conditioning value at period 1 fixed across simulations. We then construct confidence intervals from the empirical distribution of the simulated integrated variance. By setting $S$ and $N$ sufficiently large (3000 and 2880, respectively), the effects of both simulation and discretization error are negligible. This gives us a simulation based estimator of $CI(u_1, u_2 | RM_{T,M})$.

We then construct time series of length $T$ of realized volatility measures sampling the simulated data at frequency $1/M$. For the first day, we use, across all replications, the same draws used for the construction of the (simulation based) estimator of the confidence interval described above. In order to avoid boundary bias problems, we form $G_{T,M}(u_1, u_2)$ using Gaussian kernels on a logarithmic transformation of our daily series. In our base case (denoted by Case I), we simply set $X_{t+j/M} = Y_{t+j/M}$. In Case II, daily data are generated by adding microstructure noise. Namely, we generate $X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}$, where $\epsilon_{t+j/M} \sim \text{i.i.d. } N(0, \sigma^2_{\epsilon})$, and $\sigma^2_{\epsilon} = \{(0.005^2), (0.007^2), (0.014^2)\}$. A standard deviation of 0.007 corresponds to the case where the standard deviation of the noise is approximately 0.1% of the value of the asset price (this is the same percentage as that used in Aït-Sahalia and Mancini, 2008). Finally, in Case III, jumps are added by including an $\text{i.i.d. } N(0, 0.64a_{\text{jump}}\hat{\mu}_{IV})$ shock to the process for $Y_{t+j/M}$, where $a_{\text{jump}}$ is set equal to $\{3, 2, 1\}$, and $\hat{\mu}_{IV}$ is the average of (the log of) $IV_t$ over $S$. In this case, it is assumed that jumps arrive randomly with equal probability at any point in time, once every 5 days when $a_{\text{jump}} = 3$. 

13
once every 2 days $a_{jump} = 2$, and every day when $a_{jump} = 1$, on average.

We consider the interval $[u_1, u_2] = [\hat{\mu}_{IV} - \beta \hat{\sigma}_{IV}, \hat{\mu}_{IV} + \beta \hat{\sigma}_{IV}]$, where $\hat{\sigma}_{IV}$ is the standard deviation of (the log of) $IV_t$ over $S$, and $\beta = \{0.125, 0.250\}$, for different values of $T$ and $M$, i.e. $T = \{100, 300, 500\}$ and $M = \{72, 144, 288, 576\}$. Results are based upon 10,000 Monte Carlo iterations. As results for all three values of $T$ are qualitatively the same, we report our findings only for $T = 100$, for the sake of brevity.

In Tables 1-3, rejection frequencies are reported, using two-sided 5% and 10% nominal level critical values. The six columns of entries in the table contain results for $RV_{t,M}$, $BV_{t,M}$, $TPV_{t,M}$, $RV_{l,l,M}$, $\overline{RV}_{t,e,M}$, and $RK_{t,H,M}$, respectively. For construction of $RK_{t,H,M}$, we use the modified Tukey-Hanning kernel, i.e. $\kappa(x) = 0.5 \left(1 - \cos \pi \left(1 - x\right)^2\right)$, with $H$ chosen optimally according to Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). A number of clear conclusions emerge upon examination of the results.

Turning first to Table 1, where there is neither microstructure noise nor jumps, note that $RV_{t,M}$, $BV_{t,M}$, and $TPV_{t,M}$ perform approximately equally well for large values of $M$, although $RV_{t,M}$ performs marginally better than the other estimators in a number of instances, as might be expected. In particular, use of these estimators yields empirical sizes close to the nominal 5% and 10% levels in various cases, and there is a substantial improvement as both $M$ and $T$ increase. Indeed, in many cases the nominal size is achieved, or very nearly so; a finding which might be viewed as rather surprising given the small values of $M$ and $T$ used in our experiments. Overall, $RV_{t,M}$, $BV_{t,M}$ and $TPV_{t,M}$ yield more accurate confidence intervals than the other three (robust) measures, although improvements associated with using these three estimators drop off sharply for the largest two values of $M$. In particular, note that rejection frequencies at the nominal 10% level for $\overline{RV}_{t,l,M}$, $\overline{RV}_{t,e,M}$, and $RK_{t,H,M}$
are often 0.20-0.70 when $M = 72$ and 144, whereas rates for $RV_{t,M}$, $BV_{t,M}$, and $TPV_{t,M}$ are generally rather closer to 0.10. Indeed, empirical performance of $\widetilde{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$, and $RK_{t,H,M}$ is quite poor for very small values of $M$, and performance often worsens as $T$ increases (not reported in our tables), for fixed $M$. Nevertheless, the robust measures yield empirical rejection frequencies that improve quite quickly as $M$ increases, for fixed $T$. Additionally, $RK_{t,H,M}$ and $\widetilde{RV}_{t,e,M}$ perform substantially better than the other microstructure noise robust measure ($\hat{RV}_{t,l,M}$) in virtually all cases, although the relative difference in performance shrinks as $M$ increases. Finally, $RK_{t,H,M}$ and $\widetilde{RV}_{t,e,M}$ perform approximately equally well for all values of $M$ and $T$. Overall, there is clearly a need for reasonably large values of $M$ when implementing the microstructure robust realized measures in the current context. This is not surprising, given the slower rate of convergence of these estimators.

We now turn to Table 2, where microstructure noise is added to the frictionless price. It is immediate to see that $\hat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $RK_{t,H,M}$ are superior to the non robust realized measures, for large values of $M$, as expected. For example, consider Panel B in Table 2. The rejection frequencies at the nominal 10% level for $RV_{t,M}$ range from 0.16 up to 1.0, when $M = 288$, depending upon the magnitude of the noise volatility. On the other hand, comparable rejection frequencies for $\widetilde{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $RK_{t,H,M}$ range from 0.13-0.22, which indicates a marked improvement when using robust measures, as long as $M$ is large, even though $T$ is only 100. Of course, for $M$ too small, there is nothing to gain by using the robust measures. Indeed, for $M = 72$, $RV_{t,M}$ rejection frequencies are much closer to the nominal level than $\widetilde{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $RK_{t,H,M}$ rejection frequencies. This is hardly surprising, given that, from Lemma 1, $b_M$ grows as fast as $M$, in the case of $RV_{t,M}$, $BV_{t,M}$ and $TPV_{t,M}$, while it grows at a rate slower than $M$ in the case of microstructure noise robust realized measures. It follows that for empirical implementation, one may select
either a relatively small value of $M$, for which the microstructure noise effect is not too distorting, together with a non microstructure robust realized measure, or select a very large value of $M$ and a microstructure robust realized measure. Interestingly, we see in our experiments that there is little to choose between the best performing of our robust measures (i.e. $\tilde{RV}_{t,e,M}$ and $RK_{t,H,M}$), and that these two measures outperform $RV_{t,M}$ in many cases for values of $M$ as small as 144, which suggests that the relative gains associated with using robust measures are achieved very quickly as $M$ increases.

Finally, consider Table 3, where jumps are added to the price. $BV_{t,M}$ and $TPV_{t,M}$ yield similar results, both outperform all “noise-robust measures”, and their relative performance improves with the jump frequency.

In summary, the above results suggest that the asymptotic theory established in Section 2 yields reasonably sharp finite sample distributional approximations, even for small values of $T$ and $M$.

5 Empirical Illustration: Daily Volatility Predictive Intervals for Intel

In this section we construct and examine predictions of the conditional distribution of daily integrated volatility for Intel.

Data are taken from the Trade and Quotation database at the New York Stock Exchange. Our sample size covers 150 trading days starting from January 2 2002. From the original data set, we extracted 10 second and 5 minute interval data, using bid-ask midpoints and the last tick method (see Wasserfallen and Zimmermann, 1985). Provided that there is sufficient liquidity in the market, the 5 minute frequency seems to offer a reasonable compromise between minimizing the effect of microstructure noise and reaching a good approximation to integrated volatility (see Andersen, Boller-
slev, Diebold and Labys, 2001 and Andersen, Bollerslev and Lang, 1999). Hence, our choice of the two frequencies allows us to evaluate the effect of microstructure noise on the estimated predictive densities. A full trading day consists of 2340 (resp. 78) intraday returns calculated over an interval of ten seconds (resp. five minutes).

