# Predictive Inference for Integrated Volatility*

Valentina Corradi[†]
University of Warwick

Walter Distaso[‡]
Imperial College London

Norman R. Swanson[§]
Rutgers University

March 2009

**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 kernel estimators of the aforementioned quantities, based on different realized volatility measures. A set of sufficient conditions under which the estimators are asymptotically equivalent to their unfeasible counterparts, based on an unobservable volatility process, is provided. Asymptotic normality for the feasible estimators is then established. The finite sample properties of the suggested estimators are examined via a Monte Carlo study. Finally, based upon data from the New York Stock Exchange, an empirical application to volatility directional predictability is provided.

*Keywords.* Diffusions, integrated volatility, realized volatility measures, kernels, microstructure noise, conditional confidence intervals, jumps, prediction.

*JEL classification.* C22, C53, C14.

*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 $EC^2$ 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 and RES-062-23-0311, 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 oft 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 class of assets for which predictive conditional densities are relevant is that based on the use of 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, including variance and covariance swaps, overshooters, and up and downcrossers, for example (see, e.g., Carr and Lee, 2003). One of the reasons why volatility based products now form an important class of assets is the stylized fact that volatility is counter cyclical (see Schwert, 1989), suggesting the adoption of volatility exposure in order to reduce the riskiness of a portfolio.

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 poses a formidable challenge to the researcher, 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 and most widely used estimator of integrated volatility is realized volatility, concurrently proposed by Andersen, Bollerslev, Diebold and Labys (2001), and Barndorff-Nielsen and Shephard (2002).[1] 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. Important variants of realized volatility have subsequently

[1]Note that, consistent with the prevailing standard in the econometrics literature, we (mis-)use the term *volatility* in *integrated volatility*. In the financial engineering literature, *integrated volatility* is (more appropriately) called *integrated variance*.

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 power 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). The estimators due to the above authors remain consistent for integrated volatility, in the presence of jumps, and when observed prices are affected by microstructure noise. The cost of implementing these new robust estimators is either a loss of efficiency (jumps), or a slower rate of convergence (microstructure noise). 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. Exploiting the usual factorization of joint densities, our density estimator is derived as the ratio between a (nonparametric) kernel estimator of the joint density of current and future volatility, and a kernel estimator of the marginal density of current (and past) volatility. 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. Nevertheless, we require the diffusive part of the log-price process to be Brownian. In this sense, our approach might be viewed as semiparametric.

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, as each realized measure can be decomposed into integrated volatility, the object of interest, and an estimation error term. Formally,

$$RM_{t,M} = IV_t + N_{t,M}, \tag{1}$$

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 $M$ denote 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.

In this paper, 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. Given the differences in efficiency and in the rate of convergence among the considered volatility estimators, our finding that the regularity conditions vary across the different realized volatility measures is not surprising.

The idea of using a realized measure as a basis for predicting integrated volatility has been adopted in other papers. Andersen, Bollerslev, Diebold and Labys (2003), Andersen, Bollerslev and Meddahi (2004, 2005) deal with the problem of pointwise prediction of integrated volatility, using ARMA models based on the log of 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), Aït-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 (2008) 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 (2008) is not sufficient. This paper provides an answer to this sort of questions by establishing asymptotic normality for Nadaraya-Watson and local polynomial 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 differs from Corradi, Distaso and Swanson (2008) in two other important respects. First, instead of restricting our attention to the class of eigenfunction stochastic volatility models (see Meddahi, 2001), we consider 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. Second, we study the asymptotic behavior of our predictive interval and predictive density estimators when some of the assumptions are not satisfied. In particular, we analyze the effect on our estimators when volatility violates the strong-mixing assumption and instead exhibits long-range dependence, as is often observed in practice. We also analyze the consequences on conditional interval and density estimators when one uses inconsistent volatility estimators (e.g. realized volatility in the presence of jumps or microstructure noise).

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. This is done for various daily sample sizes and for a variety of intradaily frequencies. As expected, robust realized volatility measures yield substantially more accurate predictive intervals than the other measures, when data are subject to microstructure noise, for relatively large values of $M$. However, for small values of $M$, realized volatility performs the best; and in the presence of jumps, tripower variation is superior, as expected. In general, our experiment underscores the relative trade-offs between $T$ and $M$, under various different data generating assumptions.

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 the set-up. Section 3 outlines the conditional density and confidence interval estimators and establishes their asymptotic properties. Section 4 studies the applicability of the established asymptotic results to various well known realized measures, including realized volatility, power variation, two-scale and multiscale estimators and realized kernel estimators. We also study the behavior of confidence interval and predictive density estimators when we erroneously do not take into account the presence of jumps or microstructure effects. In Section 5, the results of a Monte-Carlo experiment designed to assess the finite sample accuracy of our asymptotic results are discussed. Section 6 contains an empirical illustration based upon data from the New York Stock Exchange. Finally, some concluding remarks are given in Section 7. All proofs are contained in the Appendix.

## 2 Setup

Denote the log-price of a financial asset at a continuous time $t$ as $Y_t$. Assume that the log-price process belongs to the class of Brownian semimartingale processes with jumps and write accordingly

$Y \in \mathcal{BSMJ}$. Then:

$$\mathrm{d}Y_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}J_t. \tag{2}$$

The drift $\mu_t$ is a predictable process; the diffusion term, $\sigma_t$, is a *cadlag* process, 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.

It is of interest to separate the discontinuous (due to jumps) part of $Y$, denoted by $Y^d$, from the continuous Brownian component, denoted by $Y^c$. It is well known that:

$$\langle Y \rangle_t = \langle Y^c \rangle_t + \left\langle Y^d \right\rangle_t ,$$

where $\langle \cdot \rangle$ denotes the quadratic variation process. In particular:

$$\langle Y^c \rangle_t = \int_0^t \sigma_s^2 \mathrm{d}s \ \text{ and } \ \left\langle Y^d \right\rangle_t = \sum_{0 \leq u \leq t} \Delta Y_u^2, \ \text{where } \Delta Y_t = Y_t - Y_{t-}.$$

The object of interest to the researcher is represented by the quantity on the left, integrated volatility. A special case of the class of Brownian semimartingales with jumps, which plays a key role in financial economics, is obtained when $J_t \equiv 0$, for all $t$. In this case, the log-price process belongs to the class of Brownian semimartingales and we write $Y \in \mathcal{BSM}$. This class includes the popular stochastic volatility models, which have been used extensively in theoretical and applied work.

Thus far, we have considered a market that is free from frictions. However, there is a substantial literature in financial economics that documents the presence of market distortion or friction, and that has identified several possible causes thereof (see, e.g., O'Hara, 1997). 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:

$$X = Y + \epsilon.$$

Thus, we assume that the observed transaction price can be decomposed into the "true" (efficient) price and a "noise" term which captures market microstructure effects. In order to properly manage financial risk, one is interested in the contribution to quadratic variation of the Brownian component of $Y$, hence the importance of considering volatility estimators which are consistent in the presence of market microstructure noise.

Now, in order to study integrated volatility using econometric tools, assume that there are a total of $MT$ equi-spaced observations from the process $X$, consisting of $M$ intradaily observations for $T$ days. More precisely, a sample of data is given by:

$$X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}, \ \ t = 0, \dots, T \ \text{ and } \ j = 1, \dots, M. \tag{3}$$

The object of interest, daily integrated volatility, is defined as:

$$IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s, \ \ t = 1, \dots, T. \tag{4}$$

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). In order to compute feasible estimators of conditional densities and confidence intervals, we provide sufficient conditions on the structure of the measurement error, $N_{t,M}$, ensuring that the distributions of the estimators based on realized measures and the corresponding distribution associated with the "true" (but latent) daily volatility process are asymptotically equivalent.

## 3 Main Theoretical Results

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. Extension to a finite forecast horizon $\tau > 1$ and to considering a conditioning set containing also past information is straightforward and is not considered for notational simplicity.

We analyze the properties of both kernel based and local polynomial estimators. We start from Nadaraya-Watson estimators for conditional confidence intervals:

$$\begin{aligned} &\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1|RM_{T,M}) \\ &= \frac{\frac{1}{T\xi}\sum_{t=0}^{T-1} 1_{\{u_1 \leq RM_{t+1,M} \leq u_2\}} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right)}{\frac{1}{T\xi}\sum_{t=0}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right)} \\ &= \frac{\frac{1}{T\xi}\sum_{t=0}^{T-1} 1_{\{u_1 \leq RM_{t+1,M} \leq u_2\}} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right)}{\widehat{f}_{RM_{T,M}}(RM_{T,M})}; \end{aligned} \tag{5}$$

and for conditional densities:

$$\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) = \frac{\frac{1}{T\xi_1\xi_2}\sum_{t=0}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_1}\right) K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)}{\frac{1}{T\xi_1}\sum_{t=0}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_1}\right)}.$$

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 $\sum_{j=1}^{\infty} j^{\lambda}\alpha_j^{1-2/\delta} < \infty$, with $\lambda > 1 - 2/\delta$ 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:

$$\int K(s)\mathrm{d}s = 1, \int sK(s)\mathrm{d}s = 0.$$

(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) $f_{IV_t}(\cdot)$ and, for any fixed $x$, $f_{IV_{t+1}|IV_t}(x|\cdot)$ 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 $RM_{t,M}$, $f_{IV_t}(RM_{t,M}) > 0$, $f_{IV_{t+1}|IV_t}(x|RM_{t,M}) > 0$, and

$0 < F_{IV_{t+1}|IV_t}(u|RM_{t,M}) < 1$.

**Assumption A4:** There exists a sequence $b_M$, with $b_M \to \infty$, as $M \to \infty$, such that:

$$\mathrm{E}\left(|N_{t,M}|^k\right) = O\left(b_M^{-k/2}\right), \text{ for some } k \geq 2.$$

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 $\mathcal{B}_{\sigma^2,t_1}^{t_2}$ and $\mathcal{B}_{IV,t_1}^{t_2}$ be the sigma-fields generated by $\left(\sigma_s^2, t_1 \leq s \leq t_2\right)$ and by $\left(\int_{s-1}^{s} \sigma_u^2 \mathrm{d}u, t_1 \leq s \leq t_2\right)$, respectively, and also define:

$$\alpha_{\sigma^2}(m) = \sup_n \sup_{A_1 \in \mathcal{B}_{\sigma^2,-\infty}^{n}, A_2 \in \mathcal{B}_{\sigma^2,n+m}^{\infty}} \left|\Pr\left(A_1 \cap A_2\right) - \Pr\left(A_1\right)\Pr\left(A_2\right)\right|$$

and

$$\alpha_{IV}(m) = \sup_n \sup_{A_1 \in \mathcal{B}_{IV,-\infty}^{n}, A_2 \in \mathcal{B}_{IV,n+m}^{\infty}} \left|\Pr\left(A_1 \cap A_2\right) - \Pr\left(A_1\right)\Pr\left(A_2\right)\right|.$$

Then, it is immediate to see that $\alpha_{IV}(m) \leq C\alpha_{\sigma^2}(m-1)$, for some constant $C$. Easy to verify 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.

Clearly, A1 does not rule out the possibility that volatility is a measurable function of the price of the underlying asset.

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. To take this into account, it suffices to use 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 4 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* $T^{\frac{k+3}{k}} b_M^{-1} \xi \to 0$, *then:*

$$\begin{aligned}&\sqrt{T\xi}\left(\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1|RM_{T,M})\right)\right.\\ &\left. - \left(F_{IV_{T+1}|IV_T}(u_2|RM_{T,M}) - F_{IV_{T+1}|IV_T}(u_1|RM_{T,M})\right)\right) \xrightarrow{d} \mathrm{N}\left(0, V(u_1,u_2)\right),\end{aligned}$$

*where*

$$\begin{aligned}V(u_1,u_2) = &\frac{\int K^2(u)\mathrm{d}u}{f_{IV_T}(RM_{T,M})}\left(\left(F_{IV_{T+1}|IV_T}(u_2|RM_{T,M}) - F_{IV_{T+1}|IV_T}(u_1|RM_{T,M})\right)\right.\\ &\left.\times\left(1 - \left(\left(F_{IV_{T+1}|IV_T}(u_2|RM_{T,M}) - F_{IV_{T+1}|IV_T}(u_1|RM_{T,M})\right)\right)\right)\right).\end{aligned}$$

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 ($T^{\frac{2k+6}{2k}} b_M^{-1} \xi \to 0$) on the relative rate at which both $M$ and $T$ tend to infinity.

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

$$\begin{aligned}&\widehat{V}^{-1/2}(u_1,u_2)\sqrt{T\xi}\left(\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1|RM_{T,M})\right)\right.\\ &\left. - \left(F_{IV_{T+1}|IV_T}(u_2|RM_{T,M}) - F_{IV_{T+1}|IV_T}(u_1|RM_{T,M})\right)\right) \xrightarrow{d} \mathrm{N}\left(0, 1\right),\end{aligned}$$

*where*

$$\begin{aligned}\widehat{V}(u_1,u_2) = &\frac{\int K^2(u)\mathrm{d}u}{\widehat{f}_{RM_{T,M}}(RM_{T,M})}\left(\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1|RM_{T,M})\right)\right.\\ &\left.\times\left(1 - \left(\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1|RM_{T,M})\right)\right)\right)\right).\end{aligned}$$

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:

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

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* $\xi_1,\xi_2\to 0$, $T\xi_1\xi_2\to\infty$, $T\xi_1^5\xi_2\to 0$, $T\xi_1\xi_2^5\to 0$, *and* $T^{\frac{k+1}{k-1}}b_M^{-1}\xi_1\xi_2\to 0$, *then:*

$$\begin{aligned}&\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})-f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right)\\&\xrightarrow{d}\mathrm{N}\left(0,\left(\frac{f_{IV_{T+1}|IV_T}(x|RM_{T,M})}{f_{IV_T}(RM_{T,M})}\left(\int K^2(u)\mathrm{d}u\right)^2\right)\right).\end{aligned}$$

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

$$\begin{aligned}&\left(\frac{\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})}{\widehat{f}_{RM_{T,M}}(RM_{T,M})}\left(\int K^2(u)\mathrm{d}u\right)^2\right)^{-1/2}\\&\times\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{RM_{T+1,M}|RM_T}(x|RM_{T,M})-f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right)\xrightarrow{d}\mathrm{N}(0,1).\end{aligned}$$

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:

$$\widehat{\boldsymbol{\alpha}}_{T,M}(u_1,u_2,RM_{T,M})=\arg\min_{\boldsymbol{\alpha}} Z_{T,M}(\boldsymbol{\alpha};u_1,u_2,RM_{T,M}),$$

where

$$\begin{aligned}&Z_{T,M}(\boldsymbol{\alpha};u_1,u_2,RM_{T,M})\\&=\frac{1}{T\xi}\sum_{t=0}^{T}\left(1_{\{u_1\le RM_{t+1,M}\le u_2\}}-\alpha_0-\alpha_1\left(RM_{t,M}-RM_{T,M}\right)\right)^2K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right)\end{aligned}$$

and $\boldsymbol{\alpha}=(\alpha_0,\alpha_1)'$. The local linear estimator of the conditional confidence interval is given by $\widehat{\alpha}_{0,T,M}(u_1,u_2,RM_{T,M})$. These estimators for conditional distributions have been recently used by

Aït-Sahalia, Fan and Peng (2009), for testing the correct specification of diffusion models. Such an estimator is not ensured to lie between 0 and 1 in finite samples. More complex estimators based, for example, on logistic approximations, do lie between 0 and 1 for any sample size (see, e.g., Hall, Wolff and Yao, 1999). However, they typically cannot be written in closed form.

