Corradi, Valentina; Distaso, Walter; Swanson, Norman R.

Working Paper

Predictive inference for integrated volatility

Working Paper, No. 2011-09

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Corradi, Valentina; Distaso, Walter; Swanson, Norman R. (2011) : Predictive inference for integrated volatility, Working Paper, No. 2011-09, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/59492

Predictive Inference for Integrated Volatility*

Valentina Corradi†
University of Warwick

Walter Distaso‡
Imperial College London

Norman R. Swanson§
Rutgers University

October 2010

Abstract

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

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

*We thank the Associate Editor and two referees for very constructive comments. We also 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, RES-062-23-0311 and RES-062-23-0790, and Swanson acknowledges financial support from a Rutgers University Research Council grant.

†University of Warwick, Department of Economics, Coventry CV4 7AL, UK, email: v.corradi@warwick.ac.uk.

‡Imperial College London, Business School, Exhibition Road, London SW7 2AZ, UK, email: w.distaso@imperial.ac.uk.

§Rutgers University, Department of Economics, 75 Hamilton Street, New Brunswick, NJ 08901, USA, email: nswanson@econ.rutgers.edu.

1 Introduction

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

Given the development of this new class of financial instruments, it is of interest to construct conditional (predictive) volatility densities, rather than just point forecasts thereof. In this paper, we develop a method for constructing nonparametric conditional densities and confidence intervals for daily volatility, given observed market information. We show that the proposed estimators are consistent and asymptotically normally distributed, under mild assumptions on the underlying diffusion process.

The intuition for the approach taken in the paper is the following. Since integrated volatility is unobservable, we replace it with an estimator constructed using intra day returns. Our density estimators are therefore based on a variable which is subject to measurement error. We provide sufficient conditions under which conditional density (and confidence interval) estimators based on (the unobservable) integrated volatility and ones based on realized measures are asymptotically equivalent, so that measurement error is asymptotically negligible.

The first estimator of integrated volatility is realized volatility, concurrently devel-

oped by Andersen, Bollerslev, Diebold and Labys (2001), and Barndorff-Nielsen and Shephard (2002). Several variants of realized volatility have subsequently been proposed, motivated by the need of taking into account the presence of jumps or market microstructure noise. Examples include multipower variation (Barndorff-Nielsen and Shephard, 2004), 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 and Xiu, 2010) and estimators that are robust to both jumps and noise (Fan and Wang, 2007, Podolskij and Vetter, 2009). 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. The idea of using a realized measure as a basis for predicting integrated volatility has been also adopted in other papers. Andersen, Bollerslev, Diebold and Labys (2001, 2003), Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev and Meddahi (2004, 2005) deal with the problem of pointwise prediction of integrated volatility, using ARMA models based on realized volatility. 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 (2009) was the first paper to focus on estimation of the conditional density of integrated volatility, by establishing uniform rates of convergence for kernel estimators based on realized measures. However, with regard to notions such as hedging derivatives based on volatility, the crucial question becomes how to assess the interval within which future daily volatility will fall, with a given level of confidence. In this respect, the uniform convergence result of Corradi, Distaso and Swanson (2009) is not sufficient. This paper provides an

answer to this sort of questions by establishing asymptotic normality for estimators of conditional confidence intervals. This is a substantially more challenging task, as the realized measures and hence the measurement error are arguments of the uniform kernel, which is non-differentiable, and thus standard mean value expansion tools are no longer usable. Moreover, the current paper deals with the general class of *cadlag* (right continuous with left limit) volatility processes. This makes the computation of the moment structure of the measurement error much more complicated.

In order to assess the finite sample behavior of our statistics, we carry out a Monte Carlo experiment in which *pseudo true* predictive intervals are used in conjunction with intervals based on various realized measures. An empirical application to volatility directional predictability, based on New York Stock Exchange data, highlights the potential of our method and reveals the informational content of different volatility estimators.

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

2 Setup and main results

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

$$\mathrm{d}Y_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}J_t, \tag{1}$$

where the drift μ_t is a predictable process, the diffusion term σ_t is a *cadlag* process, W_t is a standard Brownian motion and J_t denotes a finite activity jump process. This

specification is very general, and (for example) allows for jump activity in volatility, stochastic volatility and leverage effects. We introduce market frictions by assuming that the observed log-price process is given by $X = Y + \epsilon$. Finally, we assume that there are a total of MT observations from the process X, consisting of M intradaily observations for T days, viz:

$$X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}, \;\; t = 0, \ldots, T \;\text{ and }\; j = 1, \ldots, M. \tag{2}$$

Daily integrated volatility is defined as $IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s, \;\; t = 1, \ldots, T$. Since IV_t is not observable, different realized measures, based on the sample $X_{t+j/M}$, are used as proxies for IV_t. Each realized measure, $RM_{t,M}$, will have an associated estimation error $N_{t,M}$, i.e.:

$$RM_{t,M} = IV_t + N_{t,M}. \tag{3}$$

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

$$\begin{aligned}\widehat{CI}_{T,M}(u_1, u_2 | RM_{T,M}) &= \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_2|RM_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}}(u_1|RM_{T,M}) \\ &= \frac{\frac{1}{T\xi}\sum_{t=0}^{T-1} 1_{\{u_1 \le RM_{t+1,M} \le u_2\}} K\left(\frac{RM_{t,M} - RM_{T,M}}{\xi}\right)}{\frac{1}{T\xi}\sum_{t=0}^{T-1} K\left(\frac{RM_{t,M} - RM_{T,M}}{\xi}\right)}; \end{aligned} \tag{4}$$

and for conditional densities:

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

Here, K is a kernel function, and ξ, ξ_1 and ξ_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) $f(\cdot)$ and, for any fixed x, $f(x|\cdot)$ are absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}_+$, and at least twice continuously differentiable. (ii) For any fixed x, u and z, $f(z) > 0$, $f(x|z) > 0$, and $0 < F(u|z) < 1$.

Assumption A4: There exists a sequence b_M, with $b_M \to \infty$, as $M \to \infty$, such that $\mathrm{E}\left(|N_{t,M}|^k\right) = O\left(b_M^{-k/2}\right)$, 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. Assumption A1 is, for example, satisfied when the volatility process is generated by a diffusion:

$$\mathrm{d}\sigma_t^2 = b_1\left(\sigma_t^2\right)\mathrm{d}t + b_2\left(\sigma_t^2\right)\mathrm{d}W_t, \tag{5}$$

provided the drift condition in Meyn and Tweedie (1993, p.536) is satisfied:

$$2\sigma^2 b_1(\sigma^2) + b_2^2\left(\sigma^2\right) \leq -c_1\left(\sigma^2\right)^2 + c_2,$$

where c_1, c_2 are positive constants. The drift condition is met when drift and variance terms in (5) grow at most at a linear rate, and if there is mean reversion. Assumption A1 is also satisfied when a jump component is added to the diffusion in (5), see e.g. Masuda (2007). A2 and A3 are standard assumptions in the literature on nonparametric density estimation. 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 infeasible ones (based on the latent volatility). In Section 3 we shall

provide primitive conditions under which A4 is satisfied by the most commonly used realized measures. For conditional confidence intervals, we have the following result.

Theorem 1. *Let A1-A4 hold. Then:*

(i)

$$\begin{aligned} & \sqrt{T\xi}\left(\widehat{CI}_{T,M}(u_1,u_2|RM_{T,M}) - CI(u_1,u_2|RM_{T,M})\right) \\ = \; & \sqrt{V(u_1,u_2)}Z + \sqrt{T\xi^5}\frac{1}{2}C_2(K)b(u_1,u_2) + O_p\left(\xi^{-1}b_M^{-1/2} + b_M^{-1/4}T^{\frac{3}{4k}} + T^{\frac{3+k}{2k}}\xi^{1/2}b_M^{-1/2}\right), \end{aligned}$$

where Z is a standard normal, $C_2(K) = \int u^2 K(u)\mathrm{d}u$,

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

$$b(u_1,u_2) = \left.\frac{\partial^2\left(F(u_2|z) - F(u_1|z)\right)}{\partial z^2}\right|_{z=RM_{T,M}} + 2\frac{f^{(1)}(RM_{T,M})}{f(RM_{T,M})}\left.\frac{\partial\left(F(u_2|z) - F(u_1|z)\right)}{\partial z}\right|_{z=RM_{T,M}}.$$

(ii) If either (a) $T\xi^3 \to \infty$, $T\xi^5 \to 0$ and $\max\left\{b_M^{-1/2}T^{\frac{3}{2k}}, T^{\frac{3+k}{k}}\xi b_M^{-1}\right\} \to 0$ or (b) $T\xi \to \infty$, $T\xi^5 \to 0$ and $\max\left\{b_M^{-1/2}T^{\frac{3}{2k}}, T^{\frac{3+k}{k}}\xi b_M^{-1}, b_M^{-1}\xi^{-2}\right\} \to 0$, then

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

The relative orders of magnitude of M, T, ξ are discussed in Remark 6 below. 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:

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

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

Theorem 2. *Let A1-A4 hold. Then:*

(i)

$$\begin{aligned}
&\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{T,M}(x|RM_{T,M}) - f(x|RM_{T,M})\right) \\
= &\left(\sqrt{\frac{f(x|RM_{T,M})}{f(RM_{T,M})}}\int K^2(u)\mathrm{d}u\right) Z + \frac{1}{2}\sqrt{T\xi_1^5\xi_2}C_2(K)\frac{\partial^2 f(x|RM_{T,M})}{\partial x^2} \\
&+\frac{1}{2}\sqrt{T\xi_1\xi_2^5}C_2(K)\left(\frac{\partial^2 f(x|z)}{\partial z^2}\bigg|_{z=RM_{T,M}} + 2\frac{\frac{\partial f(x|z)}{\partial z}\Big|_{z=RM_{T,M}}\frac{\partial f(z)}{\partial z}\Big|_{z=RM_{T,M}}}{f(RM_{T,M})}\right) \\
&+O_p\left(\sqrt{T\xi_1\xi_2}b_M^{-1/2} + \xi_1^{-1}b_M^{-1/2} + \xi_2^{-1}b_M^{-1/2}\right).
\end{aligned}$$

(ii) If either (a) $T\xi_1^3\xi_2 \to \infty$, $T\xi_1\xi_2^3 \to \infty$, $T\xi_1^5\xi_2 \to 0$, $T\xi_1\xi_2^5 \to 0$, $T\xi_1\xi_2 b_M^{-1} \to 0$, *or (b)* $T\xi_1\xi_2 \to \infty$, $T\xi_1^5\xi_2 \to 0$, $T\xi_1\xi_2^5 \to 0$, $\max\left\{T\xi_1\xi_2 b_M^{-1}, b_M^{-1}\xi_1^{-2}, b_M^{-1}\xi_2^{-2}\right\} \to 0$, *then*