Once the different realized volatility estimators have been obtained, we have calculated predictive intervals using logs. This has the advantage of avoiding boundary bias problems. We have used a Gaussian kernel with the bandwidth chosen optimally as in Silverman (1986). Results are reported for the kernel based estimators. Local linear based results are very similar and are omitted for space reasons.

Our goal was to calculate the probability that volatility at time $T+1$ is larger than volatility at time $T$. We have done so based on a sample of $T = 100$ observations. Then we compared the prediction of the model with the actual realization at time $T + 1$. Given that our sample covers 150 days, we have a total of 50 out-of-sample comparisons. Results are reported in Table 4 for the volatility measures computed using 10 seconds returns and in Table 5 for those computed using 5 minutes returns. Since volatility has to be estimated (and is therefore subject to estimation error), for robustness purposes we have reported two out-of-sample checks: those based on the same realized measure as the one used in computing predictive intervals (column 3) and those based using a benchmark volatility measure ($RV$ using 5 minutes returns, column 4). Also, we have used two different conditioning values: the level of volatility at time $T$ and the average level of volatility over the last 5 days.

Finally, we have also used a different conditioning variable, namely realized semi-variance ($RS^-$), proposed by Barndorff-Nielsen, Kinnebrock and Shephard (2009). This was motivated by their empirical results highlighting the high informational content of such a measure of downside risk.

Several interesting conclusions emerge from analyzing the Tables. First, as ex-
pected, $RV$ and $TPV$ have better results when returns are computed every 5 minutes. Increasing the sampling frequency implies a higher noise to signal ratio and therefore non robust volatility estimators see a drop in forecasting directional changes.

Conversely, robust volatility estimators have a better performance at the higher sampling frequency. Again, this is not surprising, given that these estimators explicitly account for market microstructure noise and then it makes sense to use as many observations as possible.

Generally, conditioning on the average of volatility during the previous 5 days yields slightly better results than conditioning on the value at time $T$.

Finally, results seem to confirm the high informational content of realized semivariance. Using this measure of downside risk as a conditioning variable substantially increases directional predictability, with percentages of correct predictions as high as 0.763 (for $TSRV$ and $RK$ at 10 seconds), or 0.702 (using a benchmark volatility estimator for the out-of-sample check).
Appendix

For notational simplicity, hereafter let $u_1 = 0$ and $u_2 = u$. Also, we use $\sim$ to indicate “of the same order of magnitude”.

Proof of Theorem 1:

From Remark 6 in Hall, Wolff and Yao (1999), it follows that $\sqrt{T} \left( \hat{F}_T(u|R_{T,M}) - F(u|R_{T,M}) \right) \overset{d}{\to} N(0, V(u))$, where $V(u)$ is defined as in the statement of the theorem and we (henceforth) use the notation $\hat{F}_T(u|R_{T,M})$ to denote the unfeasible estimator, based on the latent integrated volatility.

We therefore need to show that $\sqrt{T} \left( \hat{F}_T(u|R_{T,M}) - F(u|R_{T,M}) \right) = o_p(1)$.

\[
\sqrt{T} \left( \hat{F}_T(u|R_{T,M}) - F(u|R_{T,M}) \right) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} \left( \mathbb{1}_{\{R_{t+1,1,M} \leq u\}} K \left( \frac{R_{t+1,M} - R_{T,M}}{\xi} \right) - \mathbb{1}_{\{IV_{t+1} \leq u\}} K \left( \frac{IV_t - R_{T,M}}{\xi} \right) \right). \tag{A.1}
\]

By Theorem 2.22 in Fan and Yao (2005), $\hat{F}_T(R_{T,M}) = f(R_{T,M}) + o_p(1)$ and, by A3, $f(R_{T,M}) > 0$. Turning to the numerator of (A.1), it can be expanded as

\[
\frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} \left( \mathbb{1}_{\{R_{t+1,1,M} \leq u\}} K \left( \frac{R_{t+1,M} - R_{T,M}}{\xi} \right) - \mathbb{1}_{\{IV_{t+1} \leq u\}} K \left( \frac{IV_t - R_{T,M}}{\xi} \right) \right). \tag{A.2}
\]

\[
+ \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} \left( \mathbb{1}_{\{IV_{t+1} \leq u\}} - \mathbb{1}_{\{R_{t+1,1,M} \leq u\}} \right) K \left( \frac{IV_t - R_{T,M}}{\xi} \right) \tag{A.3}
\]

\[
+ \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} \left( \mathbb{1}_{\{IV_{t+1} \leq u\}} - \mathbb{1}_{\{R_{t+1,1,M} \leq u\}} \right) \frac{K \left( \frac{R_{t+1,M} - R_{T,M}}{\xi} \right)}{K \left( \frac{IV_t - R_{T,M}}{\xi} \right)} \tag{A.4}
\]

After a first order Taylor expansion of the kernel around $IV_t$, the term in (A.2) can be written as:

\[
\frac{1}{\sqrt{T} \xi^2} \sum_{t=0}^{T-1} \mathbb{1}_{\{IV_{t+1} \leq u\}} K^{(1)} \left( \frac{IV_t - R_{T,M}}{\xi} \right) N_{t,M} \]

\[
+ O_p \left( \frac{1}{\sqrt{T} \xi^2} \sum_{t=0}^{T-1} \mathbb{1}_{\{IV_{t+1} \leq u\}} K^{(1)} \left( \frac{IV_t - R_{T,M}}{\xi} \right) N_{t,M} \right). \tag{A.5}
\]

Let $R_{t,M} = \xi^{-2} K^{(1)} \left( (IV_t - R_{T,M})/\xi \right)$. Then

\[
\left| \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{1}_{\{IV_{t+1} \leq u\}} R_{t,M} N_{t,M} \right| \leq \frac{1}{T} \sum_{t=0}^{T-1} |R_{t,M}| |N_{t,M}| \leq \sup_{t \leq T} |N_{t,M}| E |R_{t,M}|
\]

\[
+ \sup_{t \leq T} |N_{t,M}| \frac{1}{T} \sum_{t=0}^{T-1} (|R_{t,M}| - E |R_{t,M}|) = \sup_{t \leq T} |N_{t,M}| (O(1) + o_p(1)),
\]

because, by a change of variable and integration by part, $E \left( \left( R_{t,M} \right)^k \right) = O \left( (\xi^{-2})^{k-1} \right)$ and, by the central limit theorem,

\[
\frac{1}{T} \sum_{t=1}^{T-1} (|R_{t,M}| - E |R_{t,M}|) = O \left( (T \xi^2)^{-1/2} \right).
\]

19
Given assumption A4, for a positive and arbitrarily small $\varepsilon$, we have that

$$\Pr \left( \sup_{t \leq T} T^{-\frac{1}{12}} b_{M}^{1/2} |N_{t,M}| > \varepsilon \right) \leq \sum_{t=0}^{T-1} \Pr \left( T^{-\frac{1}{12}} b_{M}^{1/2} |N_{t,M}| > \varepsilon \right) \leq \frac{1}{\varepsilon} \sqrt{T} T^{-\frac{1}{2}} b_{M}^{1/2} \mathbb{E} \left( |N_{t,M}|^{k} \right) \rightarrow 0, \text{ as } T, M \rightarrow \infty.$$  

Therefore, $\sup_{t \leq T} |N_{t,M}| = O_{p} \left( T^{1/2} b_{M}^{-1/2} \right)$, and the term in (A.2) is $O_{p} \left( T^{1/2} b_{M}^{-1/2} \xi^{1/2} \right)$.