Similarly, local linear conditional density estimation (see Fan, Yao and Tong, 1996), are obtained as:

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

where

$$\begin{aligned} & S_{T,M}(\boldsymbol{\beta}; x, RM_{T,M}) \\ =& \frac{1}{T\xi_1\xi_2} \sum_{t=0}^{T} \left( K\left( \frac{RM_{t+1,M} - x}{\xi_2} \right) - \beta_0 - \beta_1 \left( RM_{t,M} - RM_{T,M} \right) \right)^2 K\left( \frac{RM_{t,M} - RM_{T,M}}{\xi_1} \right), \end{aligned}$$

and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. The conditional density is given by the estimator of the constant in the least square minimization above, $\widehat{\beta}_{0,T,M}(x, RM_{T,M})$.

We have the following result.

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

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

$$\begin{aligned} & \sqrt{T\xi} \left( \widehat{\alpha}_{0,T,M}(u_1, u_2, RM_{T,M}) - \left( F_{IV_{T+1}|IV_T}(u_2|RM_{T,M}) - F_{IV_{T+1}|IV_T}(u_1|RM_{T,M}) \right) \right) \\ & \xrightarrow{d} \mathrm{N}\left(0, V(u_1, u_2)\right). \end{aligned}$$

*(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* $T^{\frac{k+1}{k-1}} b_M^{-1} \xi_1\xi_2 \to 0$, *then:*

$$\begin{aligned} & \sqrt{T\xi_1\xi_2} \left( \widehat{\beta}_{0,T,M}(x, RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M}) \right) \\ \xrightarrow{d} & \mathrm{N}\left( 0, \left( \frac{f_{IV_{T+1}|IV_T}(x|RM_{T,M})}{f_{IV_T}(RM_{T,M})} \left( \int K^2(u)\mathrm{d}u \right)^2 \right) \right). \end{aligned}$$

From the theorem above, it is immediate to see that standard kernel and local linear estimators are asymptotically equivalent.

# 4 Applications to specific volatility estimators

We now provide primitive conditions on the moments of the drift and variance terms, as well as on the noise, which ensure that Assumption A4 is satisfied by some commonly used realized measures, which we will evaluate in the Monte Carlo and in the empirical application, namely [2]: *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 Realized Volatility* ($\widehat{RV}_{t,l,M}$, Zhang, Mykland and Aït-Sahalia, 2005), *Multi Scale Realized Volatility* ($\widetilde{RV}_{t,e,M}$, Zhang, 2006), *Realized kernels* ($RK_{t,H,M}$, based on a kernel $\kappa$ 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 $\mathrm{E}\left((\sigma_t^2)^{2(k+\delta)}\right) < \infty$ and $\mathrm{E}\left((\mu_t)^{2(k+\delta)}\right) < \infty$, 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:*

$$\mathrm{E}\left(|RV_{t,M} - IV_t|^k\right) = O(b_M^{-k/2}), \text{ with } b_M = M.$$

*(ii) If $\epsilon \equiv 0$, then:*

$$\mathrm{E}\left(|BV_{t,M} - IV_t|^k\right) = O(b_M^{-k/2}), \ \mathrm{E}\left(|TPV_{t,M} - IV_t|^k\right) = O(b_M^{-k/2}) \text{ with } b_M = M.$$

*(iii) If $J_t \equiv 0$ for all $t$, $\epsilon_t \sim i.i.d.\left(0, \sigma_\epsilon^2\right)$, $\mathrm{E}\left(\epsilon_t^{2k}\right) < \infty$, $\mathrm{E}\left(\epsilon_t Y_t\right) = 0$, and $l/M^{1/3} = O(1)$, then:*

$$\mathrm{E}\left(\left|\widehat{RV}_{t,l,M} - IV_t\right|^k\right) = O(b_M^{-k/2}) \text{ with } b_M = M^{1/3}.$$

*(iv) If $J_t \equiv 0$ for all $t$, $\epsilon_t \sim i.i.d.\left(0, \sigma_\epsilon^2\right)$, $\mathrm{E}\left(\epsilon_t^{2k}\right) < \infty$, $\mathrm{E}\left(\epsilon_t Y_t\right) = 0$, and $e/M^{1/2} = O(1)$, then:*

$$\mathrm{E}\left(\left|\widetilde{RV}_{t,e,M} - IV_t\right|^k\right) = O(b_M^{-k/2}) \text{ with } b_M = M^{1/2}.$$

*(v) If $J_t \equiv 0$ for all $t$, $\epsilon_t \sim i.i.d.\left(0, \sigma_\epsilon^2\right)$, $\mathrm{E}\left(\epsilon_t^{2k}\right) < \infty$, $\mathrm{E}\left(\epsilon_t Y_t\right) = 0$, $\kappa(0) = 1$, $\kappa(1) = \kappa^{(1)}(0) = \kappa^{(1)}(1) = 0$, and $H/M^{1/2} = O(1)$, then:*

$$\mathrm{E}\left(|RK_{t,H,M} - IV_t|^k\right) = O(b_M^{-k/2}) \text{ with } b_M = M^{1/2}.$$

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

$$\mathrm{E}\left(|RK_{t,H,M} - IV_t|^k\right) = O(b_M^{-k/2}) \text{ with } b_M = M^{1/2}.$$

[2] In Corradi and Distaso (2006), Propositions 1-3, it is shown that $\mathrm{E}(|N_{t,M}|) = O\left(b_M^{-k/2}\right)$, for $k = 2, 3, 4$, in the case of realized volatility, bipower variation and two-scale realized volatility, when the DGP is a eigenfunction stochastic volatility model.

From Lemma 1, one can see 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.

In Lemma 1, part (iii)-(v), we have considered the case of noise independently distributed and uncorrelated with the "true" efficient price. In part (vi) we allow for some dependence in the microstructure noise. In order to do that, we require 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. We thus extend the results by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), who show that, under the same assumptions on the degree of dependence and kernel type, $\mathrm{E}\left(|RK_{t,H,M} - IV_t|^2\right) = O(M^{-1/2})$.

Finally, it should be mentioned that some form of correlation between noise and price can be allowed. For example, we can follow the approach of Kalnina and Linton (2008) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and decompose microstructure noise into two parts, making appropriate assumptions on the (endogenous) component of the noise which is correlated with the price.

## 4.1 Remarks

**Remark 1.** From a practical point of view, the asymptotic normality results stated in Theorems 1, 2 and 3 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

$$\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(IV_u|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(IV_l|RM_{T,M})\right) \pm \frac{\widehat{V}^{1/2}(IV_l, IV_u) z_{\alpha/2}}{\sqrt{T\xi}},$$

with probability $1-\alpha$, where $\widehat{V}(IV_l, IV_u)$ is defined in Corollary 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 $IV_t$ gives:

$$\log(RM_{t,M}) = \log(IV_t) + \frac{1}{IV_t}N_{t,M} - \frac{1}{2}\frac{1}{IV_t^2}N_{t,M}^2 + \frac{1}{3}\frac{1}{IV_t^3}N_{t,M}^3 + \dots.$$

Under the conditions in Lemma 1, it follows that $\mathrm{E}\left(|\log(RM_{t,M}) - \log(IV_t)|^k\right) = O(b_M^{-k/2})$. 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.** All our asymptotic results have been obtained under the assumption that the volatility process is short memory. This is required as one of the conditions for the measurement error to go to zero at a fast enough rate and for the application of the central limit theorem for kernel estimators. However, there is some evidence in the literature that intergated volatility is a long memory process. For example, Andersen, Bollerslev, Diebold and Labys (2001, 2003) suggest modelling realized volatility using an ARFIMA model with a differencing coefficient equal to 0.4. Coutin and Renault (2003) show that by allowing for long-memory in the volatility process one can explain some puzzles, such as steep volatility smiles in long-term options and co-movements between implied and realized volatility. They propose a new model in which long-memory volatility is obtained via fractional integration of a short memory volatility process (such as square root volatility or log-normal volatility, as in Comte and Renault, 1998). The resulting long memory spot volatility and integrated volatility are characterized by the same autocovariance structure as an ARFIMA$(p, d_0, q)$ where the differencing parameter, $d_0$, is the one used in the fractional integration of short memory spot volatility. Thus, from, e.g., Taqqu (1975), it follows that

$$\sum_{k=0}^{T} \mathrm{E}\left((IV_0 - \mathrm{E}\left(IV_0\right))(IV_{0+k} - \mathrm{E}\left(IV_0\right)\right) = O\left(T^{2d_0}L(T)\right),$$

where $L(T)$ is a slowly varying function, and $0 < d_0 < 1/2$. Now, it is well known that kernel estimators have the same limiting distribution regardless of whether we have *i.i.d.* or strong mixing observations (e.g., Theorem 2.2 in Fan and Yao, 2005, Ch.2). Hence, the question is whether the same holds also in the case of long-range dependence. Claeskens and Hall (2002) provide sufficient conditions on the memory degree under which the integrated mean square error is the same as in the *i.i.d.* case. In the Gaussian case (or in the case of smooth functions of Gaussian processes), such conditions are equivalent to

$$\xi \sum_{k=0}^{T} \mathrm{E}\left((IV_0 - \mathrm{E}\left(IV_0\right))(IV_{0+k} - \mathrm{E}\left(IV_0\right))\right) \to 0, \text{ as } T \to \infty. \tag{6}$$

Outside the case of (functions of) Gaussian processes, Claeskens and Hall's conditions imply, but are not implied by (6). Thus, for ARFIMA processes we need that $T^{2d_0}\xi L(T) \to 0$: the stronger is the memory (i.e. the higher is $d_0$), the faster the bandwidth should go to zero. For example, for $\xi = O(T^{-1/5})$, we need $d_0 < 1/10$. Now, empirical evidence suggests that $d_0$ is often between 0.3 and 0.4. For $d_0 = 0.4$, the long-variance condition is satisfied if $\xi \to 0$ at a rate faster than $T^{4/5}$, which implies a rate of convergence for the kernel density estimator slower than $T^{1/10}$. Therefore, unless $d_0$ is very small, the rate of convergence for the estimator becomes rather slow.

## 4.2 Using Non-Robust Measures in the Presence of Jumps or Microstructure Noise

From Lemma 1 it is immediate to see that, in the presence of either jumps or microstructure effects, the moments of $N_{t,M}$ go to zero at an appropriate rate, only if one utilizes robust volatility estimators. In the Monte Carlo section, we analyze the behavior of non-robust realized measures in the presence of jumps or microstructure noise. However, it is also worthwhile to analyze this case from a theoretical perspective.

The case of jumps is relatively straightforward. Given (2), by Barndorff-Nielsen and Shephard (2004), note that:

$$BV_{t,M} - \left( IV_t + \sum_{t-1 \le u \le t} \Delta Y_u^2 \right) = o_p(1).$$

If the jumps are *i.i.d.*, then $IV_t$ and $IV_t + \sum_{t-1 \le u \le t} \Delta Y_u^2$ share the same degree of memory. Therefore, using a volatility measure not robust to jumps has the effect of producing density (and distribution) estimators for the total quadratic variation, rather than its continuous part.

The case of microstructure noise contamination is more complex. We need to distinguish among three possible cases: (i) $\mathrm{var}\big(\epsilon_{t+i/M}\big) = o(M^{-1})$, (ii) $\mathrm{var}\big(\epsilon_{t+i/M}\big) = O(M^{-1})$ and (iii) $\mathrm{var}\big(\epsilon_{t+i/M}\big) = O(a_M)$, where as $M \to \infty$, $a_M \to \pi$, with $0 \le \pi < \infty$, and $Ma_M \to \infty$. Note that the case of noise with variance independent of the sampling frequency is included in (iii), for $\pi > 0$. In all other cases, the variance of the noise approaches 0, as $M \to \infty$, albeit at different speed. The order of magnitude of the microstructure noise variance is an open empirical issue. The evidence reports that microstructure noise is very small (for example, Bandi and Russell, 2007, report values ranging from 0.87e-07 to 2.1e-07). Based on this observation, Zhang, Mykland and Aït-Sahalia (2006) argue that the noise is indeed "too small" to be considered $O_p(1)$, and derive an Edgeworth correction for the two-scale estimator of Zhang, Mykland and Aït-Sahalia (2005), under the assumption that it approaches zero as $M \to \infty$.[3]

[3] Awartani, Corradi and Distaso (2009) suggest a test for the null hypothesis of a constant noise variance versus the alternative of a variance approaching zero, at a slower rate than $\sqrt{M}$. Empirical evidence based on the DJIA stocks suggests that, at frequencies of less than one minute, the variance of the noise tends to zero.

For brevity and notational simplicity, we consider the case of independent noise. However, all statements carry through to the case in which the measurement error is geometrically mixing, as in Lemma 1, part (vi). It is immediate to see that in case (i), the effect of the microstructure noise is asymptotically negligible, and all asymptotic results carry over as in the no noise case. It is also easy to see that case (ii) can be treated in the same way as the case of jumps. In fact, as $M \to \infty$,

$$RV_{t,M} \xrightarrow{p} IV_t + 2\sigma_\epsilon^2,$$

where $\sigma_\epsilon^2 = \lim_{M\to\infty} M\text{var}\big(\epsilon_{t+i/M}\big)$ . Again, using realized volatility in the presence of this type of noise has the effect of producing distribution estimators for the total quadratic variation.