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

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

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

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

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. Similarly, local linear conditional density estimation (see Fan, Yao and Tong, 1996), are derived from:

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

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

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. (i) Then:*

$$
\begin{aligned}
&\sqrt{T\xi}\left(\widehat{\alpha}_{0,T,M}(u_1,u_2,RM_{T,M})-CI(u_1,u_2|RM_{T,M})\right) \\
=\;&\sqrt{V(u_1,u_2)}Z+\sqrt{T\xi^5}\frac{1}{2}C_2(K)\left.\frac{\partial^2\left(F(u_2|z)-F(u_1|z)\right)}{\partial z^2}\right|_{z=RM_{T,M}} \\
&+O_p\left(\xi^{-1}b_M^{-1/2}+b_M^{-1/4}T^{\frac{3}{4k}}+T^{\frac{3+k}{2k}}\xi^{1/2}b_M^{-1/2}\right),
\end{aligned}
$$

and if either (a) $T\xi^3\to\infty$, $T\xi^5\to 0$ *and* $\max\left\{b_M^{-1/2}T^{\frac{3}{2k}}, T^{\frac{3+k}{k}}\xi b_M^{-1}\right\}\to 0$ *or (b)* $T\xi\to\infty$, $T\xi^5\to 0$ *and* $\max\left\{b_M^{-1/2}T^{\frac{3}{2k}}, T^{\frac{3+k}{k}}\xi b_M^{-1}, b_M^{-1}\xi^{-2}\right\}\to 0$, *then*

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

(ii)

$$
\begin{aligned}
&\sqrt{T\xi_1\xi_2}\left(\widehat{\beta}_{0,T,M}(x,RM_{T,M})-f(x|RM_{T,M})\right) \\
=\;&\left(\sqrt{\frac{f(x|RM_{T,M})}{f(RM_{T,M})}}\int K^2(u)\mathrm{d}u\right)Z+\frac{1}{2}\sqrt{T\xi_1^5\xi_2}C_2(K)\frac{\partial^2 f(x|RM_{T,M})}{\partial x^2} \\
&+\frac{1}{2}\sqrt{T\xi_1\xi_2^5}C_2(K)\left.\frac{\partial^2 f(x|z)}{\partial z^2}\right|_{z=RM_{T,M}}+O_p\left(\sqrt{T\xi_1\xi_2}b_M^{-1/2}+\xi_1^{-1}b_M^{-1/2}+\xi_2^{-1}b_M^{-1/2}\right),
\end{aligned}
$$

and, if either (a) $T\xi_1^3\xi_2\to\infty$, $T\xi_1\xi_2^3\to\infty$, $T\xi_1^5\xi_2\to 0$, $T\xi_1\xi_2^5\to 0$, $T\xi_1\xi_2 b_M^{-1}\to 0$, *or (b)* $T\xi_1\xi_2\to\infty$, $T\xi_1^5\xi_2\to 0$, $T\xi_1\xi_2^5\to 0$, $\max\left\{T\xi_1\xi_2 b_M^{-1}, b_M^{-1}\xi_1^{-2}, b_M^{-1}\xi_1^{-2}\right\}\to 0$, *then*

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

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

3 Applications to specific volatility estimators

We now provide primitive conditions on the moments of the drift, variance and noise, which ensure that Assumption A4 is satisfied by some commonly used realized measures, namely Two Scale Realized Volatility ($\widehat{RV}_{t,l,M}$, Zhang, Mykland and Aït-Sahalia, 2005):

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

with $M = lB$. Multi Scale Realized Volatility ($\widetilde{RV}_{t,\tau,M}$, Zhang, 2006):

$$\widetilde{RV}_{t,\tau,M} = \sum_{i=1}^{\tau}\frac{a_i}{i}\left(\sum_{j=1}^{M-i}\left(X_{t+\frac{j+i}{M}} - X_{t+\frac{j}{M}}\right)^2\right) + \frac{\sum_{j=1}^{M-1}\left(X_{t+\frac{j}{M}} - X_{t+\frac{j-1}{M}}\right)^2}{M},$$

with $a_i = 12i/\tau^2\left(i/\tau - 1/2 - 1/(2\tau)\right)/\left(1 - 1/\tau^2\right)$, $\sum_{i=1}^{\tau} a_i = 1$ and $\sum_{i=1}^{\tau} a_i/i = 0$. Realized kernels ($RK_{t,H,M}$, Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008):

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

where $\gamma_h^X = \sum_{j=H}^{M-H-1}\left(X_{t+(j+1)/M} - X_{t+j/M}\right)\left(X_{t+(j+1-h)/M} - X_{t+(j-h)/M}\right)$ and κ is a kernel function defined in Lemma 1.

Lemma 1. *Let Y_t follow (1) and ϵ be defined by (2). If $J_t \equiv 0$ for all t, $\mathrm{E}\left(\left(\sigma_t^2\right)^{2(k+\delta)}\right) < \infty$ and $\mathrm{E}\left(\left(\mu_t\right)^{2(k+\delta)}\right) < \infty$, with $\delta > 2$, then there is a sequence b_M, where $b_M \to \infty$ as $M \to \infty$, such that:*

(i) If $\epsilon_t \sim i.i.d.\left(0, \sigma_\epsilon^2\right)$, $\mathrm{E}\left(\epsilon_t^{2k}\right) < \infty$, $\mathrm{E}\left(\epsilon_t Y_t\right) = 0$, and $l/M^{1/3} = O(1)$, then $\mathrm{E}\left(\left|\widehat{RV}_{t,l,M} - IV_t\right|^k\right) = O(b_M^{-k/2})$, with $b_M = M^{1/3}$.

(ii) If $\epsilon_t \sim i.i.d.\left(0, \sigma_\epsilon^2\right)$, $\mathrm{E}\left(\epsilon_t^{2k}\right) < \infty$, $\mathrm{E}\left(\epsilon_t Y_t\right) = 0$, and $\tau/M^{1/2} = O(1)$, then $\mathrm{E}\left(\left|\widetilde{RV}_{t,\tau,M} - IV_t\right|^k\right) = O(b_M^{-k/2})$, with $b_M = M^{1/2}$.

(iii) *If* $\mathrm{E}\left(\epsilon_t^{2k}\right) < \infty$, $\mathrm{E}\left(\epsilon_t Y_t\right) = 0$, $\kappa(0) = 1$, $\kappa(1) = \kappa^{(1)}(0) = \kappa^{(1)}(1) = 0$, *and* $H/M^{1/2} = O(1)$, *then* $\mathrm{E}\left(\left|RK_{t,H,M} - IV_t\right|^k\right) = O(b_M^{-k/2})$, *with* $b_M = M^{1/2}$. *If in addition* $\kappa^{(2)}(0) = 0$ *and* $\mathrm{E}\left(\epsilon_t^{4k}\right) < \infty$, *the same statement holds for* $\epsilon_{t+j/M}$ *geometrically mixing, in the sense that, for any* j, *there exists a constant* $|\rho| < 1$ *such that* $\mathrm{E}\left(\epsilon_{t+j/M} \middle| \epsilon_{t+(j-s)/M}, \ldots\right) \approx \rho^s \epsilon_{t+(j-s)/M}$.

Note that (A.11) implies that Lemma 1 holds for realized volatility and $b_M = M$. It can be shown that it also holds for power variation measures with $b_M = M$. Part *(iii)* of the Lemma allows for some serial dependence in the microstructure noise, requiring an additional condition on the kernel function, i.e. $\kappa^{(2)}(0) = 0$. Such a condition is satisfied, for example, by fifth or higher order kernels. Some form of correlation between noise and price can be allowed, following the approach of Kalnina and Linton (2008) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008).

Remark 1. From a practical point of view, the asymptotic normality results stated in the Theorems are useful, as they facilitate the construction of confidence bands around estimated conditional densities and confidence intervals. The sort of empirical problem for which these results may be useful is the following. Suppose that we want to predict the probability that integrated volatility will take a value between u_1 and u_2, say, given actual information. Then, asymptotically, $\Pr\left((u_1 \leq IV_{T+1} \leq u_2)|IV_T = RM_{T,M}\right)$ will fall in the interval $(\widehat{F}_{T,M}(u_2|RM_{T,M}) - \widehat{F}_{T,M}(u_1|RM_{T,M})) \pm \widehat{V}^{1/2}(u_1, u_2) z_{\alpha/2}/\sqrt{T\xi}$, with probability $1-\alpha$, where $\widehat{V}(u_1, u_2)$ is an estimator of $\widehat{V}(u_1, u_2)$, as defined in Theorem 1, and $z_{\alpha/2}$ denotes the $\alpha/2$ quantile of a standard normal.

Remark 2. In empirical work, volatility is often modelled and predicted with ARMA models that are constructed using logs of realized volatility. For example, Andersen, Bollerslev, Diebold and Labys (2001, 2003) use the log of realized volatility for modelling and predicting stock returns and exchange rate volatility. According to these authors, one reason for using logs is that while the distribution of realized volatility is

highly skewed to the right, the distribution of logged realized volatility is much closer to normal. It is immediate to see that a Taylor expansion of $\log(RM_{t,M})$ around IV_t ensures that $\mathrm{E}\left(\left|\log(RM_{t,M}) - \log(IV_t)\right|^k\right) = O(b_M^{-k/2})$. Therefore, the statements in the theorems above hold in the case where we are interested in predictive densities and confidence intervals for the log of integrated volatility.

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

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

where $\mathbf{K}$ is a $d-$dimensional kernel function. Extension of the results of the theorems to cover this general case is straightforward (and available upon request), but is not reported for notational simplicity. To alleviate the curse of dimensionality, one may use (weighted) averages of past realized measures. This is the route followed in the empirical section.