We now turn to (A.3), and note that it is of a smaller order of probability than

$$\frac{1}{T \varepsilon} \sum_{t=0}^{T-1} \left( 1 \{ \sup_{t \leq T} |N_{t+1,M}| \leq IV_{i+1} \leq u + \sup_{t \leq T} |N_{t+1,M}| \} \right) K \left( \frac{IV_{i} - RMT_{T,M}}{\xi} \right). \quad (A.5)$$

Let $\Omega_{T,M}$ be the complement of $\Omega_{T,M} = \{ \omega : T^{1/2} b_{M}^{-1/2} \sup_{t} |N_{t,M}| \leq \varepsilon \}$, and note that:

$$\lim_{T, M \rightarrow \infty} \sqrt{T \varepsilon} \Pr \left( \Omega_{T,M} \right) = \lim_{T, M \rightarrow \infty} \sqrt{T \varepsilon} \Pr \left( T^{1/2} b_{M}^{-1/2} \sup_{t} |N_{t,M}| > \varepsilon \right) \leq \lim_{T, M \rightarrow \infty} \sqrt{T \varepsilon} T^{-1/2} b_{M}^{k/2} O(b_{M}^{-k/2}) = 0.$$

Consequently, we can proceed by conditioning on $\Omega_{T,M}$. For all $\omega \in \Omega_{T,M}$, there exists a constant $c$, such that:

$$\frac{1}{T \varepsilon} \sum_{t=0}^{T-1} \left( 1 \{ \sup_{t \leq T} |N_{t+1,M}| \leq IV_{i+1} \leq u + \sup_{t \leq T} |N_{t+1,M}| \} \right) K \left( \frac{IV_{i} - RMT_{T,M}}{\xi} \right) \leq \frac{1}{T \varepsilon} \sum_{t=0}^{T-1} \left( 1 \{ \sup_{t \leq T} |N_{t+1,M}| \leq IV_{i+1} \leq u + \sup_{t \leq T} |N_{t+1,M}| \} \right) K \left( \frac{IV_{i} - RMT_{T,M}}{\xi} \right). \quad (A.6)$$

To simplify notation, let $d_{T,M} = c(b_{M}^{-1/2} T^{1/2})$. Then, using (A.5) and (A.6), for all $\omega \in \Omega_{T,M}$:

$$\frac{1}{T \varepsilon} \sum_{t=0}^{T-1} \left( 1 \{ IV_{i+1} \leq u \} - 1 \{ RMT_{t+1,M} \leq u \} \right) K \left( \frac{IV_{i} - RMT_{T,M}}{\xi} \right) \leq \frac{1}{T \varepsilon} \sum_{t=0}^{T-1} \left( 1 \{ IV_{i+1} \leq u \} - 1 \{ RMT_{T,M} \leq u \} \right) K \left( \frac{IV_{i} - RMT_{T,M}}{\xi} \right) \leq \frac{1}{T} \sum_{t=0}^{T-1} \left( 1 \{ IV_{i+1} \leq u \} - 1 \{ RMT_{T,M} \leq u \} \right) K \left( \frac{IV_{i} - RMT_{T,M}}{\xi} \right). \quad (A.7)$$

We start from (A.7). To simplify notation, we let $x = RMT_{T,M}$, $y_{0} = IV_{0}$, $z = IV_{1}$, and $y = (y_{0} - x) / \xi$. Given stationarity, we have that:

$$\frac{1}{\xi} \mathbb{E} \left( \left( 1 \{ IV_{i+1} \leq u \} - 1 \{ RMT_{T,M} \leq u \} \right) K \left( \frac{IV_{i} - RMT_{T,M}}{\xi} \right) \right) = \frac{1}{\xi} \int_{\mathbb{R}_{+}} \int_{u - d_{T,M}}^{u + d_{T,M}} K \left( \frac{y_{0} - x}{\xi} \right) f(z) dy_{0} dz = \int_{\mathbb{R}_{+}} \int_{u - d_{T,M}}^{u + d_{T,M}} K(y) f(x + \xi y, z) dy dz \quad (A.9)$$

$$= \int_{\mathbb{R}_{+}} K(y) dy \int_{u - d_{T,M}}^{u + d_{T,M}} f(x, z) dz (1 + O(\xi)) = O(d_{T,M})(1 + \xi).$$

20
It follows that (A.7) is $O(d_{T,M})$. Moving to (A.8), its variance is given by

$$E \left( \frac{1}{T \xi} \sum_{t=0}^{T-1} \left( I_{\{u - d_{T,M} \leq IV_{t+1} \leq u + d_{T,M} \}} \right) K \left( \frac{IV_t - RM_{T,M}}{\xi} \right) \right)^2 + O(d_{T,M}^2)$$

and the expectation above can be treated similarly to (A.9), yielding:

$$E \left( \frac{1}{T \xi} \sum_{t=0}^{T-1} \left( I_{\{u - d_{T,M} \leq IV_{t+1} \leq u + d_{T,M} \}} \right) K \left( \frac{IV_t - RM_{T,M}}{\xi} \right) \right)^2 = O\left( T^{-1} \xi^{-1} d_{T,M} \right) + O(d_{T,M}^2).$$

The sum of the terms in (A.7) and (A.8) is therefore $O\left( d_{T,M} + O \left( T^{-1/2} \xi^{-1/2} d_{T,M}^{1/2} \right) \right)$, and the term in (A.3) is $O_p(T^{1/2} \xi^{1/2} b_M^{1/2})$. Because (A.4) is of a smaller probability order than (A.2) and (A.3), it follows that (A.1) is $O_p(T^{\frac{1}{2}+\epsilon} \xi^{1/2} b_M^{-1/2})$. The statement in the theorem follows.

**Proof of Theorem 2:**

By Theorem 2.22 in Fan and Yao (2005),

$$\sqrt{T \xi} \xi_2 \left( \hat{f}_T(x|RM_{T,M}) - f(x|RM_{T,M}) \right) \overset{d}{\rightarrow} N \left( 0, \left( \frac{f(x|RM_{T,M})}{f(RM_{T,M})} \left( \int K^2(u)du \right) \right)^2 \right).$$

To prove the theorem, we need to show that

$$\sqrt{T \xi} \xi_2 \left( \hat{f}_T(x|RM_{T,M}) - f_T(x|RM_{T,M}) \right) \overset{(A.10)}{=} \sqrt{T \xi} \xi_2 \left( \frac{1}{\hat{f}_T} \sum_{t=0}^{T-1} \left( K \left( \frac{RM_{t+1,M} - RM_{T,M}}{\xi_2} \right) - K \left( \frac{IV_t - RM_{T,M}}{\xi_2} \right) \right) \right)$$

$$+ \frac{1}{\hat{f}_T(RM_{T,M})} - 1$$

$$\times \left( \frac{1}{\sqrt{T \xi} \xi_2} \sum_{t=0}^{T-1} \left( K \left( \frac{RM_{t,M} - RM_{T,M}}{\xi_2} \right) - K \left( \frac{IV_t - RM_{T,M}}{\xi_2} \right) \right) \right).$$

The second term on the right hand side (rhs) of (A.10) is of smaller probability order than the first, which can be treated similarly to the term in (A.2) in the proof of Theorem 1, and therefore is $O_p\left( T^{\frac{1}{2}+\epsilon} \xi^{1/2} b_M^{-1/2} \right)$. The statement in the theorem follows.

**Proof of Theorem 3:**

**Part (i).** From Remark 4, in Hall, Wolff and Yao (1999), $\sqrt{T \xi} \left( \hat{a}_{0,T}(u, IV_T) - a_{0,T}(u, IV_T) \right) \overset{d}{\rightarrow} N(0, V(u))$. The theorem is proved, given that $\sqrt{T \xi} \left( \hat{a}_{0,T}(u, IV_T) - \hat{a}_{0,T,M}(u, RM_{T,M}) \right) = o_p(1)$ follows straightforwardly using the same argument as in the proof of Theorem 1.

**Part (ii).** From Fan, Yao and Tong (1996, p.196),

$$\sqrt{T \xi} \xi_2 \left( \beta_{0,T}(x, IV_T) - \beta_{0,T}(x, IV_T) \right) \overset{d}{\rightarrow} N \left( 0, \frac{f(x|RM_{T,M})}{f(RM_{T,M})} \left( \int K^2(u)du \right) \right)^2$$

and the theorem is proved, because $\sqrt{T \xi} \xi_2 \left( \beta_{0,T,M}(x, RM_{T,M}) - \beta_{0,T}(x, IV_T) \right) = o_p(1)$ follows by the same argument as that used in Theorem 1.
Proof of Lemma 1:

Part (i). Realized volatility is defined as $RV_{t,M} = \sum_{j=1}^{M-1} \left( X_{t+(j+1)/M} - X_{t+j/M} \right)^2$. We begin by considering the case of zero drift. From Proposition 2.1 in Meddahi (2002),

$$
\sqrt{M} N_{t+1,M} = 2\sqrt{M} \sum_{i=0}^{M-1} \left( \sigma_{t+i/M}^2 \int_{t+i/M}^{t+(i+1)/M} \left( \int_{t+i/M}^{s} dW_u \right) dW_s \right) 
+ 2\sqrt{M} \sum_{i=0}^{M-1} \left( \sigma_{t+i/M} \int_{t+i/M}^{t+(i+1)/M} \left( \sigma_u - \sigma_{t+i/M} \right) dW_u \right) dW_s 
+ 2\sqrt{M} \sum_{i=0}^{M-1} \left( \int_{t+i/M}^{t+(i+1)/M} \left( \int_{t+i/M}^{s} dW_u \right) \sigma_{t+i/M} dW_s \right) 
+ 2\sqrt{M} \sum_{i=0}^{M-1} \left( \int_{t+i/M}^{t+(i+1)/M} \left( \sigma_u - \sigma_{t+i/M} \right) dW_u \right) \left( \sigma_s - \sigma_{t+i/M} \right) dW_s 
= 2\sqrt{M} \left( N_{t+1,M}^{(1)} + N_{t+1,M}^{(2)} + N_{t+1,M}^{(3)} + N_{t+1,M}^{(4)} \right).
$$

We consider the case of $k = 4$ ($k > 4$ can be treated in an analogous manner). Because of the Hölder continuity of a diffusion, $E \left( \left( \sqrt{M} N_{t+1,M}^{(1)} \right)^4 \right)$ is the term of largest order. To ease notation, let $\sum_{j_i} = \sum_{j_i=0}^{M-1}$ unless otherwise specified. Then:

$$
E \left( \left( \sqrt{M} N_{t+1,M}^{(1)} \right)^4 \right) = M^2 \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_4} \left[ \sigma_{t+j_1/M}^2 \sigma_{t+j_2/M}^2 \sigma_{t+j_3/M}^2 \sigma_{t+j_4/M}^2 
\times \left( \int_{t+j_1/M}^{t+(j_1+1)/M} \left( \int_{t+j_1/M}^{s} dW_u \right) dW_s \right) \left( \int_{t+j_2/M}^{t+(j_2+1)/M} \left( \int_{t+j_2/M}^{s} dW_u \right) dW_s \right) 
\times \left( \int_{t+j_3/M}^{t+(j_3+1)/M} \left( \int_{t+j_3/M}^{s} dW_u \right) dW_s \right) \left( \int_{t+j_4/M}^{t+(j_4+1)/M} \left( \int_{t+j_4/M}^{s} dW_u \right) dW_s \right) \right].
$$

For all $j_i > 0$, $i = 1, \ldots, 4$, $\int_{t+j_i/M}^{t+(j_i+1)/M} \left( \int_{t+j_i/M}^{s} dW_u \right) dW_s$ is a martingale difference sequence with respect to $\mathcal{F}_{t+j_i/M} = \sigma(W_s, s \leq t + j_i/M)$. By the law of iterated expectation, it follows that, when $j_1 \neq j_2 \neq j_3 \neq j_4$, $E(\sqrt{M} N_{t+1,M}^{(1)} \left( \mathcal{F}_{t+1/M} \right)^4) = 0$. Analogously, in the case $j_3 = j_4$, and $j_3 \neq j_2 \neq j_1$, if $j_3 < j_1$ and/or $j_3 < j_2$, then $E(\sqrt{M} N_{t+1,M}^{(1)} \left( \mathcal{F}_{t+1/M} \right)^4) = 0$. Next, consider the case of $j_3 > j_1, j_2$ and let $E_{t+j_3/M}$ denote the expectation conditional on $\mathcal{F}_{t+j_3/M}$. Because $E_{t+j_3/M} \left( \left( \int_{t+j_3/M}^{s} dW_u \right) dW_s \right)^2 = O(M^{-2})$, then by McLeish mixing inequalities it follows that

$$
E \left( \left( \sqrt{M} N_{t+1,M}^{(1)} \right)^4 \right) 
= M^2 \sum_{j_1} \sum_{j_2} \sum_{j_3} E \left[ \sigma_{t+j_1/M}^2 \sigma_{t+j_2/M}^2 \sigma_{t+j_3/M}^2 
\times \left( \int_{t+j_1/M}^{t+(j_1+1)/M} \left( \int_{t+j_1/M}^{s} dW_u \right) dW_s \right) \left( \int_{t+j_2/M}^{t+(j_2+1)/M} \left( \int_{t+j_2/M}^{s} dW_u \right) dW_s \right) 
\times \left( \int_{t+j_3/M}^{t+(j_3+1)/M} \left( \int_{t+j_3/M}^{s} dW_u \right) dW_s \right) \left( \int_{t+j_4/M}^{t+(j_4+1)/M} \left( \int_{t+j_4/M}^{s} dW_u \right) dW_s \right) \right] 
\approx \sum_{j_1} \sum_{j_2} E \left[ \sigma_{t+j_1/M}^2 \sigma_{t+j_2/M} \left( \int_{t+j_1/M}^{t+(j_1+1)/M} \left( \int_{t+j_1/M}^{s} dW_u \right) dW_s \right) \right].
$$
Because $\sigma_t^{(2+\delta)k} < \infty$, \( \sum_{j=3}^{\infty} n_{j-\max\{j_1, j_2\}}^{1/2-1/2\delta} \) is finite and \( \delta \) is defined in Lemma 1.

Next, suppose that \( j_1 = j_3 \) and \( j_2 = j_4 \), \( j_3 \neq j_4 \). Then, by the Cauchy-Schwartz inequality,

\[
E \left( \left( \int_{t}^{t+(j_1+1)/M} \left( \int_{t+j_1/M}^{t+j_1+1/M} dW_u \right) dW_s \right)^4 \right) \leq M^2 \sum_{j_1} \sum_{j_2} \left[ E \left( \sigma_{t+j_1/M}^{1/2} \sigma_{t+j_2/M}^{1/2} \right) \right]^{1/2} \times \left( E \left( \int_{t}^{t+(j_1+1)/M} \left( \int_{t+j_1/M}^{t+j_1+1/M} dW_u \right) dW_s \right)^4 \right) \leq O(1),
\]

given that \( E \left( \sigma_t^{(2+\delta)k} \right) < \infty \), \( \sum_{j=3}^{\infty} n_{j-\max\{j_1, j_2\}}^{1/2-1/2\delta} \) is finite and \( \delta \) is defined in Lemma 1.

We now analyze the case with drift. From Proposition 2.1 in Meddahi (2002), the contribution of the drift term to the measurement error, on an interval of length \( 1/M \), is given by:

\[
\sqrt{M} \sum_{j=0}^{M-1} \left( \int_{t+j/M}^{t+(j+1)/M} \mu_s ds \right)^2 + 2\sqrt{M} \sum_{j=0}^{M-1} \left( \int_{t+j/M}^{t+(j+1)/M} \mu_s ds \right) \left( \int_{t+j/M}^{t+(j+1)/M} \sigma_s dW_s \right),
\]

and its moments are of a smaller order than those of \( \sqrt{M} \sum_{j=0}^{M-1} \left( \int_{t+j/M}^{t+(j+1)/M} \sigma_u dW_u \right) \sigma_s dW_s \),

given that \( E \left( \mu_k \right) < \infty \).

**Part (ii).** \( TPV_{t+1/M} \) is given by:

\[
TPV_{t+1/M} = \left( \mu_{2/3} \right)^{-3} \sum_{j=1}^{M-3} \left| \Delta X_{t+(j+3)/M} \right|^{2/3} \left| \Delta X_{t+(j+2)/M} \right|^{2/3} \left| \Delta X_{t+(j+1)/M} \right|^{2/3}, \]

where \( \mu_k = E \left[ Z^k \right] \) and \( Z \) is a standard normal random variable. Using the results of Barndorff-Nielsen, Shephard and Winkel (2006, Section 3), we can ignore the contribution of the jump component. Let

\[
TPV_{t+(j+1)/M}(W) = \left( \mu_{2/3} \right)^{-3} \left| \Delta W_{t+(j+3)/M} \right|^{2/3} \left| \Delta W_{t+(j+2)/M} \right|^{2/3} \left| \Delta W_{t+(j+1)/M} \right|^{2/3}.
\]

From Theorem 5.1 and Lemma 5.2 in Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006), it follows that

\[
\sqrt{M} N_{t,M} = \sqrt{M} \left( TPV_{t+1/M} - \frac{M-3}{M} \sum_{j=1}^{M-3} \left( \int_{t+j/M}^{t+(j+1)/M} \right) \sigma^2_s ds \right) = \sqrt{M} \sum_{j=1}^{M-3} \left( \sigma_{t+j/M}^2 \left( TPV_{t+(j+1)/M}(W) - \frac{1}{M} \right) \right) + o_P(1).
\]

Because \( \sigma_{t+j/M}^2 \left( TPV_{t+(j+1)/M}(W) - \frac{1}{M} \right) \) is a martingale difference sequence with respect to \( \mathcal{F}_{t+(j-2)/M} = \sigma \left( X_{t+j/M}, t < t+i/M \leq t+(j-2)/M \right) \), the result follows by the same argument as in part (i) of the Lemma. The proof also holds for bipower variation, given that the difference between bipower and tripower variation is \( O_p(M^{-1/2}) \).