Turning finally to case (iii), consider predictive confidence intervals. Predictive densities can be treated along the same lines. For each $T$, $RV_{t,M}$ diverges to infinity at rate $Ma_M$. Hence, for any finite $u \in \mathbb{R}_+$, $\widehat{F}_{RV_{T+1,M}|RV_{T,M}}(u|RV_{T,M})$ tends in probability to zero, and therefore is not a consistent estimator of the conditional distribution.

Hence, the scaled difference between the confidence intervals based on realized volatility and those based on integrated volatility diverges as $M, T \to \infty$. This is confirmed by the Monte Carlo findings; in fact, our results show that that, as $M$ grows, the empirical size gets close to one.

# 5 Monte Carlo Results

In this section, we carry out a series of Monte Carlo experiments in which we assess the finite sample behavior of the conditional confidence interval estimator defined in (5) and studied in Theorem 1 and Corollary 1. In particular, for a variety of different experimental designs we begin by using the six realized measures discussed in Section 4 to construct :

$$\begin{aligned} G_{T,M}(u_1,u_2) = \widehat{V}^{-1/2}(u_1,u_2)\sqrt{T\xi}\,\Big(\Big(&\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1|RM_{T,M})\Big) \\ &- \left(F_{IV_{T+1}|IV_T}(u_2|RM_{T,M}) - F_{IV_{T+1}|IV_T}(u_1|RM_{T,M})\right)\Big) \,. \end{aligned}$$

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):

$$\begin{aligned} \mathrm{d}Y_t &= (m - \sigma_t^2/2)\mathrm{d}t + \mathrm{d}z_t + \sigma_t \mathrm{d}W_{1,t}, \\ \mathrm{d}\sigma_t^2 &= \psi\left(\upsilon - \sigma_t^2\right)\mathrm{d}t + \eta\sigma_t\mathrm{d}W_{2,t}, \end{aligned} \tag{7}$$

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 (7), we need to rely on a simulation based approach. We begin by simulating $S$ paths of length 2 (given stationarity) from (7), with a discrete interval $1/N$, and we set $S$ and $N$ chosen to be larger than the number of days $T$ and the number of intraday observations $M$, respectively. Then, we use simulated spot variance to construct (simulated) integrated variance series. We finally construct *pseudo true* confidence intervals, in the sense that for $S$ and $N$ sufficiently large, both the estimation and discretization errors are negligible. All data are generated using the Milstein approximation scheme.

We construct time series of length $T$ of realized volatility measures sampling the simulated data at frequency $1/M$, for each day. Finally, 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 $X_{t+j/M} = Y_{t+j/M}$, for $t = 1, \ldots, T$, and $j = 1, \ldots, M$. In Case II, daily data are generated by adding microstructure noise. Namely, we generate $X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}$, $t = 1, \ldots, T$, and $j = 1, \ldots, M$, where $\epsilon_{t+j/M} \sim$ *i.i.d.* $\mathrm{N}(0, \sigma_\epsilon^2)$, with $\sigma_\epsilon^2 = \{(0.005^2), (0.007^2), (0.014^2)\}$, and where 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 *i.i.d.* $\mathrm{N}(0, 0.64 a_{jump} \widehat{\mu}_{IV_t})$ shock to the process for $Y_{t+j/M}$, where $a_{jump}$ is set equal to $\{3, 2, 1\}$, and $\widehat{\mu}_{IV_t}$ is the mean of our pseudo true logged $IV_t$ values. In this case, it is assumed that jumps arrive randomly with equal probability at any point in time, once within each 5 day interval when $a_{jump} = 3$, once within each 2 day interval when $a_{jump} = 2$, and once within each 1 day interval when $a_{jump} = 1$, on average.

We set $S = 3000$. In addition, we set the interval $[u1, u2] = [\widehat{\mu}_{IV_t} - \beta\widehat{\sigma}_{IV_t}, \widehat{\mu}_{IV_t} + \beta\widehat{\sigma}_{IV_t}]$, where $\widehat{\mu}_{IV_t}$ is defined above, $\widehat{\sigma}_{IV_t}$ is the standard error of the *pseudo true* data, and $\beta = \{0.125, 0.250\}$, corresponding to confidence intervals of 0.128 and 0.181. We consider different values of $T$ and $M$, i.e. $T = \{100, 300, 500\}$ and $M = \{72, 144, 288, 576\}$. To economize on the use of space, we only report results for $T = 100$ and $T = 300$ for the case of no noise and no jumps, and for $T = 100$ for cases in which either noise or jumps are present. All results are based upon 1000 Monte Carlo iterations.

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 $RV_{t,M}$, $BV_{t,M}$, $TPV_{t,M}$, $RV_{t,l,M}$, $\widetilde{RV}_{t,e,M}$, and $RK_{t,H,M}$, respectively. For construction of $RK_{t,H,M}$, we use the modified Tukey-Hanning kernel, i.e. $\kappa(x) = 0.5\left(1 - \cos\pi\,(1-x)^2\right)$, with $H$ chosen optimally according to Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). Results for different values of $M$ are reported in different rows of the tables. 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, although $RV_{t,M}$ performs marginally better 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. As expected, $RV_{t,M}$, $BV_{t,M}$ and $TPV_{t,M}$ yield more accurate confidence intervals than the other three (robust) measures. In particular, note that rejection frequencies at the nominal 10% level for $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$, and $RK_{t,H,M}$ are often 0.20-0.40 when $M = 72$ and 144, whereas rates for $RV_{t,M}$ and $BV_{t,M}$ are generally rather closer to 0.10. Indeed, empirical performance of $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$, and $RK_{t,H,M}$ is quite poor for very small values of $M$, and performance often worsens as $T$ increases, for fixed $M$. Nevertheless, it should be stressed that the robust measures clearly yield empirical rejection frequencies that improve quite quickly as $M$ increases, for fixed $T$. Moreover, the performance of $RK_{t,H,M}$ actually improves as $T$ increases, for fixed $M$, in cases other than when $M = 72$. Additionally, $RK_{t,H,M}$ performs substantially better than the other microstructure noise robust measures in virtually all cases, although the relative difference in performance shrinks as $M$ increases. In summary, there is clearly a need for reasonably large values of $M$ when implementing the microstructure robust realized measures in this 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 simulated efficient price. It is immediate to see that $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $RK_{t,H,M}$ are superior to the non microstructure "noise-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 0.152 up to 1, when $M = 288$, depending upon the magnitude of the noise volatility. On the other hand, comparable rejection frequencies for $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $RK_{t,H,M}$ range from 0.132-0.248, 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 $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $RK_{t,H,M}$ rejection frequencies. This is as expected, given Lemma 1. In particular, recall from Lemma 1 that $b_M$ 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 thus 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 the best performing of our robust measures at small values of $M$ (i.e. $RK_{t,H,M}$) outperforms $RV_{t,M}$ 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 simulated efficient price. Note that $BV_{t,M}$ and $TPV_{t,M}$ yield similar results, and that both outperform all "noise-robust" measures, as expected. As might be expected, when the number of jumps increases, the relative performance of the "jump-robust" estimators increases. For example, when there is a jump every day, then $TPV_{t,M}$ outperforms all measures, other than $BV_{t,M}$, regardless of the value of $M$, and outperforms $BV_{t,M}$ in 13 of 16 instances, when comparing different values of $M$ as well as nominal sizes. Moreover, when jumps are less frequent, the differences between $RV_{t,M}$ rejection frequencies and nominal levels are closer to the same differences for $BV_{t,M}$ and $TPV_{t,M}$, again as expected.

In summary, the above results suggest that the asymptotic theory established in Section 3 yields reasonably sharp finite sample distributional approximations, even for small values of $T$ and $M$.

# 6 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. The rest of this section is broken into two subsections, including a discussion of the data and a discussion of our empirical findings.

## 6.1 Data Description

Data are taken from the Trade and Quotation (TAQ) database at the New York Stock Exchange (NYSE). 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, Bollerslev, 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).

### 6.2 Empirical Application: Volatility directional predictability

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 semivariance ($RS^{-}$), 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 expected, $RV$ 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 .76 (for $TSRV$ and $RK$ at 10 seconds), or .70 (using a benchmark volatility estimator for the out-of-sample check).

# 7 Concluding Remarks

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 establish asymptotic normality for Nadaraya-Watson and local polynomial estimators of conditional confidence intervals and conditional densities for daily volatility. Given that volatility is latent, we base our density estimators on realized measures. Our results are obtained following two key steps. First, given an assumption on the rate of decay to zero of the moments of the measurement error, we provide sufficient conditions on the relative rate of growth of $M$, number of intraday observations, and $T$, the number of days, under which estimators based on realized measures are asymptotically equivalent to their unfeasible counterparts based on integrated volatility. Second, we provide primitive conditions, under which realized volatility, power variation and several microstructure robust measures satisfy the assumption on the rate of decay of the measurement error. Our results apply to a general class of *cadlag* volatility processes. The finite sample behavior of the suggested procedures is analyzed via a Monte Carlo experiment, which highlights how the relative performance of five important realized measures varies with the ratio $M/T$, and with the "size" of the microstructure noise component. Finally, an empirical application based on New York Stock Exchange data is provided. We find evidence of directional predictability in volatility and highlight the informational content of the different realized measures.

# Appendix

## Proof of Theorem 1:

For notational simplicity, let $u_1 = 0$ and $u_2 = u$. From Remark 6 in Hall, Wolff and Yao (1999), it follows that $\sqrt{T\xi}\left(\widehat{F}_{IV_{T+1}|IV_T}(u|RM_{T,M}) - F_{IV_{T+1}|IV_T}(u|RM_{T,M})\right) \xrightarrow{d} \mathrm{N}\left(0, V(u)\right)$, where $V(u)$ is defined as in the statement of the theorem. Thus, it suffices to show that:

$$\sqrt{T\xi}\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u|RM_{T,M}) - \widehat{F}_{IV_{T+1}|IV_T}(u|RM_{T,M})\right) = o_p(1).$$

Now,

$$\begin{aligned}
&\sqrt{T\xi}\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u|RM_{T,M}) - \widehat{F}_{IV_{T+1}|IV_T}(u|RM_{T,M})\right)\\
&= \frac{1}{\widehat{f}_{IV_T}(RM_{T,M})} \qquad\qquad (A.1)\\
&\times\left(\frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\left(1_{\{RM_{t+1,M}\le u\}}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - 1_{\{IV_{t+1}\le u\}}K\left(\frac{IV_t - RM_{T,M}}{\xi}\right)\right)\right)\\
&+\left(\frac{\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(K\left(\frac{IV_t-RM_{T,M}}{\xi}\right) - K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right)\right)}{\frac{1}{T\xi}\sum_{t=0}^{T-1}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right)\frac{1}{T\xi}\sum_{t=0}^{T-1}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)}\right)\\
&\times\left(\frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\left(1_{\{RM_{t+1,M}\le u\}}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - 1_{\{IV_{t+1}\le u\}}K\left(\frac{IV_t - RM_{T,M}}{\xi}\right)\right)\right). \qquad (A.2)
\end{aligned}$$

We consider the term in (A.1), since (A.2) is of a smaller order. Given A1-A2, by Theorem 2.22 in Fan and Yao (2005),

$$\widehat{f}_{IV_T}(RM_{T,M}) = f_{IV_T}(RM_{T,M}) + o_p(1)$$

and, by A3, $f_{IV_T}(RM_{t,M}) > 0$. It suffices to consider the numerator. It is immediate to see that:

$$\begin{aligned}
&\frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\left(1_{\{RM_{t+1,M}\le u\}}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - 1_{\{IV_{t+1}\le u\}}K\left(\frac{IV_t - RM_{T,M}}{\xi}\right)\right)\\
&= \frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}1_{\{IV_{t+1}\le u\}}\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - K\left(\frac{IV_t - RM_{T,M}}{\xi}\right)\right) \qquad (A.3)\\
&+ \frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\left(1_{\{IV_{t+1}\le u\}} - 1_{\{RM_{t+1,M}\le u\}}\right)K\left(\frac{IV_t - RM_{T,M}}{\xi}\right) \qquad (A.4)\\
&+ \frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\left(\left(1_{\{IV_{t+1}\le u\}} - 1_{\{RM_{t+1,M}\le u\}}\right)\right.\\
&\times\left.\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - K\left(\frac{IV_t - RM_{T,M}}{\xi}\right)\right)\right). \qquad (A.5)
\end{aligned}$$

With regard to the term in (A.3), note that:

$$\begin{aligned}
&\frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}1_{\{IV_{t+1}\le u\}}\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - K\left(\frac{IV_t - RM_{T,M}}{\xi}\right)\right)\\
&= \frac{\sqrt{T\xi}}{T\xi^2}\sum_{t=0}^{T-1}1_{\{IV_{t+1}\le u\}}K^{(1)}\left(\frac{\widetilde{RM}_{t,M}-RM_{T,M}}{\xi}\right)N_{t,M}
\end{aligned}$$

where $\widetilde{RM}_{t,M} \in (RM_{t,M}, IV_t)$. Let:

$$R_{t,M} = \frac{1}{\xi^2}K^{(1)}\left(\frac{\widetilde{RM}_{t,M}-RM_{T,M}}{\xi_1}\right).$$

Now,

$$\left|\frac{1}{T}\sum_{t=0}^{T-1}R_{t,M}1_{\{IV_{t+1}\le u\}}N_{t,M}\right| \le \frac{1}{T}\sum_{t=0}^{T-1}|R_{t,M}|\,|N_{t,M}| \le \sup_{t\le T}|N_{t,M}|\,\mathrm{E}\,|R_{t,M}|$$

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

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

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

Now, given assumption A4, for a positive and arbitrarily small $\varepsilon$,

$$\begin{aligned}\Pr\left(\sup_{t\leq T}T^{-\frac{1}{k-1}}b_M^{1/2}|N_{t,M}|>\varepsilon\right)&\leq\sum_{t=0}^{T-1}\Pr\left(T^{-\frac{1}{k-1}}b_M^{1/2}|N_{t,M}|>\varepsilon\right)\\&\leq\frac{1}{\varepsilon^k}TT^{-\frac{k}{k-1}}b_M^{k/2}\mathrm{E}\left(|N_{t,M}|^k\right)\to 0,\ \text{as } T,M\to\infty.\end{aligned}$$

Thus, $\sup_{t\leq T}|N_{t,M}|=O_p\left(T^{\frac{1}{k-1}}b_M^{-1/2}\right)$, and the term in (A.3) is $O_p\left(T^{\frac{k+1}{2(k-1)}}b_M^{-1/2}\xi^{1/2}\right)$.