Remark 4. So far we have assumed that IV_t is a stationary process. This allows prediction via a state domain smoothing approach. However, if volatility is also time inhomogeneous, then a time-domain approach, or a combination of a state and time domain approach becomes necessary. Nonparametric estimation of time inhomogeneous volatility over a finite time span has been considered by Florens-Zmirou (1989). Estimation of time varying diffusions over an increasing time span is a more challenging task, as one has to control for the degree of nonstationarity, ensuring that the law of large numbers and central limit theorem still apply. Fan, Jiang, Zhang and Zhou (2003) estimate diffusions with time-varying coefficients, under the assumption that they are locally constant. Koo and Linton (2009) established asymptotic nor-

mality of nonparametric estimators of the drift and variance of a diffusion, under the assumption of local stationarity. They consider diffusion processes like:

$$X_{t,T} = \int_0^t \mu(s/T, X_{s,T})\mathrm{d}s + \int_0^t \sigma(s/T, X_{s,T})\mathrm{d}W_s.$$

$X_{t,T}$ is locally stationary if, for t/T in a neighborhood of v, it can be well approximated by a diffusion:

$$X_{v,t} = \int_0^t \mu(v, X_{v,s})\mathrm{d}s + \int_0^t \sigma(v, X_{v,s})\mathrm{d}W_s,$$

which, for any point $v \in [0,1]$, is stationary and, under usual conditions on μ and σ, $\alpha-$mixing. The corresponding integrated volatility $IV_{t,T} = \int_{t-1}^t \sigma^2(s/T, X_{s,T})\mathrm{d}s$ can be accurately approximated, for t/T in a neighborhood of v, by $IV_{v,t} = \int_{t-1}^t \sigma^2(v, X_{v,s})\mathrm{d}s$. Then, $\Pr\left(IV_{t+1,T} \leq u | IV_{t,T} = x, t/T = v\right)$ can be estimated by:

$$\widehat{F}_T\left(u|x,v\right) = \frac{\frac{1}{T}\sum_{t=1}^{T-1} 1_{\{IV_{t+1,T}\leq u\}} K\left(\frac{IV_{t,T}-x}{\xi}\right) K\left(\frac{t/T-v}{h}\right)}{\frac{1}{T}\sum_{t=1}^{T-1} K\left(\frac{IV_{t,T}-x}{\xi}\right) K\left(\frac{t/T-v}{h}\right)}. \tag{6}$$

A feasible version of (6) can be implemented by replacing $IV_{t+1,T}$ with its realized counterpart, $RM_{t,M}$. Provided the measurement error satisfies Assumption A4, the statement in Theorem 1 follows, by simply modifying the rate of convergence, in order to take into account the double domain and time smoothing. If we are interested in out of sample prediction, then we need to evaluate t/T around $v = 1$. This poses a clear boundary issue, which can be overcome by using a local polynomial estimator, with both time and domain smoothing, i.e. by computing:

$$\begin{aligned} & \widehat{\boldsymbol{\alpha}}_{T,M}(u, RM_{T,M}, v) \\ = \; & \arg\min_{\boldsymbol{\alpha}} \frac{1}{T\xi h} \sum_{t=1}^{T-1} \left(1_{\left\{RM_{t+1,M}\leq u\right\}} - \alpha_0 - \alpha_1\left(RM_{t,M} - RM_{T,M}\right) - \alpha_2\left(t/T - v\right)\right)^2 \\ & K\left(\frac{RM_{t,M} - RM_{T,M}}{\xi}\right) K\left(\frac{t/T - v}{h}\right), \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)'$ and $\widehat{\alpha}_{0,T,M}(u, x, v)$ defines the local polynomial estimator.

Remark 5. There is a well developed statistical literature on kernel estimation of densities and regression functions in the presence of measurement error. For example, in the case of *i.i.d.* observations, Stefanski (1990) and Fan (1991, 1992) derive rates of convergence for kernel marginal density estimators, and Fan and Truong (1993) for Nadaraya-Watson regression function estimators, with a one-dimensional covariate. Extension to dependent observations and to the multivariate case have been provided by Masry (1991) for joint density estimation and by Fan and Masry (1992) for regression functions. Recently, Delaigle, Fan and Carroll (2009) have derived asymptotic normality results for local polynomial estimators. This is a more challenging task, as the variable measured with error enters not only in the kernel function but also in the polynomial approximation. The common assumptions in the papers above are that the error is independent of the "true" latent variable, and that its density is known in closed form. Alternatively, the density of the error can be estimated, provided we have repeated measurements on the contaminated series, and provided the errors across different measurements are independent (see, e.g., Li and Vuong, 1998 and Schennach, 2004).

Nonparametric estimation in the presence of measurement error hinges on the idea of replacing standard kernels with deconvolution kernels, whose construction requires the knowledge, or a proper estimate, of the characteristic function of the error. Most of the papers cited above consider the case of a non-vanishing error. While deconvolution estimators ensure that the bias is of the same order as in the error-free case, the order of magnitude of the variance, hence the rate of convergence, depends on the smoothness of the error density. For smooth densities, such as the normal density, the convergence rate is logarithmic, thus making the deconvolution approach not really useful in practice. An important exception is Fan (1992, Theorem 4), who shows that, if the variance of the error approaches zero at least as fast as the

bandwidth parameter, then the error-free nonparametric optimal rate still applies.

We now explicitly compare the asymptotic properties of standard and deconvolution kernel estimators. For simplicity, we focus on the joint density. Let $\xi_1 = \xi_2 = \xi$ and $N_{t,M} = b_M^{-1/2} v_t$. Also, assume (i) $\mathrm{E}(v_t) = 0$, (ii) $\mathrm{var}(v_t) = \sigma_v^2$ and (iii) $\mathrm{E}(v_t IV_t) = 0$. Note that (i) always holds when the drift term in (1) is zero, (ii) follows from Assumption A4 for $k = 2$, and (iii) holds when there is no leverage (see, e.g., Corollary 2.1 in Meddahi, 2002). Finally, assume that characteristic function of (v_t, v_{t+1}) is known and let $\mathrm{var}(v_t, v_{t+1}) = \mathbf{\Sigma}$. Then, for $s_t = (RM_{t,M} - RM_{T,M})/\xi$ and $s_{t+1} = (RM_{t+1,M} - x)/\xi$ the deconvolution kernel estimator writes as:

$$\widehat{f}_{T,M}^{(c)}(x, RM_{T,M}) = \frac{1}{T\xi^2} \sum_{t=1}^{T-1} K^{(c)}(s_t) K^{(c)}(s_{t+1}),$$

where:

$$K^{(c)}(s_t) K^{(c)}(s_{t+1}) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-\mathrm{i}(\tau_1 s_t + \tau_2 s_{t+1})) \frac{\phi_K(\tau_1)\phi_K(\tau_2)}{\phi_v\left(\mathbf{\Sigma}^{1/2}\boldsymbol{\tau}\frac{b_M^{-1/2}}{\xi}\right)} \mathrm{d}\tau_1 \mathrm{d}\tau_2,$$

$\boldsymbol{\tau} = (\tau_1, \tau_2)'$, $\phi_K(\tau_1)$ and $\phi_K(\tau_2)$ denote, respectively, the Fourier transforms of $K^{(c)}(s_t)$ and $K^{(c)}(s_{t+1})$, and $\phi_v\left(\mathbf{\Sigma}^{1/2}\boldsymbol{\tau} b_M^{-1/2}/\xi\right)$ denotes the characteristic function of $b_M^{-1/2}(v_t, v_{t+1})$. When $\phi_v\left(\mathbf{\Sigma}^{1/2}\boldsymbol{\tau} b_M^{-1/2}/\xi\right) = 1$, the deconvolution estimator coincides with the usual kernel estimator. If $b_M^{-1/2} = O(\xi)$, then it follows from Theorem 4 in Fan (1992) that the (integrated) mean square error is of the same order as in the error free case. In particular, the squared bias is the same as in the error free case, and the variance is given by:

$$\frac{1}{(2\pi)^2} f(x, RM_{T,M}) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{|\phi_K(\tau_1)\phi_K(\tau_2)|^2}{\left|\phi_v\left(\mathbf{\Sigma}^{1/2}\boldsymbol{\tau}\frac{b_M^{-1/2}}{\xi}\right)\right|^2} \mathrm{d}\tau_1 \mathrm{d}\tau_2 + o(1).$$

Hence, if $b_M^{-1}\xi^{-2} \longrightarrow 0$, $\phi_v\left(\mathbf{\Sigma}^{1/2}\boldsymbol{\tau} b_M^{-1/2}/\xi\right) \longrightarrow 1$, and the asymptotic variance is the same as in the error-free case. In Theorem 2, we require $\max\left\{T\xi^2 b_M^{-1}, b_M^{-1}\xi^{-2}\right\} \to 0$,

which clearly implies $b_M^{-1}\xi^{-2} \to 0$. Therefore, using deconvolution methods, for any given T, we may achieve the same degree of precision with a smaller number of intraday observations, M. However, in empirical applications the use of deconvolution kernels is not practical. In addition to the fact that (i) and (iii) require zero drift and no leverage, the main obstacle is that we do not know the functional form of the density of the measurement error. Indeed, we have several measurements, as we can use different realized volatility measures. However, the measurement errors across different realized measures are highly correlated, and this makes the estimation of the error characteristic function not viable.

Remark 6. As stated in Lemma 1, k depends on the number of finite moments of the instantaneous drift and volatility function. For k large, $T^{\frac{3}{4k}} b_M^{-1/4} \to 0$, regardless of the rate at which M diverges, and $T^{\frac{3+k}{k}} \xi b_M^{-1} \simeq T\xi b_M^{-1}$.

If we express the bandwidth as a function of the number of daily observations, i.e. $\xi = cT^{\alpha}$, $\alpha \in (-1/5, -1/3)$, then we require b_M to grow at least at rate $T^{1-\alpha}$. Hence, if we ignore the measurement error, and choose the bandwidth optimally ($\xi = cT^{-1/5}$), we require b_M grow at least a rate $T^{4/5}$. If instead we allow for a certain degree of undersmoothing, e.g. $\xi = cT^{-1/3}$, it suffices that b_M grows at least as fast as $T^{2/3}$.

Alternatively, if we express the bandwidth in terms of the number of intradaily observations, i.e. $\xi = cb_M^{-\gamma}$, $\gamma < 1/2$, then we need that $Tb_M^{-(1+\gamma)} \to 0$ and b_M must diverge at least at rate $T^{2/3}$. Hence, the optimal convergence rate $T^{2/5}$ is achievable for b_M diverging faster than $T^{4/5}$. If instead b_M grows at rate $T^{2/3}$, we need to set $\xi = O(T^{-1/3})$, and can achieve only a convergence rate of $T^{1/3}$. Finally, if b_M grows at a rate slower than $T^{2/3}$, the feasible estimator is not asymptotically equivalent to the infeasible one. As established in Lemma 1, b_M differs according to the specific chosen volatility estimator. In practice, the choice of volatility estimator is an empirical question and depends on data availability, and market liquidity/activity. When T is

much larger than b_M it may be preferable to allow for some undersmoothing and set $\xi = cT^{-1/3}$.