**Part (iii).** The proof for \( TSRV \) is similar to that for \( MSRV \) below and is therefore omitted for space reasons.
Part (iv).

\[ \overline{RV}_{t,e,M} = \sum_{i=1}^{e} \frac{a_i}{i} \left( \sum_{j=1}^{M-1} (X_{t+(j+i)/M} - X_{t+j/M})^2 \right) + \frac{RV_{t,M}}{M}, \]  

where \( a_i = \frac{i}{e^2} \left( \frac{1}{2} - \frac{1}{e^2} \right). \)

The case of \( k = 2 \) is analyzed in A¨ıt-Sahalia, Mykland and Zhang (2006). Exploiting \( \sum_{i=1}^{e} \frac{a_i}{i} = 0 \) and \( \sum_{i=1}^{e} a_i = 1 \), we have that

\[ E \left( \left( \overline{RV}_{t,e,M} - IV_t \right)^k \right) \]

\[ = E \left[ \left( \left( \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-1} (Y_{t+(j+i)/M} - Y_{t+j/M})^2 - IV_t \right) - 2 \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-1} \epsilon_{t+(j+i)/M} \epsilon_{t+j/M} \right) \right. \]

\[ + 2 \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-1} (Y_{t+(j+i)/M} - Y_{t+j/M}) (\epsilon_{t+(j+i)/M} - \epsilon_{t+j/M}) \left. \right) - \left( \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=M-i}^{M} \left( \tilde{\epsilon}_{t+j/M}^2 - \sigma_t^2 \right) \right) \]

\[ + 2 (\tilde{\sigma}_t^2 - \sigma_t^2) \right]^k \].

It's enough to consider the \( k \)-th powers of the single elements of the rhs of (A.11) since, by Hölder inequality, the cross terms are of a smaller order.

It's easy to see that \( E \left( (\tilde{\sigma}_t^2 - \sigma_t^2)^k \right) = O(M^{-k/2}). \) Furthermore, because \( a_i \approx i^2/e^3, \)

\[ E \left( \left( \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-1} (\tilde{\epsilon}_{t+j/M}^2 - \sigma_t^2) \right)^k \right) \approx E \left( \left( \frac{1}{c} \sum_{j=1}^{M} (\tilde{\epsilon}_{t+j/M}^2 - \sigma_t^2) \right)^k \right) = O(e^{-k/2}), \]

so that the \( k \)-th moments of the fourth and fifth terms of the rhs of (A.11) are \( O(e^{-k/2}) \). The \( k \)-th moments of the first and third terms of the rhs of (A.11) can be analyzed using the same argument as in parts (i) and (iii) of the Lemma and, given that \( a_i/i = O(e^{-2}) \), are \( O(e^{-k/2}) \). Finally, the \( k \)-th moment of the second term of the rhs of (A.11) can be treated as the second term of the rhs of the last equality in (A.13) (in part (v) of the Lemma below), and is therefore \( O(e^{-k/2}) \). The statement follows for \( b_M = M^{1/2} \).

Part (v). \( RK_{t,H,M} = \gamma_{t,h}^{X} + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^{X} + \gamma_{t,-h}^{X} \right) \), where \( \kappa(0) = 1, \kappa(1) = 0, \) and

\[ \gamma_{t,h}^{X} = \sum_{j=1}^{M} (X_{t+j/M} - X_{t+(j-1)/M}) (X_{t+(j+h)/M} - X_{t+(j+h-1)/M}) \]  

(A.12)

The case of \( k = 2 \) has been established in Theorem 2 by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). Note that

\[ E \left( (RK_{t,H,M} - IV_t)^k \right) = E \left[ \left( \gamma_{t,0}^Y - IV_t \right) + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^Y + \gamma_{t,-h}^Y \right) \right. \]

\[ + \gamma_{t,0}^Y + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^Y + \gamma_{t,-h}^Y \right) \left( \gamma_{t,h}^Y + \gamma_{t,-h}^Y \right) \]  

\[ + \gamma_{t,0}^Y + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^Y + \gamma_{t,-h}^Y \right) \]  

\[ + \gamma_{t,0}^Y + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^Y + \gamma_{t,-h}^Y \right)^k \]  

\]
where \( Y_{t,h}^{\epsilon} = \sum_{j=1}^{M} (Y_{t+j/M} - Y_{t+(j-1)/M}) (\epsilon_{t+(j-h)/M} - \epsilon_{t+(j-1-h)/M}) \) and the other terms are defined in a similar fashion. Because \( \gamma_{t,0} + \sum_{h=1}^{H} \kappa (\frac{h-1}{H}) \left( \gamma_{t,h}^{\epsilon} + \gamma_{t,-h}^{\epsilon} \right) \) is the term of the highest order of probability, we only need to prove that \( E \left( \left( \gamma_{t,0} + \sum_{h=1}^{H} \kappa (\frac{h-1}{H}) \left( \gamma_{t,h}^{\epsilon} + \gamma_{t,-h}^{\epsilon} \right) \right)^{k} \right) = O \left( H^{-\frac{3k}{2}} M^{k/2} \right). \)

Let \( \gamma^{\epsilon} = \left( \gamma_{t,0}^{\epsilon}, \gamma_{t,1}^{\epsilon}, \gamma_{t,-1}^{\epsilon}, \ldots, \gamma_{t,h}^{\epsilon}, \gamma_{t,-h}^{\epsilon} \right) \), where for notational simplicity we have dropped the subscript \( t \). Following the proof of Theorem 1 in Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), we have that \( \gamma^{\epsilon} = \gamma_{V}^{\epsilon} + \gamma_{W}^{\epsilon} + \gamma_{Z}^{\epsilon} \), where

\[
\begin{align*}
\gamma_{V}^{\epsilon} &= 2 (V_0 - V_1, -V_0 + 2V_1 - V_2, \ldots, -V_{H-1} + 2V_H - V_{H+1})^{\epsilon}, \\
V_h &= \sum_{j=1}^{M-h-1} \epsilon_{j/M} (\epsilon_{(j+h)/M}), \\
\gamma_{W}^{\epsilon} &= (0, -W_2, \ldots, -W_{H-1} + 2W_H - W_{H+1})^{\epsilon}, \\
W_h &= \sum_{j=1}^{h-1} \epsilon_{j/M} (\epsilon_{(j-h)/M}) + \sum_{j=h+M-h}^{M} \epsilon_{j/M} (\epsilon_{(j+h)/M}), \\
\gamma_{Z}^{\epsilon} &= (Z_0 - 2Z_1, Z_1 - Z_0 + 3Z_1 - 2Z_2, \ldots, Z_{H-1} - Z_H - Z_{H+1} + 3Z_H - 2Z_{H+1})^{\epsilon}, \\
Z_h &= \epsilon_{1/M} (\epsilon_{h/M}) + \epsilon_{1} (\epsilon_{(M-h)/M}).
\end{align*}
\]

By Hölder inequality, all cross moments are of a smaller order. Hence, we only need to prove that, for \( w = (1, 1, \kappa (\frac{1}{H}), \ldots, \kappa (\frac{H-1}{H})) \), \( E(w^\epsilon)^k \), \( E(w^\epsilon W)^k \) and \( E(w^\epsilon Z)^k \) are \( O \left( H^{-k/2} \right) \). We begin with \( w^\epsilon \gamma_{V}^{\epsilon} \) and, after some algebra, get

\[
\begin{align*}
\frac{1}{2} w^\epsilon \gamma_{V}^{\epsilon} &= (V_0 - V_1) + \sum_{h=1}^{H} \kappa (\frac{h-1}{H}) (-V_{h-1} + 2V_h - V_{h+1}) \\
&= \left( 1 - \kappa (\frac{1}{H}) \right) V_1 + \sum_{h=1}^{H-2} \kappa (\frac{h-1}{H}) - 2\kappa (\frac{h}{H}) + \kappa (\frac{h+1}{H}) \right) V_{h+1} \\
&\quad + \left( 2\kappa (\frac{H-1}{H}) - \kappa (\frac{H-2}{H}) \right) V_H - \kappa (\frac{H-1}{H}) V_{H+1}. \tag{A.13}
\end{align*}
\]