In order to analyze (A.4), note that:

$$\begin{aligned}&\left|\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{IV_{t+1}\leq u\}}-1_{\{RM_{t+1,M}\leq u\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right|\\&\leq\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{u-\sup_{t\leq T}|N_{t+1,M}|\leq IV_{t+1}\leq u+\sup_{t\leq T}|N_{t+1,M}|\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right).\end{aligned}\tag{A.6}$$

Let $\Omega_{T,M}=\{\omega:T^{\frac{-3}{2k}}b_M^{1/2}\sup_t|N_{t,M}|\leq\varepsilon\}$, and note that:

$$\begin{aligned}\lim_{T,M\to\infty}\sqrt{T\xi}\left(\Pr\left(\Omega_{T,M}\right)-1\right)&=\lim_{T,M\to\infty}\sqrt{T\xi}\Pr\left(T^{\frac{-3}{2k}}b_M^{1/2}\sup_t|N_{t,M}|>\varepsilon\right)\\&\leq\lim_{T,M\to\infty}\sqrt{T\xi}TT^{-\frac{3k}{2k}}b_M^{k/2}O(b_M^{-k/2})=0.\end{aligned}$$

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

$$\begin{aligned}&\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{u-\sup_{t\leq T}|N_{t+1,M}|\leq IV_{t+1}\leq u+\sup_{t\leq T}|N_{t+1,M}|\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\\&\leq\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{u-c\varepsilon b_M^{-1/2}T^{3/2k}\leq IV_{t+1}\leq u+c\varepsilon b_M^{-1/2}T^{3/2k}\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right).\end{aligned}\tag{A.7}$$

To simplify notation, let $c\varepsilon b_M^{-1/2}T^{\frac{3}{2k}}=d_{T,M}$. Then, using (A.6) and (A.7), for all $\omega\in\Omega_{T,M}$:

$$\left|\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{IV_{t+1}\leq u\}}-1_{\{RM_{t+1,M}\leq u\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right|$$

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

$$=\frac{1}{T}\sum_{t=0}^{T-1}\mathrm{E}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\frac{1}{\xi}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right)\tag{A.8}$$

$$+\frac{1}{T}\sum_{t=0}^{T-1}\left(\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\frac{1}{\xi}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right.\tag{A.9}$$

$$\left.-\mathrm{E}\left(\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\frac{1}{\xi}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right)\right)\right).$$

We start from (A.8). As $RM_{T,M}$ is used as the evaluation point, without loss of generality let $RM_{T,M}=x$. Given stationarity, let $IV_0=y_0$, $IV_1=z$, and $y=(y_0-x)/\xi$. We then have that:

$$\frac{1}{\xi}\mathrm{E}\left(\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right)\tag{A.10}$$

$$
\begin{aligned}
&= \frac{1}{\xi}\int_{\mathbb{R}_+}\int_{u-d_{T,M}}^{u+d_{T,M}} K\left(\frac{y_0-x}{\xi}\right) f\left(y_0,z\right)\mathrm{d}y_0\mathrm{d}z \\
&= \int_{\mathbb{R}_+}\int_{u-d_{T,M}}^{u+d_{T,M}} K\left(y\right) f\left(x+\xi y,z\right)\mathrm{d}y\ \mathrm{d}z \\
&= \int_{\mathbb{R}_+} K\left(y\right)\mathrm{d}y\int_{u-d_{T,M}}^{u+d_{T,M}} f\left(x,z\right)\ \mathrm{d}z(1+O(\xi)) = O(d_{T,M})(1+\xi).
\end{aligned}
$$

Thus, (A.8) is $O(d_{T,M})$. The variance of (A.9) is given by

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

and the expectation above can be treated similarly to (A.10), yielding:

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

Thus, the sum of the terms in (A.8) and (A.9) is $O\left(d_{T,M}\right)+O\left(T^{-1/2}\xi^{-1/2}d_{T,M}^{1/2}\right)$, and the term in (A.4) is $O_p(T^{1/2}\xi^{1/2}d_{T,M})$. By noting that (A.5) is of smaller probability order than (A.3) and (A.4), it follows that (A.1) is $O_p(T^{\frac{3+k}{2k}}\xi^{1/2}b_M^{-1/2})$. The statement in the theorem follows. ■

## Proof of Theorem 2:

By Theorem 2.22 in Fan and Yao (2005),

$$
\begin{aligned}
&\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})-f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right) \\
&\xrightarrow{d}\mathrm{N}\left(0,\left(\frac{f_{IV_{T+1}|IV_T}(x|RM_{T,M})}{f_{IV_T}(RM_{T,M})}\left(\int K^2(u)\mathrm{d}u\right)^2\right)\right).
\end{aligned}
$$

Thus, it suffices to show that:

$$
\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})-\widehat{f}_{IV_{T+1}|IV_T}(x|RM_{T,M})\right)=o_p(1).
$$

Now,

$$
\begin{aligned}
&\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})-\widehat{f}_{IV_{T+1}|IV_T}(x|RM_{T,M})\right) \\
&=\frac{\sqrt{T\xi_1\xi_2}\left(\frac{1}{T\xi_1\xi_2}\sum_{t=0}^{T-1}\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_1}\right)K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)-K\left(\frac{IV_t-RM_{T,M}}{\xi_1}\right)K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\right)\right)}{\widehat{f}_{IV_T}(x|RM_{T,M})} \\
&+\left(\frac{1}{\widehat{f}_{RM_{T,M}}(RM_{T,M})}-\frac{1}{\widehat{f}_{IV_T}(RM_{T,M})}\right) \\
&\left(\frac{1}{\sqrt{T\xi_1\xi_2}}\sum_{t=0}^{T-1}\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_1}\right)K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)-K\left(\frac{IV_t-RM_{T,M}}{\xi_1}\right)K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\right)\right).
\end{aligned}
\quad (A.11)
$$

The first term on the right hand side of (A.11) can be treated the same way as the term in (A.3) in the proof of Theorem 1, and therefore it is $O_p\left(T^{\frac{k+1}{2(k-1)}}b_M^{-1/2}\xi_1^{1/2}\xi_2^{1/2}\right)$. Furthermore, the second term on the right hand side of (A.11) is of smaller probability order than the first. The statement in the theorem then follows. ■

## Proof of Theorem 3:

(i) As in the proof of Theorem 1, for notational simplicity let $u_1=0$ and $u_2=u$. From Remark 4, in Hall, Wolff and Yao (1999):

$$
\sqrt{T\xi}\left(\widehat{\alpha}_{0,T}(u,IV_T)-\alpha_{0,T}(u,IV_T)\right)\xrightarrow{d}\mathrm{N}\left(0,V(u)\right),
$$

where $V(u)$ is defined as in statement of the theorem. Thus, it suffices to show that:

$$
\sqrt{T\xi}\left(\widehat{\alpha}_{0,T}(u,IV_T)-\widehat{\alpha}_{0,T,M}(u,RM_{T,M})\right)=o_p(1).
$$

This follows straightforwardly by writing the formula of the local linear estimator and using the same argument as in the proof of Theorem 1.
(ii) Similarly to part (i), from Fan, Yao and Tong (1996, p.196),

$$\sqrt{T\xi_1\xi_2}\left(\widehat{\beta}_{0,T}(x,IV_T)-\beta_{0,T}(x,IV_T)\right)$$
$$\xrightarrow{d}\mathrm{N}\left(0,\left(\frac{f_{IV_{T+1}|IV_T}(x|RM_{T,M})}{f_{IV_T}(RM_{T,M})}\left(\int K^2(u)\mathrm{d}u\right)^2\right)\right).$$

Thus, it suffices to show that:

$$\sqrt{T\xi_1\xi_2}\left(\widehat{\beta}_{0,T,M}(x,RM_{T,M})-\widehat{\beta}_{0,T}(x,IV_T)\right)=o_p(1).$$

This follows by the same argument as that used in part (i). ■

## Proof of Lemma 1:

In the sequel, with the notation $\simeq$ we mean "of the same order of magnitude".
(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), $N_{t+1,M}=2\sum_{i=0}^{M-1}\left(\int_{t+i/M}^{t+(i+1)/M}\left(\int_{t+i/M}^{s}\sigma_u\mathrm{d}W_u\right)\sigma_s\mathrm{d}W_s\right)$. Thus,

$$\begin{aligned}\sqrt{M}N_{t+1,M}&=2\sqrt{M}\sum_{i=0}^{M-1}\left(\sigma^2_{t+i/M}\int_{t+i/M}^{t+(i+1)/M}\left(\int_{t+i/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\\&+2\sqrt{M}\sum_{i=0}^{M-1}\left(\sigma_{t+i/M}\int_{t+i/M}^{t+(i+1)/M}\left(\int_{t+i/M}^{s}\left(\sigma_u-\sigma_{t+i/M}\right)\mathrm{d}W_u\right)\mathrm{d}W_s\right)\\&+2\sqrt{M}\sum_{i=0}^{M-1}\left(\int_{t+i/M}^{t+(i+1)/M}\left(\sigma_u-\sigma_{t+i/M}\right)\left(\int_{t+i/M}^{s}\mathrm{d}W_u\right)\sigma_{t+i/M}\mathrm{d}W_s\right)\\&+2\sqrt{M}\sum_{i=0}^{M-1}\left(\int_{t+i/M}^{t+(i+1)/M}\left(\int_{t+i/M}^{s}\left(\sigma_u-\sigma_{t+i/M}\right)\mathrm{d}W_u\right)\left(\sigma_s-\sigma_{t+i/M}\right)\mathrm{d}W_s\right)\\&=2\sqrt{M}\left(N^{(1)}_{t+1,M}+N^{(2)}_{t+1,M}+N^{(3)}_{t+1,M}+N^{(4)}_{t+1,M}\right).\end{aligned}$$

Also, for the sake of notational simplicity, we consider the case of $k=4$ (the case of $k>4$ can be treated in an analogous manner). To ease notation, let $\sum_{j_i}=\sum_{j_i=0}^{M-1}$ unless otherwise specified. Then:

$$\begin{aligned}\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right)&=M^2\sum_{j_1}\sum_{j_2}\sum_{j_3}\sum_{j_4}\mathrm{E}\left[\sigma^2_{t+j_1/M}\sigma^2_{t+j_2/M}\sigma^2_{t+j_3/M}\sigma^2_{t+j_4/M}\right.\\&\times\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\\&\times\left.\left(\int_{t+j_3/M}^{t+(j_3+1)/M}\left(\int_{t+j_3/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\left(\int_{t+j_4/M}^{t+(j_4+1)/M}\left(\int_{t+j_4/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\right].\end{aligned}$$

For all $j_i>0$, $i=1,...,4$, $\int_{t+j_i/M}^{t+(j_i+1)/M}\left(\int_{t+j_i/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s$ is a martingale difference sequence with respect to $\mathcal{F}_{t+j_i/M}=\sigma\left(\sigma^2_s,W_s,s\leq t+j_i/M\right)$. Thus, by the law of iterated expectation, it follows that when $j_1\neq j_2\neq j_3\neq j_4$, $\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right)=0$.

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 $\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right)=0$. Next, consider the case of $j_3>j_1,j_2$. Let $\mathrm{E}_{t+j/M}$ be the expectation conditional on $\mathcal{F}_{t+j_i/M}$. Noting that $\mathrm{E}_{t+j_3/M}\left(\left(\int_{t+j_3/M}^{t+(j_3+1)/M}\left(\int_{t+j_3/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^2\right)=O(M^{-2})$, then by McLeish mixing inequalities it follows that

$$\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right)$$

$$
\begin{aligned}
&= M^2 \sum_{j_1} \sum_{j_2} \sum_{j_3} \mathrm{E}\left[\sigma^2_{t+j_1/M}\sigma^2_{t+j_2/M}\sigma^4_{t+j_3/M}\right.\\
&\times \left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\\
&\times \left.\mathrm{E}_{t+j_3/M}\left(\left(\int_{t+j_3/M}^{t+(j_3+1)/M}\left(\int_{t+j_3/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^2\right)\right]\\
&\simeq \sum_{j_1}\sum_{j_2}\mathrm{E}\left[\sigma^2_{t+j_1/M}\sigma^2_{t+j_2/M}\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\right.\\
&\left.\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\sum_{j_3}\mathrm{E}_{t+\max\{j_1,j_2\}/M}\left(\left|\sigma^4_{t+j_3/M}-\mathrm{E}\left(\sigma^4_{t+j_3/M}\right)\right|\right)\right]\\
&\leq \sum_{j_1}\sum_{j_2}\left(\mathrm{E}\left(\sigma^4_{t+j_1/M}\sigma^4_{t+j_2/M}\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^2\right.\right.\\
&\left.\left.\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^2\right)\right)^{1/2}\mathrm{E}\left(\sigma^4_{t+j_3/M}-\mathrm{E}\left(\sigma^4_{t+j_3/M}\right)\right)^{1/2\delta}\sum_{j_3}\alpha^{1/2-1/2\delta}_{|j3-\max\{j_1,j_2\}|}=O(1),
\end{aligned}
$$

given that $\mathrm{E}\left(\sigma_t^{(2+\delta)k}\right)<\infty$ and $\sum_{j_3}\alpha^{1/2-1/2\delta}_{|j_3-\max\{j_1,j_2\}|}<\infty$.