Remark 7. Bandwidth selection in our context is not a trivial issue. First, usual automated procedures for Nadaraya-Watson estimators (such as cross-validation) and for local polynomial estimators (e.g., the residual square criterion of Fan and Gijbels, 1995), do not necessarily lead to the choice of the same bandwidth parameter as in the case of no measurement error. In fact, given the rate conditions in Theorems 1-3, the contribution of measurement error may be negligible for some values of the bandwidth parameter and non negligible for others. As a consequence, the bandwidth minimizing a given criterion may depend on both M and T. Second, in order to ensure asymptotic equivalence between the feasible and infeasible estimators, it may be necessary to select a bandwidth approaching zero at a rate faster than the optimal one. A possible solution is a two-step procedure, in the spirit of Bandi, Corradi and Moloche (2009). Consider the case of Nadaraya-Watson estimation of the conditional distribution. In the first step, we can select a bandwidth via cross-validation, i.e.:

$$\widehat{\xi}_{T,M} = \arg\min_{\xi} \int_U \frac{1}{T} \sum_{t=1}^{T-1} \left(1_{\{RM_{t+1,M} \leq u\}} - \widehat{F}_{T,M,\xi,-t}(u|RM_{t,M}) \right)^2 \mathrm{d}u,$$

where $\widehat{F}_{T,M,\xi,-t}(u|RM_{t,M})$ is the feasible estimator of $F(u|RM_{t,M})$ constructed using bandwidth ξ, and leaving out the $t-$th observation.

Under Assumption A1, for all M and for all $\xi \to 0$, $1/\xi \sum_{s \neq t}^{T-1} K\left((RM_{s,M} - RM_{t,M})/\xi\right) = O_{a.s.}(T)$. We can therefore use the randomized procedure of Bandi, Corradi and Moloche (2009) to check whether $\widehat{\xi}_{T,M}$ satisfies the rate conditions in Theorem 1. Assuming a sufficiently large k, the additional rate condition due to measurement error reduces to $T\xi^3 \to \infty$ and $Tb_M^{-1}\xi \to 0$. Hence, we can check whether $\widehat{\xi}_{T,M}$ is large enough to ensure that $\widehat{\xi}_{T,M}^2 \sum_{t=1}^{T-1} K\left((RM_{t,M} - x)/\widehat{\xi}_{T,M}\right) \xrightarrow{a.s.} \infty$, and small enough to ensure that $\max\left\{\widehat{\xi}_{T,M}^4 \sum_{t=1}^{T-1} K\left((RM_{t,M} - x)/\widehat{\xi}_{T,M}\right), b_M^{-1} \sum_{t=1}^{T-1} K\left((RM_{t,M} - x)/\widehat{\xi}_{T,M}\right)\right\} \xrightarrow{a.s.}$

0. The outcome of the second step tells us whether we can keep $\widehat{\xi}_{T,M}$ or whether we should search for a smaller (larger) bandwidth.

4 Monte Carlo Results

In this section, we evaluate the finite sample properties of:

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

We consider the following data generating process (DGP):

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

where $W_{1,t}$ and $W_{2,t}$ are two correlated Brownian motions, with $\mathrm{corr}(W_{1,t},W_{2,t})=\rho$. Following Aït-Sahalia and Mancini (2008), we set $m=0.05$, $\psi=5$, $\upsilon=0.04$, $\eta=0.5$, and $\rho=-0.5$. Because 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), using a Milstein scheme with a discrete interval $1/N$, keeping the conditioning value at period 1 fixed across simulations. We then construct confidence intervals from the empirical distribution of the simulated integrated variance. By setting S and N sufficiently large (3000 and 2880, respectively), the effects of both simulation and discretization error are negligible. This gives us a simulation based estimator of $CI(u_1,u_2|RM_{T,M})$.

We then construct time series of length T of realized volatility measures sampling the simulated data at frequency $1/M$. For the first day, we use, across all replications, the same draws used for the construction of the (simulation based) estimator of the confidence interval described above. In order to avoid boundary bias problems, we form $G_{T,M}(u_1,u_2)$ using Gaussian kernels on a logarithmic transformation of our daily series. In our base case (denoted by Case I), we simply set

$X_{t+j/M} = Y_{t+j/M}$. In Case II, daily data are generated by adding microstructure noise. Namely, we generate $X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}$, where $\epsilon_{t+j/M} \sim i.i.d.$ N$(0, \sigma_\epsilon^2)$, and $\sigma_\epsilon^2 = \{(0.005^2), (0.007^2), (0.014^2)\}$. A standard deviation of 0.007 corresponds to the case where the standard deviation of the noise is approximately 0.1% of the value of the asset price (this is the same percentage as that used in Aït-Sahalia and Mancini, 2008). Finally, in Case III, jumps are added by including an $i.i.d.$ N$(0, 0.64 a_{jump} \widehat{\mu}_{IV})$ shock to the process for $Y_{t+j/M}$, where a_{jump} is set equal to $\{3, 2, 1\}$, and $\widehat{\mu}_{IV}$ is the average of (the log of) IV_t over S. In this case, it is assumed that jumps arrive randomly with equal probability at any point in time, once every 5 days when $a_{jump} = 3$, once every 2 days $a_{jump} = 2$, and every day when $a_{jump} = 1$, on average. We consider the interval $[u_1, u_2] = [\widehat{\mu}_{IV} - \beta\widehat{\sigma}_{IV}, \widehat{\mu}_{IV} + \beta\widehat{\sigma}_{IV}]$, where $\widehat{\sigma}_{IV}$ is the standard deviation of (the log of) IV_t over S, and $\beta = \{0.125, 0.250\}$, for different values of T and M, i.e. $T = \{100, 300, 500\}$ and $M = \{72, 144, 288, 576\}$. Results are based upon 10,000 Monte Carlo iterations. For brevity, we report our findings only for $T = 100$.

Tables 1, 3-4 report rejection frequencies using two-sided 5% and 10% nominal level critical values. In these tables, results are based on choosing the bandwidth parameter according to Silverman's (1986) rule. The six columns of entries contain results for realized volatility ($RV_{t,M}$), bi ($BV_{t,M}$) and tripower variation ($TPV_{t,M}$), $RV_{t,l,M}$, $\widetilde{RV}_{t,\tau,M}$, $RK_{t,H,M}$, 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).

Turning first to Table 1, where there is neither microstructure noise nor jumps, note that $RV_{t,M}$, $BV_{t,M}$, and $TPV_{t,M}$ perform approximately equally well for large values of M, although $RV_{t,M}$ performs marginally better than the other estimators in a number of instances, as might be expected. In particular, use of these estimators yields empirical sizes close to the nominal 5% and 10% levels in various cases,

and there is a substantial improvement as M increases. Overall, $RV_{t,M}$, $BV_{t,M}$ and $TPV_{t,M}$ yield more accurate confidence intervals than the other three (robust) measures, although improvements associated with using these three estimators drops off sharply for the largest two values of M. In particular, note that rejection frequencies at the nominal 10% level for $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,\tau,M}$, and $RK_{t,H,M}$ are often 0.20-0.70 when $M = 72$ and 144, whereas rates for $RV_{t,M}$, $BV_{t,M}$, and $TPV_{t,M}$ are generally rather closer to 0.10. Indeed, empirical performance of $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,\tau,M}$, and $RK_{t,H,M}$ is quite poor for very small values of M, but improves quite quickly as M increases. Additionally, $RK_{t,H,M}$ and $\widetilde{RV}_{t,\tau,M}$ perform substantially better than $\widehat{RV}_{t,l,M}$ in virtually all cases, although the relative difference in performance shrinks as M increases. Finally, $RK_{t,H,M}$ and $\widetilde{RV}_{t,\tau,M}$ perform approximately equally well for all values of M and T. Overall, there is clearly a need for reasonably large values of M when implementing the microstructure robust realized measures. This is not surprising, given the slower rate of convergence of these estimators.

We now turn to Table 3, where microstructure noise is added to the frictionless price. It is immediate to see that $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,\tau,M}$ and $RK_{t,H,M}$ are superior to the non robust realized measures, for large values of M, as expected. For example, consider Panel B. The rejection frequencies at the nominal 10% level for $RV_{t,M}$ range from 0.16 up to 1.0, when $M = 288$, depending upon the magnitude of the noise volatility. On the other hand, comparable rejection frequencies for $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,\tau,M}$ and $RK_{t,H,M}$ range from 0.13-0.22, which indicates a marked improvement when using robust measures, as long as M is large. 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 $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,\tau,M}$ and $RK_{t,H,M}$ rejection frequencies. This is hardly surprising, given that, from Lemma 1, b_M grows at a rate slower than M in the case of microstructure noise robust realized measures. It follows

that for empirical implementation, one may select either a relatively small value of M, for which the microstructure noise effect is not too distorting, together with a non microstructure robust realized measure, or select a very large value of M and a microstructure robust realized measure. Interestingly, we see in our experiments that there is little to choose between the best performing of our robust measures (i.e. $\widetilde{RV}_{t,\tau,M}$ and $RK_{t,H,M}$), and that these two measures outperform $RV_{t,M}$ in many cases for values of M as small as 144, which suggests that the relative gains associated with using robust measures are achieved very quickly as M increases. Finally, consider Table 4, where jumps are added to the price. $BV_{t,M}$ and $TPV_{t,M}$ yield similar results, both outperform all "noise-robust measures", and their relative performance improves with the jump frequency.

In order to check how sensitive our results are with respect to the bandwidth parameter, we have performed a Monte Carlo exercise, and results are reported in Table 2. In the table, we let the bandwidth parameter vary as a proportion of the one chose according to Silverman's (1986) rule. Hence, the case where the bandwidth proportion equals one corresponds to the entries reported in Tables 1 and 3. In the no microstructure case results seem to be fairly robust with respect to the bandwidth parameter and Silverman's bandwidth is often the best performing one. When microstructure noise is included, results become substantially worse when undersmoothing is severe and non robust measures are used. The best performing bandwidth is Silverman's for realized volatility and power variation based measures, whereas a bit of undersmoothing is beneficial for noise-robust measures.

5 Volatility Predictive Intervals for Intel

In this section we construct and examine predictions of the conditional distribution of daily integrated volatility for Intel. Data are taken from the Trade and Quotation

database at the New York Stock Exchange. Our sample size covers 150 trading days starting from January 2 2002. From the original data set, we extracted 10 second and 5 minute interval data, using bid-ask midpoints and the last tick method (see Wasserfallen and Zimmermann, 1985). Provided that there is sufficient liquidity in the market, the 5 minute frequency seems to offer a reasonable compromise between minimizing the effect of microstructure noise and reaching a good approximation to integrated volatility (see Andersen, 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). 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 5 for the volatility measures computed using 10 seconds returns and in Table 6 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) and a different conditioning variable, namely realized semivariance ($RS^{-}_{t,M} = \sum_{i=0}^{M-1} \left(X_{t+(i+1)/M} - X_{t+i/M}\right)^2 1_{\left\{X_{t+(i+1)/M} - X_{t+i/M} \leq 0\right\}}$), 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. Under mild regularity conditions, $RS^{-}_{t,M} = 1/2RV_{t,M} + o_p(M^{-1/2})$, hence it behaves as realized volatility.