Because \( \epsilon \) is i.i.d., \( E(V_k) = O(M^{k/2}) \), and \( E \left( \sum_{h=1}^{H} V_h \right)^k = O \left( H^{k/2} M^{k/2} \right) \). Therefore,

\[
\begin{align*}
\frac{1}{2^k} E \left( (w^\epsilon \gamma_{V}^{\epsilon})^k \right) \tag{A.14} \\
&\approx \left( 1 - \kappa (\frac{1}{H}) \right)^k E (V_k) \tag{A.15} \\
&\quad + \sum_{h_1=1}^{H} \sum_{h_2=1}^{H} \ldots \sum_{h_k=1}^{H} \left( \kappa (\frac{h_1-1}{H}) - 2\kappa (\frac{h_1}{H}) + \kappa (\frac{h_1+1}{H}) \right)^2 \ldots \\
&\quad \ldots \left( \kappa (\frac{h_{k-2}-1}{H}) - 2\kappa (\frac{h_{k-2}}{H}) + \kappa (\frac{h_{k-2}+1}{H}) \right)^2 E (V_{h_{k+1}}^2) \ldots E (V_{h_{k+2}}^2) \tag{A.16} \\
&\quad + \left( 2\kappa (\frac{H-1}{H}) - \kappa (\frac{H-2}{H}) \right)^k E (V_H^k) \tag{A.17} \\
&\quad + \kappa (\frac{H-1}{H})^k E (V_{H+1}^k). \tag{A.18}
\end{align*}
\]

Exploiting the properties of the kernel \( \kappa \), and applying standard Taylor expansions, it follows that

(A.15), (A.17) and (A.18) are \( O \left( M^{k/2} H^{-2k} \right) = O(H^{-1}) \), while (A.16) is \( O \left( H^{-\frac{3k}{2} M^{k/2}} \right) \). Hence,
(A.14) is $O\left(H^{-3/2} M^{k/2}\right)$. By a similar argument, $E\left((w'\gamma_{iW})^k\right)$ and $E\left((w'\gamma_{2W})^k\right)$ are $O\left(H^{-k/2}\right)$. Then, for $H = M^{1/2}$ and $b_M = M^{1/2}$ the statement follows.

**Part (ei).** The case of $k = 2$ has been established in Proposition 4 by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). We consider the case of $k = 4$ (the proof for higher $k$ follows straightforwardly). We need to show that $E\left((w'\gamma_{iW})^k\right)$ is $O\left(H^{-k/2}\right)$. Contrary to the i.i.d. case, $E(V_h) \neq 0$ and $E(V_h V_{h'}) \neq 0$ for $h \neq h'$. Without loss of generality, let $\epsilon_{ji/M} = \rho(\gamma_{j-1}/M + u_j/M)$, with $u_j/M \sim i.i.d. (0, \sigma^2)$. As a result of stationarity, we have suppressed the subscript $t$. After a Taylor expansion of the first two terms of the rhs of (A.13) around zero and of the last two terms around one, we obtain

$$
\frac{1}{2} w' \gamma_{iW} = \frac{\kappa(3)(0)}{H^3} V_1 + \frac{\kappa(3)(0)}{H^3} \sum_{h=1}^{H-2} h V_{h+1} + \frac{2\kappa(2)(1)}{H^2} V_H + \frac{\kappa(2)(1)}{2H^2} V_{H+1}, \tag{A.19}
$$

where $V_h$ is defined in part (ui) of the Lemma. We only consider the second term of the rhs of (A.19), given that the others are of a smaller order. We have that

$$
E\left(\left(\frac{\kappa(3)(0)}{H^3} \sum_{h=1}^{H} h V_h\right)^4\right) = \frac{\kappa(3)(0)^4}{H^4} \sum_{h_1=1}^{H} \sum_{h_2=1}^{h_1} \sum_{h_3=1}^{h_2} \sum_{h_4=1}^{h_3} E\left(\epsilon_{j_1/M} \epsilon_{j_2/M} \epsilon_{j_3/M} \epsilon_{j_4/M}\right). \tag{A.20}
$$

Clearly, the leading term of the sum is when $h_1 \neq h_2 \neq h_3 \neq h_4$. First, consider the case of $j_1 = j_2$ and $j_3 = j_4$, with $j_1 \neq j_3$. It’s easy to see that

$$
\frac{1}{H^4} \sum_{h_1=1}^{H} \sum_{h_2=1}^{h_1} \sum_{h_3=1}^{h_2} \sum_{h_4=1}^{h_3} E\left(\epsilon_{j_1/M} \epsilon_{j_2/M} \epsilon_{j_3/M} \epsilon_{j_4/M}\right) = O(H^{-2}).
$$

Next, we consider the case of $j_1 \neq j_2 \neq j_3 \neq j_4$ (the case of $j_1 = j_2 = j_3 \neq j_4$ follows by a similar argument). After a sequential application of the law of the iterated expectation, we have

$$
E\left(\left(\frac{\kappa(3)(0)}{H^3} \sum_{h=1}^{H} h V_h\right)^4\right) = \frac{\kappa(3)(0)^4}{H^4} \sum_{h_1=1}^{H} \sum_{h_2=1}^{h_1} \sum_{h_3=1}^{h_2} \sum_{h_4=1}^{h_3} \left(E\left(\epsilon_{j_1/M} E_{j_2/M} \left(\epsilon_{j_1+h_1}/M\right) E_{j_3/M} \left(\epsilon_{j_1+h_2}/M\right)\right) E_{j_4/M} \left(\epsilon_{j_1+h_3}/M\right) E_{j_4/M} \left(\epsilon_{j_1+h_4}/M\right)\right) \tag{A.21}
$$

The conditional expectations in (A.20) are easily computed. For example,

$$
E_{j_3+h_3/M} \left(\epsilon_{j_4/M}^2\right) = \rho^2(j_4-j_3-h_3) \epsilon_{j_3+h_3/M}^2 + O(1).
$$

The other conditional expectations can be calculated similarly. Plugging the expressions back in (A.20), after some algebra, we have

$$
E\left(\left(\frac{\kappa(3)(0)}{H^3} \sum_{h=1}^{H} h V_h\right)^4\right) = O(H^{-4}).
$$

The statement in the Lemma then follows.
References


Table 1: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case I: No Microstructure Noise or Jumps in DGP

<table>
<thead>
<tr>
<th>M</th>
<th>$RV_t,M$</th>
<th>$BV_t,M$</th>
<th>$TPV_t,M$</th>
<th>$\bar{RV}_{t,l,M}$</th>
<th>$\bar{RV}_{t,e,M}$</th>
<th>$RK_{t,H,M}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>0.095</td>
<td>0.105</td>
<td>0.122</td>
<td>0.455</td>
<td>0.203</td>
<td>0.200</td>
</tr>
<tr>
<td>144</td>
<td>0.090</td>
<td>0.089</td>
<td>0.094</td>
<td>0.257</td>
<td>0.142</td>
<td>0.126</td>
</tr>
<tr>
<td>288</td>
<td>0.082</td>
<td>0.084</td>
<td>0.082</td>
<td>0.156</td>
<td>0.106</td>
<td>0.098</td>
</tr>
<tr>
<td>576</td>
<td>0.078</td>
<td>0.083</td>
<td>0.080</td>
<td>0.103</td>
<td>0.088</td>
<td>0.089</td>
</tr>
</tbody>
</table>

Nominal Size = 5%

<table>
<thead>
<tr>
<th>Interval = $\bar{\mu}<em>{IV} + 0.125\hat{\sigma}</em>{IV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
</tr>
<tr>
<td>144</td>
</tr>
<tr>
<td>288</td>
</tr>
<tr>
<td>576</td>
</tr>
</tbody>
</table>

Nominal Size = 10%

<table>
<thead>
<tr>
<th>Interval = $\bar{\mu}<em>{IV} + 0.250\hat{\sigma}</em>{IV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
</tr>
<tr>
<td>144</td>
</tr>
<tr>
<td>288</td>
</tr>
<tr>
<td>576</td>
</tr>
</tbody>
</table>