Now, suppose that $j_1=j_3$ and $j_2=j_4$, $j_3\neq j_4$. Then, by Cauchy-Schwartz inequality,

$$
\begin{aligned}
&\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right)\\
&\leq M^2\sum_{j_1}\sum_{j_2}\left[\left(\mathrm{E}\left(\sigma^8_{t+j_1/M}\sigma^8_{t+j_2/M}\right)\right)^{1/2}\right.\\
&\left.\left(\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^4\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^4\right)\right)^{1/2}\right]=O(1).
\end{aligned}
$$

As $\mathrm{E}_{t+j_2/M}\left(\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^3\right)=0$, in the case of $j_2=j_3=j_4$, then $\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right)=0$, regardless of whether $j_1>j_2$ or $j_1<2$. Finally, it is immediate to see that the fourth moment above in the case of $j_1=j_2=j_3=j_4$ cannot be of a larger order than that in the case of $j_1=j_3$ and $j_2=j_4$, with $j_3\neq j_4$. Because of the Hölder continuity of a diffusion, $\mathrm{E}\left(\left(\sqrt{M}N^{(i)}_{t+1,M}\right)^4\right)$ cannot be of larger order of magnitude than $\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right)$, for $i=2,3,4$.

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:

$$
\sqrt{M}\sum_{j=0}^{M-1}\left(\int_{t+j/M}^{t+(j+1)/M}\mu_s\mathrm{d}s\right)^2+2\sqrt{M}\sum_{j=0}^{M-1}\left(\int_{t+j/M}^{t+(j+1)/M}\mu_s\mathrm{d}s\right)\left(\int_{t+j/M}^{t+(j+1)/M}\sigma_s\mathrm{d}W_s\right),
$$

and its moments cannot be of larger order than those of $\sqrt{M}\sum_{j=0}^{M-1}\left(\int_{t+j/M}^{t+(j+1)/M}\left(\int_{t+j/M}^{s}\sigma_u\mathrm{d}W_u\right)\sigma_s\mathrm{d}W_s\right)$, given that $\mathrm{E}\left((\mu_t)^{2(k+\delta)}\right)<\infty$.

(ii) $TPV_{t,M}=\left(\mu_{2/3}\right)^{-3}\sum_{j=1}^{M-3}\left|\Delta X_{(j+3)/M}\right|^{2/3}\left|\Delta X_{(j+2)/M}\right|^{2/3}\left|\Delta X_{(j+1)/M}\right|^{2/3}$, where $\mu_k=\mathrm{E}\left|Z^k\right|$ and $Z$ is a standard normal random variable. By Barndorff-Nielsen, Shephard and Winkel (2006, Section 3), we can ignore the contribution of the jump component. Let

$$
\begin{aligned}
TPV_{t+(j+1)/M}(W)&=\left(\mu_{2/3}\right)^{-3}\left|\Delta W_{t+(j+3)/M}\right|^{2/3}\left|\Delta W_{t+(j+2)/M}\right|^{2/3}\left|\Delta W_{t+(j+1)/M}\right|^{2/3},\\
TPV_{t+(j+1)/M}(X)&=\left(\mu_{2/3}\right)^{-3}\left|\Delta X_{t+(j+3)/M}\right|^{2/3}\left|\Delta X_{t+(j+2)/M}\right|^{2/3}\left|\Delta X_{t+(j+1)/M}\right|^{2/3}
\end{aligned}
$$

and $\mathrm{E}_{t+j/M}$ be the expectation conditional on $\mathcal{F}_{t+j/M}=\sigma\left(X_{t+i/M},\ t<t+i/M\leq t+j/M\right)$. We can write

$$
\sqrt{M}N_{t,M}=\sqrt{M}\sum_{j=1}^{M-3}\left(TPV_{t+(j+1)/M}(X)-\int_{t+j/M}^{t+(j+1)/M}\sigma^2_s\mathrm{d}s\right)
$$

$$= \sqrt{M} \sum_{j=1}^{M-3} \left( TPV_{t+(j+1)/M}(X) - \mathrm{E}_{t+j/M} \left( TPV_{t+(j+1)/M}(X) \right) \right) \tag{A.12}$$

$$+ \sqrt{M} \sum_{j=1}^{M-3} \left( \mathrm{E}_{t+j/M} \left( TPV_{t+(j+1)/M}(X) \right) - \sigma^2_{t+j/M} \mathrm{E}_{t+j/M} \left( TPV_{t+(j+1)/M}(W) \right) \right) \tag{A.13}$$

$$+ \sqrt{M} \sum_{j=1}^{M-3} \left( \sigma^2_{t+j/M} \mathrm{E}_{t+j/M} \left( TPV_{t+(j+1)/M}(W) \right) - \int_{t+j/M}^{t+(j+1)/M} \sigma^2_s \mathrm{d}s \right). \tag{A.14}$$

From Theorem 5.1 and Lemma 5.2 in Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006), it follows that (A.13) and (A.14) are $o_p(1)$, and that (A.12) can be expressed as follows:

$$\sqrt{M} \sum_{j=1}^{M-3} \left( \sigma^2_{t+j/M} \left( TPV_{t+(j+1)/M}(W) - \frac{1}{M} \right) \right) + o_p(1).$$

As $\sigma^2_{t+j/M} \left( TPV_{t+(j+1)/M}(W) - \frac{1}{M} \right)$ is a martingale difference sequence with respsct to $\mathcal{F}_{t+(j-2)/M}$, the result comes by the same argument as in part (i). The proof holds also for bipower variation, given that the difference between bipower and tripower variation is $O_p(M^{-1/2})$.
(iii)

$$\widehat{RV}_{t,l,M} = RV^{avg}_{t,l,M} - 2l\widehat{\sigma}^2_\epsilon, \text{ where } \widehat{\sigma}^2_\epsilon = \frac{RV_{t,M}}{2M} = \frac{1}{2M} \sum_{j=1}^{M-1} \left( X_{t+j/M} - X_{t+(j-1)/M} \right)^2$$

and

$$RV^{avg}_{t,l,M} = \frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left( X_{t+(jB+b)/M} - X_{t+((j-1)B+b)/M} \right)^2.$$

Here, $Bl = M$, $l = O(M^{1/3})$, $l$ denotes the subsample size, and $B$ denotes the number of subsamples. We can write:

$$\mathrm{E}\left( \left( \widehat{RV}_{t,l,M} - IV_t \right)^k \right)$$

$$\simeq \mathrm{E}\left( \left( \frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left( Y_{t+(jB+b)/M} - Y_{t+((j-1)B+b)/M} \right)^2 - IV_t \right)^k \right) \tag{A.15}$$

$$+ \mathrm{E}\left( \left( \frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left( Y_{t+(jB+b)/M} - Y_{t+((j-1)B+b)/M} \right) \left( \epsilon_{t+(jB+b)/M} - \epsilon_{t+((j-1)B+b)/M} \right) \right)^k \right) \tag{A.16}$$

$$+ \mathrm{E}\left( \left( \frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left( \left( \epsilon_{t+(jB+b)/M} - \epsilon_{t+((j-1)B+b)/M} \right)^2 - 2\widehat{\sigma}^2_\epsilon \right) \right)^k \right) \tag{A.17}$$

As $2\widehat{\sigma}^2_\epsilon - \mathrm{E}\left( \left( \epsilon_{t+(jB+b)/M} - \epsilon_{t+((j-1)B+b)/M} \right)^2 \right) = O_p(M^{-1/2})$, uniformly in $t$, and by assumption $\mathrm{E}(\epsilon^{2k}_{t+j/M}) < \infty$, then the term in (A.17) is $O(l^{k/2}/B^{k/2}) = (b_M^{-k/2})$, with $b_M = M^{1/3}$. Given the independence between the noise and the price, (A.16) is $O(B^{-k/2})$.

We are left with (A.15). From the proof of Theorem 2 in Zhang, Mykland and Aït-Sahalia (2005), assuming no drift, (A.15) writes as

$$\mathrm{E}\left( \left( 2 \sum_{j=1}^{M-1} \left( Y_{t+(j+1)/M} - Y_{t+j/M} \right) \sum_{i=1}^{\min(B,j)} \left( 1 - \frac{j}{B} \right) \left( Y_{t+(j-i+1)/M} - Y_{t+(j-i)/M} \right) \right)^k \right) (1 + o(1))$$

$$= \mathrm{E}\left( \left( \sum_{j=1}^{M-1} \int_{t+j/M}^{t+(j+1)/M} \sigma_s \mathrm{d}W_s \sum_{i=1}^{\min(B,j)} \left( 1 - \frac{j}{B} \right) \int_{t+(j-i)/M}^{t+(j-i+1)/M} \sigma_u \mathrm{d}W_u \right)^k \right)$$

$$= O(B^{k/2} M^{-k/2}) = O(l^{k/2}) = O(b_M^{-k/2}) \tag{A.18}$$

for $b_M = M^{1/3}$. The order of magnitude of (A.18), comes by the same argument as in part (i), once we notice that

$$\mathrm{E}_{t+(j-B)/M} \left( \sigma_{t+j/M} \int_{t+j/M}^{t+(j+1)/M} \mathrm{d}W_s \sum_{i=1}^{\min(B,j)} \left( 1 - \frac{j}{B} \right) \sigma_{t+(j-i)/M} \int_{t+(j-i)/M}^{t+(j-i+1)/M} \mathrm{d}W_u \right) = 0.$$

The case with drift can be treated as in Part (i).
(iv)

$$\widetilde{RV}_{t,e,M} = \sum_{i=1}^{e} \frac{a_i}{i} \left( \sum_{j=1}^{M-i} \left( X_{t+(j+i)/M} - X_{t+j/M} \right)^2 \right) + \frac{RV_{t,M}}{M}, \text{ where } a_i = 12 \frac{i}{e^2} \frac{\left( \frac{i}{e} - \frac{1}{2} - \frac{1}{2e} \right)}{\left( 1 - \frac{1}{e^2} \right)}.$$

The case of $k=2$ is treated in Aït-Sahalia, Mykland and Zhang (2006). Recalling that $\sum_{i=1}^{e} \frac{a_i}{i} = 0$ and $\sum_{i=1}^{e} a_i = 1$,

$$\begin{aligned}
&\mathrm{E}\left( \left( \widetilde{RV}_{t,e,M} - IV_t \right)^k \right) \\
&= \mathrm{E}\left( \left( \left( \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \left( Y_{t+(j+i)/M} - Y_{t+j/M} \right)^2 - IV_t \right) - 2 \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \epsilon_{t+(j+i)/M} \epsilon_{t+j/M} \right.\right. \\
&+ 2 \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \left( Y_{t+(j+i)/M} - Y_{t+j/M} \right) \left( \epsilon_{t+(j+i)/M} - \epsilon_{t+j/M} \right) - \left( \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=M-i}^{M} \left( \epsilon_{t+j/M}^2 - \sigma_\epsilon^2 \right) \right) \\
&\left.\left. +2 \left( \widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2 \right) \right)^k \right)
\end{aligned} \tag{A.19}$$

It is immediate to see that $\mathrm{E}\left( \left( \widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2 \right)^k \right) = O(M^{-k/2})$. Note also that $a_i \simeq i^2/e^3$. Therefore,

$$\mathrm{E}\left( \left( \sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{i} \left( \epsilon_{t+j/M}^2 - \sigma_\epsilon^2 \right) \right)^k \right) \simeq \mathrm{E}\left( \left( \frac{1}{e} \sum_{j=1}^{i} \left( \epsilon_{t+j/M}^2 - \sigma_\epsilon^2 \right) \right)^k \right) = O(e^{-k/2}),$$

so that the $k$-th moments of the fourth and fifth terms of the right hand side (rhs) of (A.19) are $O(e^{-k/2})$. The $k-$th moments of the first and third terms of the rhs of (A.19) can be treated by the same argument as in Part (i) and (iii) and thus, by noting that $a_i/i = O(e^{-2})$, are $O(e^{-k/2})$. The $k-$th moment of the second term can be treated as the second term of the rhs of the last equality in (A.21) (in Part (v) below), by noting that the weights are of the same order of magnitude ($O(e^{-2})$). It follows that this term is also $O(e^{-k/2})$. Finally, note that by Hölder inequality, the contribution of the cross terms cannot be of a larger order. Thus, the statement follows for $b_M = M^{1/2}$.
(v)

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

$$\gamma_{t,h}^X = \sum_{j=1}^{M} \left( X_{t+j/M} - X_{t+(j-1)/M} \right) \left( X_{t+(j-h)/M} - X_{t+(j-1-h)/M} \right). \tag{A.20}$$

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

$$\begin{aligned}
\mathrm{E}\left( \left( RK_{t,H,M} - IV_t \right)^k \right) = \mathrm{E}\Bigg( \Bigg( & \left( \gamma_{t,0}^Y - IV_t \right) + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^Y + \gamma_{t,-h}^Y \right) \\
& + \gamma_{t,0}^{Y,\epsilon} + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^{Y,\epsilon} + \gamma_{t,-h}^{Y,\epsilon} \right) + \gamma_{t,0}^{\epsilon,Y} + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^{\epsilon,Y} + \gamma_{t,-h}^{\epsilon,Y} \right) \\
& + \gamma_{t,0}^{\epsilon} + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^{\epsilon} + \gamma_{t,-h}^{\epsilon} \right) \Bigg)^k \Bigg),
\end{aligned}$$

with $\gamma_{t,h}^{Y,\epsilon} = \sum_{j=1}^{M} \left( Y_{t+j/M} - Y_{t+(j-1)/M} \right) \left( \epsilon_{t+(j-h)/M} - \epsilon_{t+(j-1-h)/M} \right)$ and the other terms defined as in (A.20). As $\gamma_{t,0}^{\epsilon} + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^{\epsilon} + \gamma_{t,-h}^{\epsilon} \right)$ is the term of the highest order of magnitude, we just need to show that $\mathrm{E}\left( \left( \gamma_{t,0}^{\epsilon} + \sum_{h=1}^{H} \kappa \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}^{\epsilon} + \gamma_{t,-h}^{\epsilon} \right) \right)^k \right) = O\left( H^{-\frac{3k}{2}} M^{k/2} \right)$.