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 0.763 (for $\widehat{RV}$ and RK at 10 seconds), or 0.702 (using a benchmark volatility estimator for the out-of-sample check).

Appendix

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

Proof of Theorem 1:

Part *(i)*. From Remark 6 in Hall, Wolff and Yao (1999), it follows that:

$$\widehat{F}_T(u|RM_{T,M}) - F(u|RM_{T,M}) = \frac{1}{\sqrt{T\xi}}\sqrt{V(u)}Z + \xi^2\frac{1}{2}C_2(K)b(u),$$

where $V(u)$ and $b(u)$ are defined as in the statement of the theorem and we (henceforth) use the notation $\widehat{F}_T(u|RM_{T,M})$ to denote the infeasible estimator, based on the latent integrated volatility. We therefore need to show that $\sqrt{T\xi}\left(\widehat{F}_{T,M}(u|RM_{T,M}) - \widehat{F}_T(u|RM_{T,M})\right) = o_p(1)$.

$$\begin{aligned}
&\sqrt{T\xi}\left(\widehat{F}_{T,M}(u|RM_{T,M}) - \widehat{F}_T(u|RM_{T,M})\right) \\
&= \frac{\frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\left(1_{\{RM_{t+1,M}\leq u\}}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - 1_{\{IV_{t+1}\leq u\}}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right)}{\widehat{f}_T(RM_{T,M})} \\
&= \frac{\frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\left(1_{\{RM_{t+1,M}\leq u\}}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - 1_{\{IV_{t+1}\leq u\}}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right)}{f(RM_{T,M})}\left(1+o_p(1)\right),
\end{aligned}$$

given that, by Theorem 2.22 in Fan and Yao (2005), $\widehat{f}_T(RM_{T,M}) = f(RM_{T,M}) + o_p(1)$. Because, by A3, $f(RM_{t,M}) > 0$, it suffices to show that:

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

We can expand the left hand side above:

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

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

$$-\frac{1}{\sqrt{T\xi}}\sum_{t=0}^{T-1}\Bigg(\left(1_{\{IV_{t+1}\leq u\}} - 1_{\{RM_{t+1,M}\leq u\}}\right) \times\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi}\right) - K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right)\Bigg). \tag{A.4}$$

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

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

Let $R_{t,\xi} = \xi^{-2} 1_{\{IV_{t+1} \leq u\}} K^{(1)}\left((IV_t - RM_{T,M})/\xi\right)$, $\overline{R}_{t,\xi} = R_{t,\xi} - \mathrm{E}\left(R_{t,\xi}\right)$, and $\overline{N}_{t,M} = N_{t,M} - \mathrm{E}\left(N_{t,M}\right)$. Then:

$$\left|\frac{1}{T}\sum_{t=0}^{T-1} R_{t,\xi} N_{t,M}\right| \leq \underbrace{\left|\frac{1}{T}\sum_{t=0}^{T-1} \overline{R}_{t,\xi}\overline{N}_{t,M}\right|}_{A_{T,M,\xi}} + \underbrace{\left|\frac{1}{T}\sum_{t=0}^{T-1} \overline{R}_{t,\xi}\mathrm{E}\left(N_{t,M}\right)\right|}_{B_{T,M,\xi}} + \underbrace{\left|\frac{1}{T}\sum_{t=0}^{T-1} \mathrm{E}\left(R_{t,\xi}\right)\overline{N}_{t,M}\right|}_{C_{T,M,\xi}} + \underbrace{\mathrm{E}\left(R_{t,\xi}\right)\mathrm{E}\left(N_{t,M}\right)}_{D_{T,M,\xi}}.$$

Given Assumption A4, and given that $\mathrm{E}\left(R_{t,\xi}\right) = O(1)$, $D_{T,M,\xi} = O\left(b_M^{-1/2}\right)$, and $C_{T,M,\xi} = O_p\left(T^{-1/2} b_M^{-1/2}\right)$. By the central limit theorem $\frac{1}{T}\sum_{t=0}^{T-1} \overline{R}_{t,\xi} = O_p\left(T^{-1/2}\xi^{-3/2}\right)$, implying that $B_{T,M,\xi} = O_p\left(b_M^{-1/2} T^{-1/2}\xi^{-3/2}\right)$. Finally:

$$A_{T,M,\xi} = \left|\frac{1}{T}\sum_{t=0}^{T-1}\left(\overline{R}_{t,\xi}\overline{N}_{t,M} - \mathrm{E}\left(\overline{R}_{t,\xi}\overline{N}_{t,M}\right)\right)\right| + \left|\mathrm{E}\left(\overline{R}_{t,\xi}\overline{N}_{t,M}\right)\right|.$$

To simplify notation, let $y_0 = IV_t$, $x = RM_{T,M}$, $z = IV_{t+1}$, $v = b_M^{-1/2} N_{t,M}$, $y = (y_0 - x)/\xi$. Given Assumptions A2 and A4, by a change of variable, Fubini theorem and integration by parts, we have that:

$$\begin{aligned}
& \mathrm{E}\left(\overline{R}_{t,\xi}\overline{N}_{t,M}\right) \\
= \;& \mathrm{E}\left(R_{t,\xi} N_{t,M}\right) + O(b_M^{-1/2}) \\
= \;& \frac{1}{b_M^{1/2}}\int_V \int_Z \int_{Y_0} \frac{1}{\xi^2} 1_{\{z \leq u\}} K^{(1)}\left(\frac{y_0 - x}{\xi}\right) v f(y_0, v, z) \mathrm{d}y_0 \mathrm{d}z \mathrm{d}v + O(b_M^{-1/2}) \\
= \;& \frac{1}{b_M^{1/2}}\int_V \int_Z \left(\int_Y \frac{1}{\xi} 1_{\{z \leq u\}} K^{(1)}(y)\, v f(x + y\xi, v, z) \mathrm{d}y\right) \mathrm{d}z \mathrm{d}v + O(b_M^{-1/2}) \\
= \;& \frac{1}{b_M^{1/2}}\int_V \int_Z \left(\left(\left[\frac{K(y)}{\xi} 1_{\{z \leq u\}} v f(x + y\xi, v, z)\right]_Y - \int_Y K(y) 1_{\{z \leq u\}} v f_y^{(1)}(x + y\xi, z, v)\right) \mathrm{d}y\right) \mathrm{d}z \mathrm{d}v + O(b_M^{-1/2}) \\
= \;& \frac{1}{b_M^{1/2}}\int_V \int_Z \left(-\int_Y K(y) 1_{\{z \leq u\}} v f_y^{(1)}(x + y\xi, z, v)\, \mathrm{d}y\right) \mathrm{d}z \mathrm{d}v + O(b_M^{-1/2}) \\
= \;& \frac{1}{b_M^{1/2}}\int_V \int_Z -1_{\{z \leq u\}} v f_y^{(1)}(x, z, v)\, \mathrm{d}z \mathrm{d}v\,(1 + O(\xi)) + O(b_M^{-1/2}) = O(b_M^{-1/2}).
\end{aligned}$$

Finally, given Assumptions A1 and A4, $\mathrm{var}\left(\frac{1}{T}\sum_{t=0}^{T-1}\left(\overline{R}_{t,\xi}\overline{N}_{t,M} - \mathrm{E}\left(\overline{R}_{t,\xi}\overline{N}_{t,M}\right)\right)\right) = b_M^{-1} T^{-1} \xi^{-3}$, and the term in (A.2) is $O_p\left(T^{1/2}\xi^{1/2} b_M^{-1/2} + \xi^{-1} b_M^{-1/2}\right)$.

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

$$\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{u - \sup_{t \leq T}|N_{t+1,M}| \leq IV_{t+1} \leq u + \sup_{t \leq T}|N_{t+1,M}|\}}\right) K\left(\frac{IV_t - RM_{T,M}}{\xi}\right). \tag{A.5}$$

Let $\overline{\Omega}_{T,M}$ be the complement of $\Omega_{T,M} = \{\omega : T^{\frac{-3}{2k}} b_M^{1/2} \sup_t |N_{t,M}| \leq c\}$, and note that, given assumption A4, for a positive c, we have:

$$\begin{aligned}
\sqrt{T\xi}\Pr\left(\overline{\Omega}_{T,M}\right) &= \sqrt{T\xi}\Pr\left(T^{\frac{-3}{2k}} b_M^{1/2} \sup_t |N_{t,M}| > c\right) \leq \sqrt{T\xi}\sum_{t=0}^{T-1}\Pr\left(T^{-\frac{3}{2k}} b_M^{1/2} |N_{t,M}| > c\right) \\
&\leq T^{3/2}\sqrt{\xi} T^{-\frac{3k}{2k}} c^{-k} b_M^{k/2} \mathrm{E}\left(|N_{t,M}|^k\right) = o(1).
\end{aligned}$$

Consequently, we can proceed by conditioning on $\Omega_{T,M}$. For all $\omega \in \Omega_{T,M}$:

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

$$\leq \frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{u-cb_M^{-1/2}T^{3/2k}\leq IV_{t+1}\leq u+cb_M^{-1/2}T^{3/2k}\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right). \tag{A.6}$$

To simplify notation, let $d_{T,M}=cb_M^{-1/2}T^{\frac{3}{2k}}$. Then, using (A.5) and (A.6), for all $\omega\in\Omega_{T,M}$:

$$\begin{aligned}
&\left|\frac{1}{T\xi}\sum_{t=0}^{T-1}\left(1_{\{IV_{t+1}\leq u\}}-1_{\{RM_{t+1,M}\leq u\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right| \\
&\leq \frac{1}{T\xi}\sum_{t=0}^{T-1}1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right) \\
&= \frac{1}{T}\sum_{t=0}^{T-1}\mathrm{E}\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\frac{1}{\xi}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right) \qquad \text{(A.7)} \\
&+ \frac{1}{T}\sum_{t=0}^{T-1}\left(\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\right)\frac{1}{\xi}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right. \qquad \text{(A.8)} \\
&\qquad\left.-\mathrm{E}\left(\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\frac{1}{\xi}K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right)\right)\right).
\end{aligned}$$