Notes: Entries denote rejection frequencies based on the construction of $G_{T,M}(u_1, u_2)$. In particular, values of $G_{T,M}(u_1, u_2)$ are compared with 5% and 10% nominal size critical values of the standard normal distribution. We use “pseudo true” IV values in place of actual IV values when constructing $G_{T,M}(u_1, u_2)$, as discussed in Section 4. Results are reported for various realized measures, for different values of $M$, and two daily sample sizes. The interval over which the statistics are calculated is $[u_1, u_2] = [\bar{\mu}_{IV} - \beta\hat{\sigma}_{IV}, \bar{\mu}_{IV} + \beta\hat{\sigma}_{IV}]$, where $\bar{\mu}_{IV}$ and $\hat{\sigma}_{IV}$ are the mean and standard error of the pseudo true data, and $\beta = \{0.125, 0.250\}$. All experiments are based on samples of 100 daily observations and 10,000 Monte Carlo iterations. See Section 4 for further details.
Table 2: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case II: Microstructure Noise in DGP

### Panel A: Interval = $\hat{\mu}_{IV} + 0.125\hat{\sigma}_{IV}$

<table>
<thead>
<tr>
<th>M</th>
<th>RV_{t,M}</th>
<th>BV_{t,M}</th>
<th>TPV_{t,M}</th>
<th>RV_{t,l,M}</th>
<th>RV_{t,e,M}</th>
<th>RK_{t,H,M}</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>0.075</td>
<td>0.083</td>
<td>0.085</td>
<td>0.457</td>
<td>0.203</td>
<td>0.199</td>
</tr>
<tr>
<td>144</td>
<td>0.073</td>
<td>0.075</td>
<td>0.072</td>
<td>0.260</td>
<td>0.140</td>
<td>0.125</td>
</tr>
<tr>
<td>288</td>
<td>0.152</td>
<td>0.163</td>
<td>0.152</td>
<td>0.159</td>
<td>0.114</td>
<td>0.103</td>
</tr>
<tr>
<td>576</td>
<td>0.980</td>
<td>0.995</td>
<td>0.991</td>
<td>0.102</td>
<td>0.089</td>
<td>0.087</td>
</tr>
</tbody>
</table>

**Nominal Size = 10%**

| 72  | 0.121    | 0.129    | 0.127     | 0.530      | 0.258      | 0.256      |
| 144 | 0.118    | 0.123    | 0.120     | 0.323      | 0.190      | 0.173      |
| 288 | 0.196    | 0.210    | 0.197     | 0.210      | 0.158      | 0.149      |
| 576 | 0.983    | 0.996    | 0.992     | 0.150      | 0.138      | 0.130      |

### Panel B: Interval = $\hat{\mu}_{IV} + 0.250\hat{\sigma}_{IV}$

<table>
<thead>
<tr>
<th>M</th>
<th>RV_{t,M}</th>
<th>BV_{t,M}</th>
<th>TPV_{t,M}</th>
<th>RV_{t,l,M}</th>
<th>RV_{t,e,M}</th>
<th>RK_{t,H,M}</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>0.076</td>
<td>0.084</td>
<td>0.082</td>
<td>0.456</td>
<td>0.209</td>
<td>0.206</td>
</tr>
<tr>
<td>144</td>
<td>0.130</td>
<td>0.133</td>
<td>0.121</td>
<td>0.262</td>
<td>0.137</td>
<td>0.130</td>
</tr>
<tr>
<td>288</td>
<td>0.940</td>
<td>0.967</td>
<td>0.949</td>
<td>0.162</td>
<td>0.115</td>
<td>0.102</td>
</tr>
<tr>
<td>576</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.002</td>
<td>0.080</td>
<td>0.087</td>
</tr>
</tbody>
</table>

**Nominal Size = 10%**

| 72  | 0.123    | 0.127    | 0.129     | 0.532      | 0.269      | 0.262      |
| 144 | 0.167    | 0.175    | 0.163     | 0.326      | 0.190      | 0.179      |
| 288 | 0.971    | 0.971    | 0.955     | 0.214      | 0.148      | 0.148      |
| 576 | 1.000    | 1.000    | 1.000     | 1.000      | 1.000      | 1.000      |

Notes: See notes to Table 1. The interval over which the statistics are calculated is $[u_1, u_2] = [\hat{\mu}_{IV} - \hat{\sigma}_{IV}, \hat{\mu}_{IV} + \hat{\sigma}_{IV}]$, where $\hat{\mu}_{IV}$ and $\hat{\sigma}_{IV}$ are the mean and standard error of the pseudo true data, and $\beta = \{0.125, 0.250\}$. All experiments are based on samples of 100 daily observations and 10,000 Monte Carlo iterations.
Table 3: Conditional Confidence Interval Accuracy Assessment: Level Experiments

**Case III: Jumps in DGP**

### Panel A: Interval = $\hat{\mu}_{IV} + 0.125\hat{\sigma}_{IV}$

**Nominal Size = 5%**

<table>
<thead>
<tr>
<th></th>
<th>72</th>
<th>144</th>
<th>288</th>
<th>576</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RV$_{t,M}$</td>
<td>0.239</td>
<td>0.197</td>
<td>0.194</td>
<td>0.622</td>
</tr>
<tr>
<td>BV$_{t,M}$</td>
<td>0.196</td>
<td>0.136</td>
<td>0.131</td>
<td>0.317</td>
</tr>
<tr>
<td>TPV$_{t,M}$</td>
<td>0.126</td>
<td>0.122</td>
<td>0.243</td>
<td>0.237</td>
</tr>
<tr>
<td>$\hat{\mu}_{IV}$</td>
<td>0.239</td>
<td>0.197</td>
<td>0.194</td>
<td>0.622</td>
</tr>
<tr>
<td>$\hat{\sigma}_{IV}$</td>
<td>0.196</td>
<td>0.136</td>
<td>0.131</td>
<td>0.317</td>
</tr>
<tr>
<td>$\tilde{R}_{IV}$</td>
<td>0.239</td>
<td>0.197</td>
<td>0.194</td>
<td>0.622</td>
</tr>
<tr>
<td>$\tilde{V}_{t,M}$</td>
<td>0.196</td>
<td>0.136</td>
<td>0.131</td>
<td>0.317</td>
</tr>
<tr>
<td>$\hat{\mu}<em>{IV} + 0.125\hat{\sigma}</em>{IV}$</td>
<td>0.239</td>
<td>0.197</td>
<td>0.194</td>
<td>0.622</td>
</tr>
</tbody>
</table>

**Nominal Size = 10%**

<table>
<thead>
<tr>
<th></th>
<th>72</th>
<th>144</th>
<th>288</th>
<th>576</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RV$_{t,M}$</td>
<td>0.259</td>
<td>0.259</td>
<td>0.253</td>
<td>0.695</td>
</tr>
<tr>
<td>BV$_{t,M}$</td>
<td>0.204</td>
<td>0.204</td>
<td>0.204</td>
<td>0.509</td>
</tr>
<tr>
<td>TPV$_{t,M}$</td>
<td>0.192</td>
<td>0.192</td>
<td>0.192</td>
<td>0.390</td>
</tr>
<tr>
<td>$\hat{\mu}_{IV}$</td>
<td>0.259</td>
<td>0.259</td>
<td>0.253</td>
<td>0.695</td>
</tr>
<tr>
<td>$\hat{\sigma}_{IV}$</td>
<td>0.204</td>
<td>0.204</td>
<td>0.204</td>
<td>0.509</td>
</tr>
<tr>
<td>$\tilde{R}_{IV}$</td>
<td>0.259</td>
<td>0.259</td>
<td>0.253</td>
<td>0.695</td>
</tr>
<tr>
<td>$\tilde{V}_{t,M}$</td>
<td>0.204</td>
<td>0.204</td>
<td>0.204</td>
<td>0.509</td>
</tr>
<tr>
<td>$\hat{\mu}<em>{IV} + 0.125\hat{\sigma}</em>{IV}$</td>
<td>0.259</td>
<td>0.259</td>
<td>0.253</td>
<td>0.695</td>
</tr>
</tbody>
</table>