Let $\boldsymbol{\gamma}_t^{\epsilon} = \left( \gamma_{t,0}^{\epsilon}, \gamma_{t,1}^{\epsilon} + \gamma_{t,-1}^{\epsilon}, \ldots, \gamma_{t,h}^{\epsilon} + \gamma_{t,-h}^{\epsilon} \right)'$. For notational simplicity, hereafter we drop the subscript $t$. Following the proof of Theorem 1 in Barndorff-Nielsen, Hansen, Lunde and Shephard (2008),

$$\boldsymbol{\gamma}^{\epsilon} = \boldsymbol{\gamma}_V^{\epsilon} + \boldsymbol{\gamma}_W^{\epsilon} + \boldsymbol{\gamma}_Z^{\epsilon},$$

where

$$\boldsymbol{\gamma}_V^{\epsilon} = 2 \left( V_0 - V_1, -V_0 + 2V_1 - V_2, \ldots, -V_{h-1} + 2V_h - V_{h+1}, \ldots, -V_{H-1} + 2V_H - V_{H+1} \right)',$$

with $V_h = \sum_{j=1}^{M-h-1} \epsilon_{t+j/M}\epsilon_{t+(j+h)/M}$,

$$\boldsymbol{\gamma}_W^\epsilon = (0, -W_2, 2W_2 - W_3, \ldots, -W_{h-1} + 2W_h - W_{h+1}, \ldots, -W_{H-1} + 2W_H - W_{H+1})',$$

with $W_h = \sum_{j=1}^{h-1} \epsilon_{t+j/M}\epsilon_{t+(j-h)/M} + \sum_{j=M-h+1}^{M} \epsilon_{t+j/M}\epsilon_{t+(j+h)/M}$, and finally

$$\boldsymbol{\gamma}_Z^\epsilon = (Z_0 - 2Z_1, Z_{-1} - Z_0 + 3Z_1 - 2Z_2, Z_{-2} - Z_{-1} - Z_1 + 3Z_2 - 2Z_3, \ldots, Z_{-H} - Z_{-H+1} - Z_{H-1} + 3Z_H - 2Z_{H+1})',$$

with $Z_h = \epsilon_t\epsilon_{t+h/M} + \epsilon_{t+1}\epsilon_{t+(M-h)/M}$. As all cross moments can be dealt with by Hölder inequality, it suffices to show that, for $\mathbf{w} = \left(1, 1, \kappa\left(\frac{1}{H}\right), \ldots, \kappa\left(\frac{H-1}{H}\right)\right)'$, $\mathrm{E}\left(\mathbf{w}'\boldsymbol{\gamma}_V^\epsilon\right)^k, \mathrm{E}\left(\mathbf{w}'\boldsymbol{\gamma}_W^\epsilon\right)^k, \mathrm{E}\left(\mathbf{w}'\boldsymbol{\gamma}_Z^\epsilon\right)^k = O\left(H^{-k/2}\right)$. We begin with $\mathbf{w}'\boldsymbol{\gamma}_V^\epsilon$. After a few simple manipulations,

$$\begin{aligned}
&\frac{1}{2}\mathbf{w}'\boldsymbol{\gamma}_V^\epsilon \\
&= (V_0 - V_1) + \sum_{h=1}^{H} \kappa\left(\frac{h-1}{H}\right)(-V_{h-1} + 2V_h - V_{h+1}) \\
&= \left(1 - \kappa\left(\frac{1}{H}\right)\right)V_1 + \sum_{h=1}^{H-2}\left(\kappa\left(\frac{h-1}{H}\right) - 2\kappa\left(\frac{h}{H}\right) + \kappa\left(\frac{h+1}{H}\right)\right)V_{h+1} \\
&+ \left(2\kappa\left(\frac{H-1}{H}\right) - \kappa\left(\frac{H-2}{H}\right)\right)V_H - \kappa\left(\frac{H-1}{H}\right)V_{H+1}
\end{aligned} \tag{A.21}$$

As $\epsilon$ is *i.i.d.*, $\mathrm{E}\left(V_h^k\right) = O(M^{k/2})$, and $\mathrm{E}(V_jV_kV_l) = 0$ unless $i = j = l$, it follows that

$$\mathrm{E}\left(\sum_{h=1}^{H} V_h\right)^k \simeq \sum_{h_1=1}^{H}\sum_{h_2=1}^{H}\cdots\sum_{h_{k/2}=1}^{H} \mathrm{E}\left(V_{h_1}^2\right)\mathrm{E}\left(V_{h_2}^2\right)\ldots\mathrm{E}\left(V_{h_{k/2}}^2\right) = O\left(H^{k/2}M^{k/2}\right)$$

Thus,

$$\frac{1}{2^k}\mathrm{E}\left(\left(\mathbf{w}'\boldsymbol{\gamma}_V^\epsilon\right)^k\right) \tag{A.22}$$

$$\simeq \left(1 - \kappa\left(\frac{1}{H}\right)\right)^k \mathrm{E}\left(V_1^k\right) \tag{A.23}$$

$$+ \sum_{h_1=1}^{H}\sum_{h_2=1}^{H}\cdots\sum_{h_{k/2}=1}^{H}\left(\left(\kappa\left(\frac{h_1-1}{H}\right) - 2\kappa\left(\frac{h_1}{H}\right) + \kappa\left(\frac{h_1+1}{H}\right)\right)^2\ldots\right. \tag{A.24}$$

$$\left.\ldots\left(\kappa\left(\frac{h_{k/2}-1}{H}\right) - 2\kappa\left(\frac{h_{k/2}}{H}\right) + \kappa\left(\frac{h_{k/2}+1}{H}\right)\right)^2\right)\mathrm{E}\left(V_{h_1+1}^2\right)\ldots\mathrm{E}\left(V_{h_{k/2}+1}^2\right)$$

$$+ \left(2\kappa\left(\frac{H-1}{H}\right) - \kappa\left(\frac{H-2}{H}\right)\right)^k \mathrm{E}\left(V_H^k\right) \tag{A.25}$$

$$+ \kappa\left(\frac{H-1}{H}\right)^k \mathrm{E}\left(V_{H+1}^k\right) \tag{A.26}$$

Exploiting the properties of the kernel $\kappa$, and employing standard Taylor expansions it follows that (A.23), (A.25) and (A.26) are $O\left(M^{k/2}H^{-2k}\right) = O(H^{-1})$, while (A.24) is $O\left(H^{-\frac{3k}{2}}M^{k/2}\right)$. Hence, (A.22) is $O\left(H^{-\frac{3k}{2}}M^{k/2}\right)$. Now, by a similar argument it follows that $\mathrm{E}\left(\left(\mathbf{w}'\boldsymbol{\gamma}_W^\epsilon\right)^k\right), \mathrm{E}\left(\left(\mathbf{w}'\boldsymbol{\gamma}_Z^\epsilon\right)^k\right) = O\left(H^{-k/2}\right)$. Then, for $H = M^{1/2}$ and $b_M = M^{1/2}$ the statement follows.

(vi) The case of $k = 2$ has been established by Proposition 4 in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). It suffices to show that $\mathrm{E}\left(\left(\mathbf{w}'\boldsymbol{\gamma}_V^\epsilon\right)^k\right)$, as defined in (A.22), is $O\left(H^{-k/2}\right)$. Contrary to the *i.i.d.* case, $\mathrm{E}\left(V_h\right) \neq 0$ and $\mathrm{E}\left(V_hV_{h'}\right) \neq 0$ for $h \neq h'$. For notational simplicity, we consider the case of $k = 4$ (the case of higher $k$ follows straightforwardly). Because of stationarity, hereafter we suppress the subscript $t$. We also assume, for sake of clarity but without loss of generality, that $\epsilon_{j/M} = \rho\epsilon_{(j-1)/M} + u_{j/M}$, with $u_{j/M} \sim i.i.d.(0, \sigma_u^2)$.

Recalling that $\kappa^{(1)}(0) = \kappa^{(2)}(0) = \kappa(1) = \kappa^{(1)}(1) = 0$, via a Taylor expansion of the first two terms around zero and of the last two terms around one,

$$\frac{1}{2}\mathbf{w}'\boldsymbol{\gamma}_V^\epsilon = \frac{\kappa^{(3)}(0)}{H^3}V_1 + \frac{1}{H^3}\sum_{h=1}^{H-2}\kappa^{(3)}(0)hV_{h+1} + \frac{2\kappa^{(2)}(1)}{H^2}V_H + \frac{\kappa^{(2)}(1)}{2H^2}V_{H+1}, \tag{A.27}$$

where $V_h$ is defined as in part (v). Starting from the third term in (A.27),

$$\begin{aligned}
&\mathrm{E}\left(\left(\frac{\kappa^{(2)}(1)}{H^2}V_H\right)^4\right)\\
&\simeq \frac{\kappa^{(2)}(1)^4}{H^8}\sum_{j=1}^{M}\sum_{j_2>j_1+H}^{M}\sum_{j_3>j_2+H}^{M}\sum_{j_4>j_3+H}^{M}\mathrm{E}\left(\epsilon_{j_1/M}\epsilon_{(j_1+H)/M}\epsilon_{j_2/M}\epsilon_{(j_2+H)/M}\epsilon_{j_3/M}\epsilon_{(j_3+H)/M}\epsilon_{j_4/M}\epsilon_{(j_4+H)/M}\right)\\
&+\frac{\kappa^{(2)}(1)^4}{H^8}\sum_{j=1}^{M}\sum_{\substack{j_1<j_2\\ j_2<j_1+H}}^{M}\sum_{\substack{j_2<j_3\\ j_3<j_2+H}}^{M}\sum_{\substack{j_3<j_4\\ j_4<j_3+H}}^{M}\mathrm{E}\left(\epsilon_{j_1/M}\epsilon_{(j_1+H)/M}\epsilon_{j_2/M}\epsilon_{(j_2+H)/M}\epsilon_{j_3/M}\epsilon_{(j_3+H)/M}\epsilon_{j_4/M}\epsilon_{(j_4+H)/M}\right)\\
&\simeq \frac{\kappa^{(2)}(1)^4}{H^8}\rho^H\sum_{j=1}^{M}\sum_{j_2>j_1+H}^{M}\sum_{j_3>j_2+H}^{M}\sum_{j_4>j_3+H}^{M}\mathrm{E}\left(\epsilon_{j_1/M}\epsilon_{(j_1+H)/M}\epsilon_{j_2/M}\epsilon_{(j_2+H)/M}\epsilon_{j_3/M}\epsilon_{(j_3+H)/M}\epsilon_{j_4/M}g_{j_4/M}\right)\\
&+\frac{\kappa^{(2)}(1)^4}{H^8}O(H^{5/2})=o(H^{-2}),
\end{aligned}$$

given that $M=O(H^2)$ and $\rho^H=o(H^{-2})$. Thus, the fourth moment of the first and fourth terms of the rhs of (A.27) are also $o(H^{-2})$. Finally, the fourth moment of the second term in (A.27) is given by

$$\begin{aligned}
&\mathrm{E}\left(\left(\frac{\kappa^{(3)}(0)}{H^3}\sum_{h=1}^{H}hV_{n,h}\right)^4\right)\\
&=\frac{\left(\kappa^{(3)}(0)\right)^4}{H^{12}}\sum_{h_1=1}^{H}h_1\sum_{h_2=1}^{H}h_2\sum_{h_3=1}^{H}h_3\sum_{h_4=1}^{H}h_4\\
&\times\sum_{j_1=1}^{M}\sum_{j_2=1}^{M}\sum_{j_3=1}^{M}\sum_{j_4=1}^{M}\mathrm{E}\left(\epsilon_{j_1/M}\epsilon_{(j_1+h_1)/M}\epsilon_{j_2/M}\epsilon_{(j_2+h_2)/M}\epsilon_{j_3/M}\epsilon_{(j_3+h_3)/M}\epsilon_{j_4/M}\epsilon_{(j_4+h_4)/M}\right).
\end{aligned}$$

As $\sum_{h_1=1}^{H}h_1^2=O(H^3)$ and $\left(\sum_{h_1=1}^{H}h_1\right)^2=O(H^4)$, the most complex case 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 is immediate to see that

$$\begin{aligned}
&\frac{1}{H^{12}}\sum_{h_1=1}^{H}h_1\sum_{h_2>h_1}^{H}h_2\sum_{h_3>h_2}^{H}h_3\sum_{h_4>h_3}^{H}h_4\sum_{j_1=1}^{M}\sum_{j_3>j_1}^{M}\\
&\times\mathrm{E}\left(\epsilon_{j_1/M}^2\epsilon_{j_3/M}^2\epsilon_{(j_1+h_1)/M}\epsilon_{(j_1+h_2)/M}\epsilon_{(j_3+h_3)/M}\epsilon_{(j_3+h_4)/M}\right)=O(H^{-2}).
\end{aligned}$$

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). By the law of the iterated expectation,

$$\begin{aligned}
&\mathrm{E}\left(\epsilon_{j_1/M}\epsilon_{(j_1+h_1)/M}\epsilon_{j_2/M}\epsilon_{(j_2+h_2)/M}\epsilon_{j_3/M}\epsilon_{(j_3+h_3)/M}\epsilon_{j_4/M}\epsilon_{(j_4+h_4)/M}\right)\\
&=\Big(\mathrm{E}\Big(\epsilon_{j_1/M}\mathrm{E}_{j_1/M}\left(\epsilon_{(j_1+h_1)/M}\right)\mathrm{E}_{(j_1+h_1)/M}\left(\epsilon_{j_2/M}\right)\mathrm{E}_{j_2/M}\left(\epsilon_{(j_2+h_2)/M}\right)\\
&\mathrm{E}_{(j_2+h_2)/M}\left(\epsilon_{j_3/M}\right)\mathrm{E}_{j_3/M}\left(\epsilon_{(j_3+h_3)/M}\right)\mathrm{E}_{(j_3+h_3)/M}\left(\epsilon_{j_4/M}\right)\mathrm{E}_{j_4/M}\left(\epsilon_{(j_4+h_4)/M}\right)\Big)\Big).
\end{aligned}$$