We start from (A.7). Given stationarity, we have that:

$$\begin{aligned}
&\frac{1}{\xi}\mathrm{E}\left(\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\right)K\left(\frac{IV_t-RM_{T,M}}{\xi}\right)\right) \qquad \text{(A.9)} \\
&= \frac{1}{\xi}\int_{Y_0}\int_{u-d_{T,M}}^{u+d_{T,M}}K\left(\frac{y_0-x}{\xi}\right)f(y_0,z)\,\mathrm{d}y_0\mathrm{d}z = \int_Y\int_{u-d_{T,M}}^{u+d_{T,M}}K(y)\,f(x+\xi y,z)\,\mathrm{d}y\mathrm{d}z \\
&= \int_Y K(y)\,\mathrm{d}y\int_{u-d_{T,M}}^{u+d_{T,M}}f(x,z)\;\mathrm{d}z(1+O(\xi)) = O(d_{T,M})(1+\xi).
\end{aligned}$$

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

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

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

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

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

Part *(ii)*. Because $b_M^{-1}\xi^{-2}=(Tb_M^{-1}\xi^{-1})/(T\xi^3)$, if $T\xi^3\to\infty$, then $b_M^{-1}\xi^{-2}\to 0$ for $Tb_M^{-1}\xi^{-1}\to 0$, but the latter condition is implied by $T^{\frac{3+k}{k}}\xi b_M^{-1}$. The statement then follows from the proof of part (i). ■

Proof of Theorem 2:

Part *(i)*. From, e.g., Hansen (2004):

$$\widehat{f}_T(x|RM_{T,M})-f(x|RM_{T,M}) \quad = \quad \frac{1}{\sqrt{T\xi_1\xi_2}}\left(\sqrt{\frac{f(x|RM_{T,M})}{f(RM_{T,M})}}\int K^2(u)\mathrm{d}u\right)Z+\frac{1}{2}\xi_1^2C_2(K)\frac{\partial^2 f(x|RM_{T,M})}{\partial x^2}$$

$$+\frac{1}{2}\xi_2^2 C_2(K)\left(\left.\frac{\partial^2 f(x|z)}{\partial z^2}\right|_{z=RM_{T,M}}+2\frac{\left.\frac{\partial f(x|z)}{\partial z}\right|_{z=RM_{T,M}}\left.\frac{\partial f(z)}{\partial z}\right|_{z=RM_{T,M}}}{f(RM_{T,M})}\right).$$

To prove the theorem, we need to show that $\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{T,M}(x|RM_{T,M})-\widehat{f}_T(x|RM_{T,M})\right)=o_p(1)$.

$$\begin{aligned}&\sqrt{T\xi_1\xi_2}\left(\widehat{f}_{T,M}(x|RM_{T,M})-\widehat{f}_T(x|RM_{T,M})\right) \qquad \text{(A.10)}\\ &=\frac{\sqrt{T\xi_1\xi_2}\left(\frac{1}{T\xi_1\xi_2}\sum_{t=0}^{T-1}\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_1}\right)K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)-K\left(\frac{IV_t-RM_{T,M}}{\xi_1}\right)K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\right)\right)}{\widehat{f}_T(x|RM_{T,M})}\\ &+\left(\frac{1}{\widehat{f}_{T,M}(RM_{T,M})}-\frac{1}{\widehat{f}_T(RM_{T,M})}\right)\\ &\times\left(\frac{1}{\sqrt{T\xi_1\xi_2}}\sum_{t=0}^{T-1}\left(K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_1}\right)K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)-K\left(\frac{IV_t-RM_{T,M}}{\xi_1}\right)K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\right)\right).\end{aligned}$$

The second term on the right hand side (rhs) of (A.10) is of smaller probability order than the first, which can be treated similarly to the term in (A.2) in the proof of Theorem 1, and therefore is $O_p\left(Tb_M^{-1}\xi_1\xi_2+b_M\xi_1^{-2}+b_M\xi_2^{-2}\right)$. The statement in the theorem follows.
Part *(ii)*. Similar to part *(ii)* in the proof of Theorem 1. ■

Proof of Theorem 3:

Part *(i)*. From Remark 4, in Hall, Wolff and Yao (1999):

$$\widehat{\alpha}_{0,T,M}(u,RM_{T,M})-F(u|RM_{T,M})=\frac{1}{\sqrt{T\xi}}\sqrt{V(u)}Z+\xi^2\frac{1}{2}C_2(K)\left.\frac{\partial^2 F(u|z)}{\partial z^2}\right|_{z=RM_{T,M}}.$$

The theorem is proved, given that $\sqrt{T\xi}\left(\widehat{\alpha}_{0,T}(u,IV_T)-\widehat{\alpha}_{0,T,M}(u,RM_{T,M})\right)=o_p(1)$ follows straightforwardly using the same argument as in the proof of Theorem 1.
Part *(ii)*. From Fan, Yao and Tong (1996, p.196):

$$\begin{aligned}&\widehat{\beta}_{0,T}(x,RM_{T,M})-f(x|RM_{T,M})\\ =\;&\frac{1}{\sqrt{T\xi_1\xi_2}}\left(\sqrt{\frac{f(x|RM_{T,M})}{f(RM_{T,M})}}\int K^2(u)\mathrm{d}u\right)Z+\frac{1}{2}\xi_1^2C_2(K)\frac{\partial^2 f(x|RM_{T,M})}{\partial x^2}+\frac{1}{2}\xi_2^2C_2(K)\left.\frac{\partial^2 f(x|z)}{\partial z^2}\right|_{z=RM_{T,M}},\end{aligned}$$

and the theorem is proved, because $\sqrt{T\xi_1\xi_2}\left(\widehat{\beta}_{0,T,M}(x,RM_{T,M})-\widehat{\beta}_{0,T}(x,IV_T)\right)=o_p(1)$, by the same argument as that used in Theorem 1. ■

Proof of Lemma 1:

Before moving to the proof of (i)-(iii), it is useful to show that:

$$\mathrm{E}\left(\sum_{j=1}^{M-1}\left(Y_{t+(j+1)/M}-Y_{t+j/M}\right)^2-IV_t\right)^k=O\left(M^{-k/2}\right). \qquad \text{(A.11)}$$

We start with the case of zero drift. From Proposition 2.1 in Meddahi (2002):

$$N_{t,M}=2\sqrt{M}\underbrace{\sum_{i=0}^{M-1}\left(\sigma_{t+i/M}^2\int_{t+i/M}^{t+(i+1)/M}\left(\int_{t+i/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)}_{N_{t,M}^{(1)}}$$

$$+ 2\sqrt{M}\underbrace{\sum_{i=0}^{M-1}\left(\sigma_{t+i/M}\int_{t+i/M}^{t+(i+1)/M}\left(\int_{t+i/M}^{s}\left(\sigma_u-\sigma_{t+i/M}\right)\mathrm{d}W_u\right)\mathrm{d}W_s\right)}_{N_{t,M}^{(2)}}$$

$$+ 2\sqrt{M}\underbrace{\sum_{i=0}^{M-1}\left(\int_{t+i/M}^{t+(i+1)/M}\left(\sigma_u-\sigma_{t+i/M}\right)\left(\int_{t+i/M}^{s}\mathrm{d}W_u\right)\sigma_{t+i/M}\mathrm{d}W_s\right)}_{N_{t,M}^{(3)}}$$

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

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

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

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

$$\begin{aligned}&\mathrm{E}\left(\left(\sqrt{M}N_{t,M}^{(1)}\right)^4\right)\\ &= M^2\sum_{j_1}\sum_{j_2}\sum_{j_3}\mathrm{E}\Bigg[\sigma_{t+j_1/M}^2\sigma_{t+j_2/M}^2\sigma_{t+j_3/M}^4\\ &\quad\times\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\\ &\quad\times\mathrm{E}_{t+j_3/M}\left(\left(\int_{t+j_3/M}^{t+(j_3+1)/M}\left(\int_{t+j_3/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^2\right)\Bigg]\\ &\simeq \sum_{j_1}\sum_{j_2}\mathrm{E}\Bigg[\sigma_{t+j_1/M}^2\sigma_{t+j_2/M}^2\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\\ &\quad\times\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\sum_{j_3}\mathrm{E}_{t+\max\{j_1,j_2\}/M}\left(\left|\sigma_{t+j_3/M}^4-\mathrm{E}\left(\sigma_{t+j_3/M}^4\right)\right|\right)\Bigg]\end{aligned}$$

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

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

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

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

The fourth moment in the case above is of larger order than in the case $j_1=j_2=j_3=j_4$. Finally, for $j_2=j_3=j_4$, it follows trivially that $\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t,M}\right)^4\right)=0$.

We now analyze the case with drift. From Proposition 2.1 in Meddahi (2002), the contribution of the drift term to the measurement error 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 are of a smaller 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 $\mathrm{E}\left((\mu_t)^{2(k+\delta)}\right)<\infty$.

Part *(i)*. Note that:

$$\mathrm{E}\left(\left(\widehat{RV}_{t,l,M}-IV_t\right)^k\right)$$
$$\simeq \mathrm{E}\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(Y_{t+(jB+b)/M}-Y_{t+((j-1)B+b)/M}\right)^2-IV_t\right)^k\right) \tag{A.12}$$
$$+\,\mathrm{E}\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(Y_{t+(jB+b)/M}-Y_{t+((j-1)B+b)/M}\right)\left(\epsilon_{t+(jB+b)/M}-\epsilon_{t+((j-1)B+b)/M}\right)\right)^k\right) \tag{A.13}$$
$$+\,\mathrm{E}\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(\epsilon_{t+(jB+b)/M}-\epsilon_{t+((j-1)B+b)/M}\right)^2-\frac{l}{M}\sum_{j=1}^{M}\left(\epsilon_{t+j/M}-\epsilon_{t+(j-1)/M}\right)^2\right)^k\right). \tag{A.14}$$

Because $\mathrm{E}(\epsilon^{2k}_{t+j/M})<\infty$, the term in (A.14) is $O(l^{k/2}/B^{k/2})=O(b_M^{-k/2})$ for $b_M=M^{1/3}$. Given the independence between the noise and the price, (A.13) is of order $O(B^{-k/2})$. We are left with (A.12):

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

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

$$+\,\mathrm{E}\left(\left(\sum_{j=1}^{M}\left(Y_{t+j/M}-Y_{t+(j-1)/M}\right)^2-IV_t\right)^k\right). \quad \text{(A.16)}$$

From (A.11), it follows that the term in (A.16) is $O(M^{-k/2})$. With regard to (A.15), from the proof of Theorem 2 in Zhang, Mykland and Aït-Sahalia (2005):

$$\begin{aligned}&\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(Y_{t+(jB+b)/M}-Y_{t+((j-1)B+b)/M}\right)^2-\sum_{j=1}^{M}\left(Y_{t+j/M}-Y_{t+(j-1)/M}\right)^2\\&=2\sum_{j=1}^{M-1}\left(Y_{t+(j+1)/M}-Y_{t+j/M}\right)\sum_{i=1}^{B\wedge j}\left(1-\frac{j}{B}\right)\left(Y_{t+(j-i+1)/M}-Y_{t+(j-i)/M}\right)+O(B/M)\end{aligned}$$

and:

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

by the same argument as that used to show (A.11). The statement then follows.
Part *(ii)*. The case of $k=2$ is analyzed in Aït-Sahalia, Mykland and Zhang (2006). Exploiting $\sum_{i=1}^{\tau}\frac{a_i}{i}=0$ and $\sum_{i=1}^{\tau}a_i=1$, we have that:

$$\mathrm{E}\left(\left(\widetilde{RV}_{t,\tau,M}-IV_t\right)^k\right) \quad \text{(A.17)}$$

$$\begin{aligned}&=\mathrm{E}\left[\left(\sum_{i=1}^{\tau}a_i\left(\frac{1}{i}\sum_{j=1}^{M-i}\left(Y_{t+(j+i)/M}-Y_{t+j/M}\right)^2-IV_t\right)-2\sum_{i=1}^{\tau}\frac{a_i}{i}\sum_{j=1}^{M-i}\epsilon_{t+(j+i)/M}\epsilon_{t+j/M}\right.\right.\\&\quad+2\sum_{i=1}^{\tau}a_i\left(\sum_{j=1}^{M-i}\frac{1}{i}\left(Y_{t+(j+i)/M}-Y_{t+j/M}\right)\left(\epsilon_{t+(j+i)/M}-\epsilon_{t+j/M}\right)\right)-\left(\sum_{i=1}^{\tau}\frac{a_i}{i}\sum_{j=M-i}^{M}\left(\epsilon^2_{t+j/M}-\sigma^2_\epsilon\right)\right)\\&\quad\left.\left.+2\left(\widehat{\sigma}^2_\epsilon-\sigma^2_\epsilon\right)\right)^k\right].\end{aligned}$$

It's enough to consider the k-th powers of the single elements of the rhs of (A.17) since, by Hölder inequality, the cross terms are of a smaller order. $\mathrm{E}\left(\left(\widehat{\sigma}^2_\epsilon-\sigma^2_\epsilon\right)^k\right)=O(M^{-k/2})$. Furthermore, because $a_i\simeq i^2/\tau^3$:

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

so that the k-th moments of the fourth and fifth terms of the rhs of (A.17) are $O(\tau^{-k/2})$. Because of (A.11), the $k-$th moments of the first term in (A.17) is $O\left((\tau/M)^{k/2}\right)=O(\tau^{-k/2})$. Given $\mathrm{E}\left(\epsilon_t Y_t\right)=0$, the third term of the rhs of (A.17) is also $O(\tau^{-k/2})$ because of (A.11). Finally, the $k-$th moment of the second term of the rhs of (A.17) can be treated as the second term of the rhs of the last equality in (A.18) (in part *(iii)* of the Lemma below), and is therefore $O(\tau^{-k/2})$. Because $\tau=O(M^{1/2})$, the statement follows for $b_M=M^{1/2}$.

Part *(iii)*. The case of $k = 2$ has been established in Theorem 2 by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). We begin with the case of ϵ_t independently distributed. Note that:

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

where $\gamma^{Y,\epsilon}_{t,h} = \sum_{j=1}^{M}\left(Y_{t+j/M} - Y_{t+(j-1)/M}\right)\left(\epsilon_{t+(j-h)/M} - \epsilon_{t+(j-1-h)/M}\right)$ and the other terms are defined in a similar fashion. We only need to show that $\mathrm{E}\left(\left(\gamma^{\epsilon}_{t,0} + \sum_{h=1}^{H}\kappa\left(\frac{h-1}{H}\right)\left(\gamma^{\epsilon}_{t,h} + \gamma^{\epsilon}_{t,-h}\right)\right)^k\right) = O\left(H^{-\frac{3k}{2}}M^{k/2}\right) = O(M^{-1/2})$ for $H = M^{1/2}$, since the other terms can be treated as in part *(ii)*.

Let $\boldsymbol{\gamma}^{\epsilon} = \left(\gamma^{\epsilon}_{t,0}, \gamma^{\epsilon}_{t,1} + \gamma^{\epsilon}_{t,-1}, \ldots, \gamma^{\epsilon}_{t,h} + \gamma^{\epsilon}_{t,-h}\right)'$, where for notational simplicity we have dropped the subscript t. Following the proof of Theorem 1 in Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), we have that $\boldsymbol{\gamma}^{\epsilon} = \boldsymbol{\gamma}^{\epsilon}_V + \boldsymbol{\gamma}^{\epsilon}_W + \boldsymbol{\gamma}^{\epsilon}_Z$, where:

$$\begin{aligned}\boldsymbol{\gamma}^{\epsilon}_V &= 2\left(V_0 - V_1, -V_0 + 2V_1 - V_2, \ldots, -V_{H-1} + 2V_H - V_{H+1}\right)',\\ V_h &= \sum_{j=1}^{M-h-1}\epsilon_{j/M}\epsilon_{(j+h)/M},\\ \boldsymbol{\gamma}^{\epsilon}_W &= \left(0, -W_2, \ldots, -W_{H-1} + 2W_H - W_{H+1}\right)',\\ W_h &= \sum_{j=1}^{h-1}\epsilon_{j/M}\epsilon_{(j-h)/M} + \sum_{j=M-h+1}^{M}\epsilon_{j/M}\epsilon_{(j+h)/M},\\ \boldsymbol{\gamma}^{\epsilon}_Z &= \left(Z_0 - 2Z_1, Z_{-1} - Z_0 + 3Z_1 - 2Z_2, \ldots, Z_{-H} - Z_{-H+1} - Z_{H-1} + 3Z_H - 2Z_{H+1}\right)',\\ Z_h &= \epsilon_{1/M}\epsilon_{h/M} + \epsilon_1\epsilon_{(M-h)/M}.\end{aligned}$$

By Hölder inequality, all cross moments are of a smaller order. Hence, we only need to prove that, for $\mathbf{w} = \left(1, 1, \kappa\left(\frac{1}{H}\right), \ldots, \kappa\left(\frac{H-1}{H}\right)\right)'$, $\mathrm{E}\left(\mathbf{w}'\boldsymbol{\gamma}^{\epsilon}_V\right)^k$, $\mathrm{E}\left(\mathbf{w}'\boldsymbol{\gamma}^{\epsilon}_W\right)^k$ and $\mathrm{E}\left(\mathbf{w}'\boldsymbol{\gamma}^{\epsilon}_Z\right)^k$ are $O\left(H^{-k/2}\right)$. We begin with $\mathbf{w}'\boldsymbol{\gamma}^{\epsilon}_V$ and, after some algebra, get:

$$\begin{aligned}\frac{1}{2}\mathbf{w}'\boldsymbol{\gamma}^{\epsilon}_V &= \left(V_0 - V_1\right) + \sum_{h=1}^{H}\kappa\left(\frac{h-1}{H}\right)\left(-V_{h-1} + 2V_h - V_{h+1}\right)\\ &= \left(1 - \kappa\left(\frac{1}{H}\right)\right)V_1 + \sum_{h=1}^{H-2}\left(\kappa\left(\frac{h-1}{H}\right) - 2\kappa\left(\frac{h}{H}\right) + \kappa\left(\frac{h+1}{H}\right)\right)V_{h+1}\\ &+ \left(2\kappa\left(\frac{H-1}{H}\right) - \kappa\left(\frac{H-2}{H}\right)\right)V_H - \kappa\left(\frac{H-1}{H}\right)V_{H+1}. \qquad \text{(A.18)}\end{aligned}$$

Because ϵ is *i.i.d.*, $\mathrm{E}\left(V^k_h\right) = O(M^{k/2})$, and $\mathrm{E}\left(\sum_{h=1}^{H}V_h\right)^k = O\left(H^{k/2}M^{k/2}\right)$. Therefore:

$$\begin{aligned}&\frac{1}{2^k}\mathrm{E}\left(\left(\mathbf{w}'\boldsymbol{\gamma}^{\epsilon}_V\right)^k\right) \qquad \text{(A.19)}\\ &\simeq \left(1 - \kappa\left(\frac{1}{H}\right)\right)^k\mathrm{E}\left(V^k_1\right) \qquad \text{(A.20)}\\ &+ \sum_{h_1=1}^{H}\sum_{h_2=1}^{H}\cdots\sum_{h_{k/2}=1}^{H}\left(\left(\kappa\left(\frac{h_1-1}{H}\right) - 2\kappa\left(\frac{h_1}{H}\right) + \kappa\left(\frac{h_1+1}{H}\right)\right)^2\cdots \qquad \text{(A.21)}\end{aligned}$$

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

$$+\left(2\kappa\left(\frac{H-1}{H}\right)-\kappa\left(\frac{H-2}{H}\right)\right)^k\mathrm{E}\left(V_H^k\right) \tag{A.22}$$

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

Exploiting the properties of the kernel κ, and applying standard Taylor expansions, it follows that (A.20), (A.22) and (A.23) are $O\left(M^{k/2}H^{-2k}\right)=O(H^{-1})$, while (A.21) is $O\left(H^{-\frac{3k}{2}}M^{k/2}\right)$. Hence, (A.19) is $O\left(H^{-\frac{3k}{2}}M^{k/2}\right)$. By a similar argument, $\mathrm{E}\left((\mathbf{w}'\boldsymbol{\gamma}_W^\epsilon)^k\right)$ and $\mathrm{E}\left((\mathbf{w}'\boldsymbol{\gamma}_Z^\epsilon)^k\right)$ are $O\left(H^{-k/2}\right)$. Then, for $H=M^{1/2}$ and $b_M=M^{1/2}$ the statement follows.