### Panel B: Interval = $\hat{\mu}_{IV} + 0.250\hat{\sigma}_{IV}$

**Nominal Size = 5%**

<table>
<thead>
<tr>
<th></th>
<th>72</th>
<th>144</th>
<th>288</th>
<th>576</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RV$_{t,M}$</td>
<td>0.471</td>
<td>0.224</td>
<td>0.206</td>
<td>0.645</td>
</tr>
<tr>
<td>BV$_{t,M}$</td>
<td>0.445</td>
<td>0.172</td>
<td>0.163</td>
<td>0.532</td>
</tr>
<tr>
<td>TPV$_{t,M}$</td>
<td>0.446</td>
<td>0.145</td>
<td>0.134</td>
<td>0.463</td>
</tr>
<tr>
<td>$\hat{\mu}_{IV}$</td>
<td>0.440</td>
<td>0.128</td>
<td>0.131</td>
<td>0.413</td>
</tr>
<tr>
<td>$\hat{\sigma}_{IV}$</td>
<td>0.471</td>
<td>0.224</td>
<td>0.206</td>
<td>0.645</td>
</tr>
<tr>
<td>$\tilde{R}_{IV}$</td>
<td>0.445</td>
<td>0.172</td>
<td>0.163</td>
<td>0.532</td>
</tr>
<tr>
<td>$\tilde{V}_{t,M}$</td>
<td>0.446</td>
<td>0.145</td>
<td>0.134</td>
<td>0.463</td>
</tr>
<tr>
<td>$\hat{\mu}<em>{IV} + 0.250\hat{\sigma}</em>{IV}$</td>
<td>0.440</td>
<td>0.128</td>
<td>0.131</td>
<td>0.413</td>
</tr>
</tbody>
</table>

**Nominal Size = 10%**

<table>
<thead>
<tr>
<th></th>
<th>72</th>
<th>144</th>
<th>288</th>
<th>576</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RV$_{t,M}$</td>
<td>0.471</td>
<td>0.224</td>
<td>0.206</td>
<td>0.645</td>
</tr>
<tr>
<td>BV$_{t,M}$</td>
<td>0.445</td>
<td>0.172</td>
<td>0.163</td>
<td>0.532</td>
</tr>
<tr>
<td>TPV$_{t,M}$</td>
<td>0.446</td>
<td>0.145</td>
<td>0.134</td>
<td>0.463</td>
</tr>
<tr>
<td>$\hat{\mu}_{IV}$</td>
<td>0.440</td>
<td>0.128</td>
<td>0.131</td>
<td>0.413</td>
</tr>
<tr>
<td>$\hat{\sigma}_{IV}$</td>
<td>0.471</td>
<td>0.224</td>
<td>0.206</td>
<td>0.645</td>
</tr>
<tr>
<td>$\tilde{R}_{IV}$</td>
<td>0.445</td>
<td>0.172</td>
<td>0.163</td>
<td>0.532</td>
</tr>
<tr>
<td>$\tilde{V}_{t,M}$</td>
<td>0.446</td>
<td>0.145</td>
<td>0.134</td>
<td>0.463</td>
</tr>
<tr>
<td>$\hat{\mu}<em>{IV} + 0.250\hat{\sigma}</em>{IV}$</td>
<td>0.440</td>
<td>0.128</td>
<td>0.131</td>
<td>0.413</td>
</tr>
</tbody>
</table>

### Notes:
- See notes to Table 1. The interval over which the statistics are calculated is $[\mu_1, \mu_2] = [\hat{\mu}_{IV} - \beta \hat{\sigma}_{IV}, \hat{\mu}_{IV} + \beta \hat{\sigma}_{IV}]$, where $\hat{\mu}_{IV}$ and $\hat{\sigma}_{IV}$ are the mean and standard error of the pseudo true data, and $\beta = \{0.125, 0.250\}$.
- All experiments are based on samples of 100 daily observations and 10,000 Monte Carlo iterations.
### Table 4: Directional predictions results: $M = 2340$.  

<table>
<thead>
<tr>
<th>Realized Measure</th>
<th>Conditioning Variable</th>
<th>Percentage of correct predictions using the same Realized Measure</th>
<th>Percentage of correct predictions using a benchmark Measure$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RV$</td>
<td>$RV_T$</td>
<td>0.521</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>$RV_T$</td>
<td>0.578</td>
<td>0.426</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.543</td>
<td>0.461</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.58</td>
<td>0.426</td>
</tr>
<tr>
<td>$TPV$</td>
<td>$TPV_T$</td>
<td>0.501</td>
<td>0.418</td>
</tr>
<tr>
<td></td>
<td>$TPV_T$</td>
<td>0.562</td>
<td>0.378</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.539</td>
<td>0.360</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.562</td>
<td>0.378</td>
</tr>
<tr>
<td>$TSRV$</td>
<td>$TSRV_T$</td>
<td>0.503</td>
<td>0.476</td>
</tr>
<tr>
<td></td>
<td>$TSRV_T$</td>
<td>0.602</td>
<td>0.544</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.763</td>
<td>0.702</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.659</td>
<td>0.602</td>
</tr>
<tr>
<td>$MSRV$</td>
<td>$MSRV_T$</td>
<td>0.522</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td>$MSRV_T$</td>
<td>0.618</td>
<td>0.563</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.744</td>
<td>0.682</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.569</td>
<td>0.603</td>
</tr>
<tr>
<td>$RK$</td>
<td>$RK_T$</td>
<td>0.522</td>
<td>0.461</td>
</tr>
<tr>
<td></td>
<td>$RK_T$</td>
<td>0.577</td>
<td>0.522</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.737</td>
<td>0.701</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.674</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>$MSRV_T$</td>
<td>0.541</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td>$MSRV_T$</td>
<td>0.620</td>
<td>0.579</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.661</td>
<td>0.662</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.618</td>
<td>0.620</td>
</tr>
</tbody>
</table>

$^a$ Notes: this Table reports the percentage of correct directional volatility predictions for different conditioning variables and different volatility estimators constructed using 10 seconds returns. In Column 2, the use of an overline denotes the fact that the conditioning value is taken an average over the previous 5 days ($T - 4$ to $T$). Column 3 reports results obtained using the same volatility measure for both predictive probabilities and out-of-sample checks. Column 4 reports results obtained using a benchmark measure ($RV$ at 5 minutes frequency) for the out-of-sample checks.

### Table 5: Directional predictions results: $M = 78$.  

<table>
<thead>
<tr>
<th>Realized Measure</th>
<th>Conditioning Variable</th>
<th>Percentage of correct predictions using the same Realized Measure</th>
<th>Percentage of correct predictions using a benchmark Measure$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RV$</td>
<td>$RV_T$</td>
<td>0.503</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td>$RV_T$</td>
<td>0.578</td>
<td>0.579</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.601</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.618</td>
<td>0.620</td>
</tr>
<tr>
<td>$TPV$</td>
<td>$TPV_T$</td>
<td>0.583</td>
<td>0.541</td>
</tr>
<tr>
<td></td>
<td>$TPV_T$</td>
<td>0.720</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.681</td>
<td>0.578</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.702</td>
<td>0.601</td>
</tr>
<tr>
<td>$TSRV$</td>
<td>$TSRV_T$</td>
<td>0.541</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td>$TSRV_T$</td>
<td>0.578</td>
<td>0.579</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.620</td>
<td>0.619</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.661</td>
<td>0.662</td>
</tr>
<tr>
<td>$MSRV$</td>
<td>$MSRV_T$</td>
<td>0.364</td>
<td>0.482</td>
</tr>
<tr>
<td></td>
<td>$MSRV_T$</td>
<td>0.639</td>
<td>0.561</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.617</td>
<td>0.584</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>.615</td>
<td>0.578</td>
</tr>
<tr>
<td>$RK$</td>
<td>$RK_T$</td>
<td>0.460</td>
<td>0.583</td>
</tr>
<tr>
<td></td>
<td>$RK_T$</td>
<td>0.704</td>
<td>0.601</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.681</td>
<td>0.655</td>
</tr>
<tr>
<td></td>
<td>$RS_T$</td>
<td>0.656</td>
<td>0.658</td>
</tr>
</tbody>
</table>

$^a$ Notes: this Table reports the percentage of correct directional volatility predictions for different conditioning variables and different volatility estimators constructed using 5 minutes returns. In Column 2, the use of an overline denotes the fact that the conditioning value is taken an average over the previous 5 days ($T - 4$ to $T$). Column 3 reports results obtained using the same volatility measure for both predictive probabilities and out-of-sample checks. Column 4 reports results obtained using a benchmark measure ($RV$ at 5 minutes frequency) for the out-of-sample checks.