The conditional expectations above can be simplified as follows:

$$\begin{aligned}
\mathrm{E}_{j_4/M}\left(\epsilon_{(j_4+h_4)/M}\right)&=\rho^{h_4}\epsilon_{j_4/M},\\
\mathrm{E}_{(j_3+h_3)/M}\left(\epsilon_{j_4/M}^2\right)&=\left(\rho^{2(j_4-j_3-h_3)}\epsilon_{(j_3+h_3)/M}^2+2\sum_{k=1}^{j_4-j_3-h_3}\rho^k\sigma_u^2\right)\\
&=\rho^{2(j_4-j_3-h_3)}\epsilon_{(j_3+h_3)/M}^2+c,\\
\mathrm{E}_{j_3/M}\left(\epsilon_{(j_3+h_3)/M}^3\right)&=\rho^{3h_3}\epsilon_{j_3/M}^3+6\sum_{k=1}^{2h_3}\rho^{k+h_3-1}\epsilon_{j_3/M}\sigma_u^2+\sum_{k=1}^{h_3}\rho^{k-1}\mathrm{E}(u_1^4)\\
&=\rho^{3h_3}\epsilon_{j_3/M}^3+c\rho^{h_3}\epsilon_{j_3/M}+c,\\
\mathrm{E}_{(j_2+h_2)/M}\left(\epsilon_{j_3/M}^4\right)&=\rho^{4(j_3-j_2-h_2)}\epsilon_{(j_2+h_2)/M}^4+c\sum_{k=1}^{2}\rho^{k(j_3-j_2-h_2)}\epsilon_{(j_2+h_2)/M}^k+c,
\end{aligned}$$

$$\mathrm{E}_{j_2/M}\left(\epsilon^5_{(j_2+h_2)/M}\right) = \rho^{5h_2}\epsilon^5_{j_2/M} + c\sum_{k=1}^{3}\rho^{kh_2}\epsilon^k_{j_2/M} + c,$$

$$\mathrm{E}_{(j_1+h_1)/M}\left(\epsilon^6_{j_2/M}\right) = \rho^{6(j_2-j_1-h_1)}\epsilon^6_{(j_1+h_1)/M} + c\sum_{k=1}^{4}\rho^{k(j_2-j_1-h_1)}\epsilon^k_{j_2/M} + c,$$

$$\mathrm{E}_{j_1/M}\left(\epsilon^7_{(j_1+h_1)/M}\right) = \rho^{7h_1}\epsilon^7_{j_2/M} + c\sum_{k=1}^{5}\rho^{kh_1}\epsilon^k_{j_2/M} + c,$$

where $c$ denotes a generic $O(1)$ term. Given the results above, it follows that

$$\begin{aligned}
&\frac{1}{H^{12}}\sum_{h_1=1}^{H} h_1 \sum_{h_2>h_1}^{H} h_2 \sum_{h_3>h_2}^{H} h_3 \sum_{h_4>h_3}^{H} h_4 \sum_{j_1=1}^{M}\sum_{j_2>j_1+h_1}^{M}\sum_{j_3>j_2+h_2}^{M}\sum_{j_4>j_3+h_3}^{M} \\
&\times \mathrm{E}\left(\epsilon_{j_1/M}\epsilon_{(j_1+h_1)/M}\epsilon_{j_2/M}\epsilon_{(j_2+h_2)/M}\epsilon_{j_3/M}\epsilon_{(j_3+h_3)/M}\epsilon_{j_4/M}\epsilon_{(j_4+h_4)/M}\right) \\
&= \frac{c}{H^{12}}M^4\sum_{h_1=1}^{H} h_1\rho^{h_1}\sum_{h_2>h_1}^{H} h_2\rho^{h_2}\sum_{h_3>h_2}^{H} h_3\rho^{h_3}\sum_{h_4>h_3}^{H} h_4\rho^{h_4}\mathrm{E}\left(\epsilon^2_{j_1/M}\right) \\
&+ \dots \\
&+ \frac{c}{H^{12}}M^2\sum_{h_1=1}^{H} h_1\rho^{h_1}\sum_{h_2>h_1}^{H} h_2\rho^{h_2}\sum_{h_3>h_2}^{H} h_3\rho^{3h_3}\sum_{h_4>h_3}^{H} h_4\rho^{h_4}\sum_{j_1=1}^{M}\sum_{j_2>j_1+h_1}^{M}\rho^{j_2-j_1-h_1}\mathrm{E}\left(\epsilon^4_{j_1/M}\right) \\
&+ \dots \\
&+ \frac{1}{H^{12}}\sum_{h_1=1}^{H} h_1\rho^{7h_1}\sum_{h_2>h_1}^{H} h_2\rho^{5h_2}\sum_{h_3>h_2}^{H} h_3\rho^{3h_3}\sum_{h_4>h_3}^{H} h_4\rho^{h_4}\sum_{j_1=1}^{M}\sum_{j_2>j_1+h_1}^{M}\rho^{6(j_2-j_1-h_1)} \\
&\sum_{j_3>j_2+h_2}^{M}\rho^{4(j_3-j_2-h_2)}\sum_{j_4>j_3+h_3}^{M}\rho^{2(j_4-j_3-h_3)}\mathrm{E}\left(\epsilon^8_{j_1/M}\right) \\
&= O(H^{-4})(1+o(1)) = o(H^{-2})
\end{aligned}$$

The statement in the Lemma then follows. ■

# References


Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2005). How Often to Sample a Continuous Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies*, 18, 351-416.

Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2006). Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise. Manuscript, Princeton University.

Aït-Sahalia, Y., J. Fan and H. Peng (2009). Forthcoming, *Journal of the American Statistical Association.*

Aït-Sahalia, Y., and L. Mancini (2008). Out of Sample Forecasts of Quadratic Variation. *Journal of Econometrics*, 147, 17-33.

Andersen, T.G., T. Bollerslev, P. F. Christoffersen and F.X. Diebold (2006). Practical Volatility and Correlation Modeling for Financial Market Risk Management. In M. Carey and R. Stulz (eds.), *Risks of Financial Institutions*. University of Chicago Press for NBER, Chicago.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96, 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003). Modelling and Forecasting Realized Volatility. *Econometrica*, 71, 579-626

Andersen, T.G., T. Bollerslev and S. Lang (1999). Forecasting Financial Market Volatility. Sample Frequency vs Forecast Horizon. *Journal of Empirical Finance*, 6, 457-477.

Andersen, T.G., T. Bollerslev and N. Meddahi (2004). Analytic Evaluation of Volatility Forecasts. *International Economic Review*, 45, 1079-1110.

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

Andersen, T.G., T. Bollerslev and N. Meddahi (2006). Market Microstructure Noise and Realized Volatility Forecasting, Technical Report, University of Montreal.

Awartani, B.M., V. Corradi and W. Distaso (2009). Testing Market Microstructure Effects, with an Application to the Dow Jones Industrial Average Stocks. *Journal of Business and Economic Statistics,* forthcoming.

Bandi, F.M. and J.R. Russell (2007). Market Microstructure Noise, integrated variance Estimators, and the Accuracy of Asymptotic Approximations. Working Paper, University of Chicago, Graduate School of Business.

Barndorff-Nielsen O.E., S.E. Graversen, J. Jacod, M. Podolskij and N. Shephard (2006). A Central Limit Theorem for Realized Power and Bipower Variations of Continuous Semimartingales, in *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*, Springer and Verlag, Y. Kabanov and R. Lipster (eds.), New York.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2008). Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. *Econometrica,* 76, 1481-1536.

Barndorff-Nielsen, O.E., S. Kinnebrouck and N. Shephard (2009). 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.E. and N. Shephard (2002). Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility. *Journal of the Royal Statistical Society, Ser. B*, 64, 243-280.

Barndorff-Nielsen, O.E. and N. Shephard (2004). Power and Bipower Variation with Stochastic Volatility and Jumps (with discussion). *Journal of Financial Econometrics*, 2, 1-48.

Barndorff-Nielsen O.E., N. Shephard and M. Winkel (2006). Limit Theorems for Multipower Variation in the Presence of Jumps. *Stochastic Processes and Their Applications*, 116, 796-806.

Carr, P. and R. Lee (2003). Trading Autocorrelation. Manuscript, New York University.

Claeskens, G. and P. Hall (2002). Effect of Dependence on Stochastic Measures of Accuracy of Density Estimators. *Annals of Statistics,* 30, 431-454.

Comte, F. and E. Renault (1998). Long Memory in Stochastic Volatility Models. *Mathematical Finance,* 8, 291-323.

Comte, F., L. Coutin and E. Renault (2003). Affine Fractional Stochastic Volatility Models with Application to Option Pricing, Working Paper. Cirano-Cireq University of Montreal.

Corradi, V. and W. Distaso (2006). Semiparametric Comparison of Stochastic Volatility Models via Realized Measures. *Review of Economic Studies*, 73, 635-668.

Corradi, V., W. Distaso and N.R. Swanson (2008). Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures. *Journal of Econometrics,* forthcoming.

Diebold F.X., T.A. Gunther and A.S. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39, 863-883.

Fan, J. and Q. Yao (2005). *Nonlinear Time Series: Parametric and Nonparametric Methods.* Springer-Verlag, New York.

Fan, J., Q. Yao and H. Tong (1996). Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems. *Biometrika,* 83, 189-206.

Ghysels, E., and A. Sinko (2006). Volatility Forecasting and Microstructure Noise. Manuscript, University of North Carolina.

Hall, P., R.C.L. Wolff and Q. Yao (1999). Methods for Estimating a Conditional Distribution Function. *Journal of the American Statistical Association,* 94, 154-163.

Kalnina, I. and O. Linton (2008). Estimating Quadratic Variation Consistently in the Presence of Endogenous and Diurnal Measurement Error. *Journal of Econometrics,* 147, 47-59.

Masuda, H. (2004). Ergodicity and Exponential $\beta-$mixing for a Strong Solution of Levy-Driven Stochastic Differential Equations. MHF 2004-19, Kyushu University.

Meddahi, N. (2001). An Eigenfunction Approach for Volatility Modeling. Manuscript, University of Montreal.

Meddahi, N. (2002). A Theoretical Comparison between Integrated and Realized Volatility. *Journal of Applied Econometrics*, 17, 475-508.

Meyn, S.P. and R.L. Tweedie (1993). Stability of Markovian Processes 3: Foster-Lyapunov Criteria for Continuous Time Processes. *Advance in Applied Probability,* 25, 518-548.

Müller, H. G. (1991). Smooth Optimum Kernel Estimators near Endpoints. *Biometrika*, 78, 521-530.

O'Hara, M. (1997). *Market Microstructure Theory.* Blackwell, Cambridge.

Schwert, G.W. (1989). Why Does Stock Market Volatility Change Over Time? *Journal of Finance*, 44, 1115-1153.

Silverman, B.W. (1986). *Density Estimation.* Chapman & Hall, New York.

Taqqu, M.S. (1975). Weak Convergence to Fractional Brownian Motion and to Rosenblatt Process. *Zeitschrift fur Wahrscheinlichkitstheorie und Verwandte Gebiete,* 31, 287-302.

Wasserfallen, W. and H. Zimmermann (1985). The behavior of intraday exchange rates. *Journal of Banking and Finance*, 9, 55-72.

Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2005). A Tale of Two Time Scales: Determining integrated volatility with Noisy High Frequency Data. *Journal of the American Statistical Association*, 100, 1394-1411.

Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2006). Edgeworth Expansion for Realized Volatility and Related Estimators. Working Paper. Princeton University.

Zhang, L. (2006). Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. *Bernoulli,* 12, 1019-1043.

Table 1: Conditional Confidence Interval Accuracy Assessment: Level Experiments

**Case I: No Microstructure Noise or Jumps in DGP**[a]

| $M$ | $RV_{t,M}$ | $BV_{t,M}$ | $TPV_{t,M}$ | $\widehat{RV}_{t,l,M}$ | $\widetilde{RV}_{t,e,M}$ | $RK_{t,H,M}$ |
|---|---|---|---|---|---|---|
| *Interval* $= \widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$ | | | | | | |
| *Sample Size = 100 Daily Realized Measure Observations* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.091 | 0.108 | 0.116 | 0.453 | 0.309 | 0.199 |
| 144 | 0.093 | 0.086 | 0.090 | 0.254 | 0.241 | 0.142 |
| 288 | 0.081 | 0.086 | 0.085 | 0.165 | 0.182 | 0.089 |
| 576 | 0.074 | 0.081 | 0.082 | 0.109 | 0.140 | 0.084 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.136 | 0.157 | 0.166 | 0.527 | 0.374 | 0.255 |
| 144 | 0.140 | 0.142 | 0.142 | 0.318 | 0.298 | 0.180 |
| 288 | 0.128 | 0.130 | 0.134 | 0.218 | 0.230 | 0.142 |
| 576 | 0.124 | 0.136 | 0.129 | 0.155 | 0.189 | 0.130 |
| *Sample Size = 300 Daily Realized Measure Observations* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.070 | 0.093 | 0.105 | 0.727 | 0.470 | 0.253 |
| 144 | 0.068 | 0.078 | 0.070 | 0.342 | 0.332 | 0.108 |
| 288 | 0.080 | 0.074 | 0.067 | 0.184 | 0.217 | 0.066 |
| 576 | 0.056 | 0.056 | 0.068 | 0.091 | 0.144 | 0.069 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.113 | 0.140 | 0.151 | 0.795 | 0.559 | 0.332 |
| 144 | 0.113 | 0.130 | 0.114 | 0.435 | 0.409 | 0.151 |
| 288 | 0.116 | 0.114 | 0.108 | 0.245 | 0.282 | 0.109 |
| 576 | 0.103 | 0.113 | 0.115 | 0.133 | 0.207 | 0.111 |

[a] 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 5. 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 $[u1, u2] = [\widehat{\mu}_{IV_t} - \beta\widehat{\sigma}_{IV_t}, \widehat{\mu}_{IV_t} + \beta\widehat{\sigma}_{IV_t}]$, where $\widehat{\mu}_{IV_t}$ and $\widehat{\sigma}_{IV_t}$ are the mean and standard error of the pseudo true data, and $\beta = 0.125$. See Section 5 for further details.

Table 2: Conditional Confidence Interval Accuracy Assessment: Level Experiments

**Case II: Microstructure Noise in DGP**[a]

| $M$ | $RV_{t,M}$ | $BV_{t,M}$ | $TPV_{t,M}$ | $\widehat{RV}_{t,l,M}$ | $\widetilde{RV}_{t,e,M}$ | $RK_{t,H,M}$ |
|---|---|---|---|---|---|---|
| *Panel A: Interval* $= \widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$ | | | | | | |
| *Noise has* $\mathrm{N}(0, 0.005^2)$ *distribution* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.064 | 0.080 | 0.077 | 0.466 | 0.320 | 0.215 |
| 144 | 0.075 | 0.075 | 0.073 | 0.252 | 0.224 | 0.131 |
| 288 | 0.155 | 0.171 | 0.137 | 0.175 | 0.182 | 0.108 |
| 576 | 0.977 | 0.990 | 0.987 | 0.109 | 0.139 | 0.096 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.095 | 0.126 | 0.119 | 0.538 | 0.370 | 0.264 |
| 144 | 0.115 | 0.113 | 0.128 | 0.316 | 0.288 | 0.183 |
| 288 | 0.198 | 0.208 | 0.181 | 0.227 | 0.230 | 0.155 |
| 576 | 0.979 | 0.990 | 0.988 | 0.155 | 0.189 | 0.135 |
| *Noise has* $\mathrm{N}(0, 0.007^2)$ *distribution* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.057 | 0.085 | 0.070 | 0.472 | 0.323 | 0.194 |
| 144 | 0.131 | 0.122 | 0.123 | 0.260 | 0.240 | 0.132 |
| 288 | 0.932 | 0.967 | 0.942 | 0.173 | 0.179 | 0.119 |
| 576 | 1.000 | 1.000 | 1.000 | 0.111 | 0.137 | 0.100 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.105 | 0.136 | 0.121 | 0.539 | 0.376 | 0.247 |
| 144 | 0.173 | 0.176 | 0.173 | 0.325 | 0.301 | 0.185 |
| 288 | 0.943 | 0.972 | 0.947 | 0.222 | 0.235 | 0.163 |
| 576 | 1.000 | 1.000 | 1.000 | 0.158 | 0.189 | 0.144 |
| *Noise has* $\mathrm{N}(0, 0.014^2)$ *distribution* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.739 | 0.719 | 0.642 | 0.478 | 0.321 | 0.228 |
| 144 | 1.000 | 1.000 | 1.000 | 0.272 | 0.236 | 0.167 |
| 288 | 1.000 | 1.000 | 1.000 | 0.170 | 0.170 | 0.116 |
| 576 | 1.000 | 1.000 | 1.000 | 0.123 | 0.148 | 0.097 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.764 | 0.751 | 0.682 | 0.545 | 0.398 | 0.287 |
| 144 | 1.000 | 1.000 | 1.000 | 0.329 | 0.297 | 0.226 |
| 288 | 1.000 | 1.000 | 1.000 | 0.218 | 0.222 | 0.159 |
| 576 | 1.000 | 1.000 | 1.000 | 0.170 | 0.198 | 0.145 |
| *Panel B: Interval* $= \widehat{\mu}_{IV_t} + 0.250\widehat{\sigma}_{IV_t}$ | | | | | | |
| *Noise has* $\mathrm{N}(0, 0.005^2)$ *distribution* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.070 | 0.071 | 0.062 | 0.625 | 0.407 | 0.225 |
| 144 | 0.079 | 0.090 | 0.076 | 0.330 | 0.271 | 0.108 |
| 288 | 0.104 | 0.137 | 0.113 | 0.168 | 0.177 | 0.076 |
| 576 | 0.961 | 0.987 | 0.981 | 0.073 | 0.130 | 0.065 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.116 | 0.120 | 0.103 | 0.719 | 0.476 | 0.282 |
| 144 | 0.139 | 0.138 | 0.131 | 0.389 | 0.343 | 0.163 |
| 288 | 0.152 | 0.186 | 0.162 | 0.229 | 0.248 | 0.132 |
| 576 | 0.970 | 0.988 | 0.987 | 0.118 | 0.192 | 0.120 |
| *Noise has* $\mathrm{N}(0, 0.007^2)$ *distribution* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.059 | 0.074 | 0.060 | 0.621 | 0.405 | 0.232 |
| 144 | 0.083 | 0.106 | 0.083 | 0.332 | 0.273 | 0.105 |
| 288 | 0.934 | 0.966 | 0.939 | 0.160 | 0.185 | 0.080 |
| 576 | 1.000 | 1.000 | 1.000 | 0.075 | 0.128 | 0.083 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.110 | 0.142 | 0.101 | 0.716 | 0.481 | 0.303 |
| 144 | 0.122 | 0.157 | 0.123 | 0.398 | 0.337 | 0.157 |
| 288 | 0.952 | 0.971 | 0.945 | 0.233 | 0.237 | 0.133 |
| 576 | 1.000 | 1.000 | 1.000 | 0.121 | 0.187 | 0.130 |
| *Noise has* $\mathrm{N}(0, 0.014^2)$ *distribution* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.782 | 0.792 | 0.732 | 0.643 | 0.412 | 0.251 |
| 144 | 1.000 | 1.000 | 1.000 | 0.326 | 0.274 | 0.149 |
| 288 | 1.000 | 1.000 | 1.000 | 0.171 | 0.171 | 0.103 |
| 576 | 1.000 | 1.000 | 1.000 | 0.091 | 0.128 | 0.079 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.832 | 0.844 | 0.789 | 0.724 | 0.483 | 0.309 |
| 144 | 1.000 | 1.000 | 1.000 | 0.408 | 0.362 | 0.199 |
| 288 | 1.000 | 1.000 | 1.000 | 0.232 | 0.241 | 0.156 |
| 576 | 1.000 | 1.000 | 1.000 | 0.138 | 0.186 | 0.133 |

[a] Notes: See notes to Table 1. The interval over which the statistics are calculated is $[u1, u2] = [\widehat{\mu}_{IV_t} - \beta\widehat{\sigma}_{IV_t}, \widehat{\mu}_{IV_t} + \beta\widehat{\sigma}_{IV_t}]$, where $\widehat{\mu}_{IV_t}$ and $\widehat{\sigma}_{IV_t}$ 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.

Table 3: Conditional Confidence Interval Accuracy Assessment: Level Experiments

**Case III: Jumps in DGP**[a]

| $M$ | $RV_{t,M}$ | $BV_{t,M}$ | $TPV_{t,M}$ | $\widehat{RV}_{t,l,M}$ | $\widetilde{RV}_{t,e,M}$ | $RK_{t,H,M}$ |
|---|---|---|---|---|---|---|
| *Panel A: Interval* $= \widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$ | | | | | | |
| *One i.i.d.* $\mathrm{N}(0, 3*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.237 | 0.194 | 0.209 | 0.639 | 0.509 | 0.396 |
| 144 | 0.210 | 0.171 | 0.145 | 0.436 | 0.409 | 0.283 |
| 288 | 0.191 | 0.138 | 0.134 | 0.332 | 0.350 | 0.216 |
| 576 | 0.181 | 0.143 | 0.108 | 0.246 | 0.299 | 0.194 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.296 | 0.272 | 0.276 | 0.721 | 0.588 | 0.467 |
| 144 | 0.255 | 0.228 | 0.196 | 0.502 | 0.491 | 0.340 |
| 288 | 0.235 | 0.182 | 0.184 | 0.398 | 0.427 | 0.283 |
| 576 | 0.230 | 0.184 | 0.164 | 0.308 | 0.370 | 0.252 |
| *One i.i.d.* $\mathrm{N}(0, 2*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 2 Days* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.380 | 0.244 | 0.230 | 0.660 | 0.580 | 0.486 |
| 144 | 0.348 | 0.169 | 0.152 | 0.511 | 0.518 | 0.386 |
| 288 | 0.323 | 0.129 | 0.122 | 0.415 | 0.460 | 0.370 |
| 576 | 0.303 | 0.129 | 0.141 | 0.366 | 0.424 | 0.340 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.450 | 0.301 | 0.289 | 0.734 | 0.648 | 0.576 |
| 144 | 0.410 | 0.215 | 0.202 | 0.597 | 0.597 | 0.474 |
| 288 | 0.378 | 0.185 | 0.169 | 0.484 | 0.529 | 0.444 |
| 576 | 0.368 | 0.174 | 0.182 | 0.447 | 0.503 | 0.413 |
| *One i.i.d.* $\mathrm{N}(0, 0.64*\widehat{\mu}_{IV_t})$ *Jump Every Day* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.478 | 0.250 | 0.201 | 0.663 | 0.606 | 0.570 |
| 144 | 0.452 | 0.184 | 0.171 | 0.492 | 0.584 | 0.534 |
| 288 | 0.432 | 0.134 | 0.138 | 0.473 | 0.552 | 0.467 |
| 576 | 0.430 | 0.119 | 0.129 | 0.425 | 0.512 | 0.462 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.548 | 0.310 | 0.258 | 0.730 | 0.675 | 0.653 |
| 144 | 0.512 | 0.229 | 0.211 | 0.563 | 0.663 | 0.605 |
| 288 | 0.502 | 0.187 | 0.178 | 0.548 | 0.621 | 0.545 |
| 576 | 0.512 | 0.168 | 0.177 | 0.489 | 0.595 | 0.542 |
| *Panel B: Interval* $= \widehat{\mu}_{IV_t} + 0.250\widehat{\sigma}_{IV_t}$ | | | | | | |
| *One i.i.d.* $\mathrm{N}(0, 3*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.325 | 0.244 | 0.260 | 0.833 | 0.760 | 0.575 |
| 144 | 0.274 | 0.156 | 0.156 | 0.666 | 0.648 | 0.412 |
| 288 | 0.231 | 0.138 | 0.137 | 0.464 | 0.502 | 0.303 |
| 576 | 0.235 | 0.110 | 0.119 | 0.322 | 0.447 | 0.263 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.415 | 0.308 | 0.328 | 0.880 | 0.834 | 0.655 |
| 144 | 0.368 | 0.205 | 0.213 | 0.754 | 0.720 | 0.509 |
| 288 | 0.322 | 0.204 | 0.199 | 0.558 | 0.593 | 0.381 |
| 576 | 0.305 | 0.157 | 0.171 | 0.401 | 0.528 | 0.346 |
| *One i.i.d.* $\mathrm{N}(0, 2*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 2 Days* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.499 | 0.248 | 0.229 | 0.851 | 0.794 | 0.666 |
| 144 | 0.425 | 0.153 | 0.145 | 0.689 | 0.709 | 0.571 |
| 288 | 0.421 | 0.118 | 0.116 | 0.538 | 0.631 | 0.506 |
| 576 | 0.377 | 0.089 | 0.107 | 0.500 | 0.581 | 0.444 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.589 | 0.328 | 0.303 | 0.909 | 0.858 | 0.745 |
| 144 | 0.526 | 0.207 | 0.186 | 0.761 | 0.782 | 0.653 |
| 288 | 0.505 | 0.161 | 0.160 | 0.624 | 0.714 | 0.590 |
| 576 | 0.461 | 0.136 | 0.161 | 0.585 | 0.647 | 0.523 |
| *One i.i.d.* $\mathrm{N}(0, 0.64*\widehat{\mu}_{IV_t})$ *Jump Every Day* | | | | | | |
| *Nominal Size = 5%* | | | | | | |
| 72 | 0.686 | 0.296 | 0.253 | 0.874 | 0.847 | 0.812 |
| 144 | 0.615 | 0.189 | 0.174 | 0.735 | 0.786 | 0.711 |
| 288 | 0.635 | 0.163 | 0.139 | 0.663 | 0.757 | 0.656 |
| 576 | 0.608 | 0.137 | 0.113 | 0.583 | 0.728 | 0.651 |
| *Nominal Size = 10%* | | | | | | |
| 72 | 0.752 | 0.372 | 0.316 | 0.914 | 0.897 | 0.860 |
| 144 | 0.701 | 0.260 | 0.248 | 0.814 | 0.841 | 0.799 |
| 288 | 0.719 | 0.223 | 0.203 | 0.751 | 0.830 | 0.742 |
| 576 | 0.686 | 0.187 | 0.173 | 0.660 | 0.790 | 0.721 |

[a] Notes: See notes to Table 1. The interval over which the statistics are calculated is $[u1, u2] = [\widehat{\mu}_{IV_t} - \beta\widehat{\sigma}_{IV_t}, \widehat{\mu}_{IV_t} + \beta\widehat{\sigma}_{IV_t}]$, where $\widehat{\mu}_{IV_t}$ and $\widehat{\sigma}_{IV_t}$ 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.

Table 4: Directional predictions results: $M = 2340$.

| Realized Measure | Conditioning Variable | Percentage of correct predictions using the same Realized Measure | Percentage of correct predictions using a benchmark Measure[a] |
|---|---|---|---|
| $RV$ | $RV_T$ | .52 | .44 |
| | $\overline{RV}_T$ | .58 | .42 |
| | $RS_T^-$ | .54 | .46 |
| | $\overline{RS}_T^-$ | .58 | .42 |
| $TPV$ | $TPV_T$ | .50 | .32 |
| | $\overline{TPV}_T$ | .56 | .38 |
| | $RS_T^-$ | .54 | .36 |
| | $\overline{RS}_T^-$ | .56 | .38 |
| $TSRV$ | $TSRV_T$ | .50 | .48 |
| | $\overline{TSRV}_T$ | .60 | .54 |
| | $RS_T^-$ | .76 | .70 |
| | $\overline{RS}_T^-$ | .66 | .60 |
| $MSRV$ | $MSRV_T$ | .52 | .50 |
| | $\overline{MSRV}_T$ | .62 | .56 |
| | $RS_T^-$ | .74 | .68 |
| | $\overline{RS}_T^-$ | .66 | .60 |
| $RK$ | $RK_T$ | .52 | .46 |
| | $\overline{RK}_T$ | .58 | .52 |
| | $RS_T^-$ | .76 | .70 |
| | $\overline{RS}_T^-$ | .67 | .60 |

[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: $M = 78$.

| Realized Measure | Conditioning Variable | Percentage of correct predictions using the same Realized Measure | Percentage of correct predictions using a benchmark Measure[a] |
|---|---|---|---|
| $RV$ | $RV_T$ | .50 | .50 |
| | $\overline{RV}_T$ | .58 | .58 |
| | $RS_T^-$ | .60 | .60 |
| | $\overline{RS}_T^-$ | .62 | .62 |
| $TPV$ | $TPV_T$ | .68 | .54 |
| | $\overline{TPV}_T$ | .72 | .66 |
| | $RS_T^-$ | .68 | .58 |
| | $\overline{RS}_T^-$ | .70 | .60 |
| $TSRV$ | $TSRV_T$ | .54 | .52 |
| | $\overline{TSRV}_T$ | .58 | .58 |
| | $RS_T^-$ | .62 | .62 |
| | $\overline{RS}_T^-$ | .66 | .66 |
| $MSRV$ | $MSRV_T$ | .56 | .48 |
| | $\overline{MSRV}_T$ | .64 | .56 |
| | $RS_T^-$ | .62 | .58 |
| | $\overline{RS}_T^-$ | .62 | .58 |
| $RK$ | $RK_T$ | .66 | .58 |
| | $\overline{RK}_T$ | .70 | .60 |
| | $RS_T^-$ | .68 | .66 |
| | $\overline{RS}_T^-$ | .66 | .66 |

[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.