We now turn to the case of dependent noise. Contrary to the *i.i.d.* case, $\mathrm{E}(V_h)\neq 0$ and $\mathrm{E}(V_hV_{h'})\neq 0$ for $h\neq h'$. Without loss of generality, let $\epsilon_{j/M}=\rho\epsilon_{(j-1)/M}+u_{j/M}$, with $u_{j/M}\sim$ *i.i.d.*$(0,\sigma_u^2)$, where, because of stationarity, we have suppressed the subscript t. After a Taylor expansion of the first two terms of the rhs of (A.18) around zero and of the last two terms around one, we obtain:

$$\frac{1}{2}\mathbf{w}'\boldsymbol{\gamma}_V^\epsilon=\frac{\kappa^{(3)}(0)}{H^3}V_1+\frac{\kappa^{(3)}(0)}{H^3}\sum_{h=1}^{H-2}hV_{h+1}+\frac{2\kappa^{(2)}(1)}{H^2}V_H+\frac{\kappa^{(2)}(1)}{2H^2}V_{H+1}. \tag{A.24}$$

We only consider the second term of the rhs of (A.24), given that the others are of a smaller order. We have that:

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

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

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

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

$$\mathrm{E}\left(\left(\frac{\kappa^{(3)}(0)}{H^3}\sum_{h=1}^{H}hV_h\right)^4\right)$$
$$=\frac{\left(\kappa^{(3)}(0)\right)^4}{H^{12}}\sum_{h_1=1}^{H}h_1\sum_{h_2>h_1}^{H}h_2\sum_{h_3>h_2}^{H}h_3\sum_{h_4>h_3}^{H}h_4$$
$$\times\sum_{j_1=1}^{M}\sum_{j_2=1}^{M}\sum_{j_3=1}^{M}\sum_{j_4=1}^{M}\Big(\mathrm{E}\left(\epsilon_{j_1/M}\mathrm{E}_{j_1/M}\left(\epsilon_{(j_1+h_1)/M}\right)\mathrm{E}_{(j_1+h_1)/M}\left(\epsilon_{j_2/M}\right)\mathrm{E}_{j_2/M}\left(\epsilon_{(j_2+h_2)/M}\right)\right.$$
$$\left.\mathrm{E}_{(j_2+h_2)/M}\left(\epsilon_{j_3/M}\right)\mathrm{E}_{j_3/M}\left(\epsilon_{(j_3+h_3)/M}\right)\mathrm{E}_{(j_3+h_3)/M}\left(\epsilon_{j_4/M}\right)\mathrm{E}_{j_4/M}\left(\epsilon_{(j_4+h_4)/M}\right)\right)\Big). \tag{A.25}$$

The conditional expectations in (A.25) are easily computed. For example:

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

The other conditional expectations can be calculated similarly. Plugging the expressions back in (A.25), after some algebra, we have:

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

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. Forthcoming, *Journal of Econometrics*.

Aït-Sahalia, Y., J. Fan and H. Peng (2009). Nonparametric Transition-Based Tests for Jump-Diffusions. *Journal of the American Statistical Association*, 104, 1102-1116.

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 & Economic Statistics,* 27, 251-265.

Bandi, F.M., V. Corradi and G. Molche (2009). Bandwidth Selection for Continuous Time Markov Processes. John Hopkins University, Working Paper.

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 Y. Kabanov and R. Lipster (eds.), *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*. Springer and Verlag, 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.

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 (2009). Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures. *Journal of Econometrics*, 150, 119-138.

Delaigle, A., J. Fan and R. Carroll (2009). Design-Adaptive Local Polynomial Estimator for the Error-in-Variables Problem. *Journal of the American Statistical Association,* 104, 348-359.

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. (1991). On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *Annals of Statistics*, 19, 1257-1272.

Fan, J. (1992). Deconvolution with Supersmooth Distributions, *Canadian Journal of Statistics,* 20, 155-169.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Fan, J., J.C. Jiang, C.M. Zhang and Z. Zhou (2003). Time-Dependent Diffusion Models for Term Structure Dynamics. *Statistica Sinica*, 13, 965-992.

Fan, J. and E. Masry (1992). Multivariate Regression Estimation with Errors-in-Variables: Asymptotic Normality for Mixing Processes. *Journal of Multivariate Analysis*, 43, 237-271.

Fan, J. and Y.K. Truong (1993). Nonparametric Regression with Errors in Variables. *Annals of Statistics*, 21, 1900-1925

Fan, J. and Y. Whang (2007). Multi-Scale Jumo and Volatility Analysis for High-Frequency Financial Data. *Journal of the American Statistical Association,* 102, 1349-1362.

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.

Florens-Zmirou, D. (1989). Estimation of the Coefficient of Diffusion from Discretized Observations. *Comptes Rendus de l'Academie de Science-Series I Mathematique*, 309, 195-200.

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.

Hansen, B.E. (2004). Nonparametric Conditional Density Estimation. University of Wisconsin, Working Paper

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.

Kinnebrock, S. and M. Podolskij (2008). A note on the central limit theorem for bipower variation of general functions. *Stochastic Processes and Their Applications*, 118, 1056-1070.

Koo, B. and O. Linton (2009). Semiparametric Estimation of Locally Stationary Diffusion Models. London School of Economics, Working Paper.

Li, T. and Q. Vuong (1998). Nonparametric Estimation of the Measurement Error Model using Multiple Indicators. *Journal of Multivariate Analysis*, 65, 139-165.

Masry, E. (1991). Multivariate Probability Density Deconvolution for Stationary Random Processes. *IEEE Transaction on Information Theory*, 37, 1105-1115.

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

Masuda, H. (2007). Ergodicity and Exponential $\beta-$Mixing Bounds for Multivariate Diffusions with Jumps. *Stochastic Processes and Their Applications*, 117, 35-56.

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

Meddahi, N. (2003). ARMA Representation of Integrated and Realized Variances. *Econometrics Journal*, 6, 334-355.

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.

Podolskij, M. and M. Vetter (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and Their Applications*, 119, 2803-2831.

Schennach, S.M. (2004). Nonparametric Regression in the Presence of Measurement Error. *Econometric Theory,* 20, 1046–1093.

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

Stefanski, L.A. (1990). Rates of Convergence of Some Estimators in a Class od Deconvolution Problems. *Statistics and Probability Letters*, 9, 229-235.

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

Xiu, D. (2010). Quasi-Maximum Likelihood Estimation of Volatility with High Frequency Data. *Journal of Econometrics*, 159, 235-250.

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. (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	$RV_{t,M}$	$BV_{t,M}$	$TPV_{t,M}$	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,\tau,M}$	$RK_{t,H,M}$
			$Interval = \widehat{\mu}_{IV} + 0.125\widehat{\sigma}_{IV}$			
Nominal Size = 5%						
72	0.095	0.105	0.122	0.455	0.203	0.200
144	0.090	0.089	0.094	0.257	0.142	0.126
288	0.082	0.084	0.082	0.156	0.106	0.098
576	0.078	0.083	0.080	0.103	0.088	0.089
Nominal Size = 10%						
72	0.141	0.151	0.168	0.531	0.255	0.260
144	0.137	0.138	0.142	0.321	0.191	0.172
288	0.130	0.129	0.128	0.213	0.154	0.144
576	0.126	0.133	0.128	0.152	0.133	0.134
			$Interval = \widehat{\mu}_{IV} + 0.250\widehat{\sigma}_{IV}$			
Nominal Size = 5%						
72	0.076	0.084	0.098	0.609	0.182	0.212
144	0.076	0.074	0.077	0.313	0.112	0.107
288	0.075	0.078	0.075	0.150	0.080	0.079
576	0.081	0.079	0.075	0.083	0.074	0.075
Nominal Size = 10%						
72	0.123	0.132	0.146	0.690	0.247	0.275
144	0.131	0.123	0.122	0.394	0.162	0.158
288	0.127	0.128	0.125	0.211	0.128	0.125
576	0.136	0.132	0.129	0.128	0.126	0.123

[a] Notes: Entries denote rejection frequencies based on comparing $G_{T,M}(u_1, u_2)$ to 5% and 10% critical values of the standard normal distribution. We use "pseudo true" IV values for $G_{T,M}(u_1, u_2)$, as discussed in Section 4. The interval $[u_1, u_2]$ is $[\widehat{\mu}_{IV} - \beta\widehat{\sigma}_{IV}, \widehat{\mu}_{IV} + \beta\widehat{\sigma}_{IV}]$, where $\widehat{\mu}_{IV}$ and $\widehat{\sigma}_{IV}$ are the mean and standard error of the pseudo true data, and $\beta = \{0.125, 0.250\}$. All experiments are based on several values of M, $T = 100$ and 10,000 Monte Carlo iterations. See Section 4 for further details.

Table 2: Conditional Confidence Interval Accuracy Assessment: Bandwidth Sensitivity

Bandwidth[a]	$RV_{t,M}$	$BV_{t,M}$	$TPV_{t,M}$	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,\tau,M}$	$RK_{t,H,M}$
			No Noise			
Nominal Size = 5%						
1/3	0.098	0.102	0.105	0.160	0.126	0.117
1/2	0.085	0.091	0.090	0.152	0.117	0.107
2/3	0.080	0.082	0.082	0.149	0.112	0.099
3/4	0.080	0.082	0.081	0.150	0.111	0.099
1	0.082	0.084	0.082	0.156	0.106	0.098
3/2	0.091	0.089	0.090	0.167	0.119	0.104
Nominal Size = 10%						
1/3	0.138	0.139	0.140	0.203	0.170	0.156
1/2	0.127	0.131	0.130	0.198	0.161	0.146
2/3	0.124	0.126	0.129	0.199	0.159	0.147
3/4	0.125	0.125	0.129	0.199	0.159	0.144
1	0.130	0.129	0.128	0.213	0.154	0.144
3/2	0.141	0.138	0.139	0.219	0.169	0.156
			Noise has $\mathrm{N}(0, 0.005^2)$ *distribution*			
Nominal Size = 5%						
1/3	0.354	0.374	0.351	0.148	0.122	0.122
1/2	0.243	0.265	0.245	0.143	0.113	0.111
2/3	0.187	0.204	0.189	0.138	0.109	0.103
3/4	0.170	0.186	0.172	0.138	0.107	0.103
1	0.152	0.163	0.152	0.159	0.114	0.103
3/2	0.163	0.177	0.164	0.161	0.115	0.106
Nominal Size = 10%						
1/3	0.383	0.398	0.377	0.194	0.159	0.160
1/2	0.274	0.297	0.275	0.185	0.153	0.153
2/3	0.221	0.239	0.227	0.188	0.149	0.147
3/4	0.206	0.224	0.211	0.189	0.150	0.145
1	0.196	0.210	0.197	0.210	0.158	0.149
3/2	0.217	0.229	0.213	0.209	0.163	0.156

[a] Notes: See notes to Table 1. The bandwidth column reports the ratio between the used bandwidth and the one calculated using Silverman's (1986) rule. The interval $[u_1, u_2]$ is $[\widehat{\mu}_{IV} - .125\widehat{\sigma}_{IV}, \widehat{\mu}_{IV} + .125\widehat{\sigma}_{IV}]$. All experiments are based on $T = 100$, $M = 288$ and 10,000 Monte Carlo iterations.

Table 3: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case II: Microstructure Noise in DGP[a]

M	$RV_{t,M}$	$BV_{t,M}$	$TPV_{t,M}$	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,\tau,M}$	$RK_{t,H,M}$
Panel A: Interval $= \widehat{\mu}_{IV} + 0.125\widehat{\sigma}_{IV}$						
Noise has $\mathrm{N}(0, 0.005^2)$ *distribution*						
Nominal Size = 5%						
72	0.075	0.083	0.085	0.457	0.203	0.199
144	0.073	0.075	0.072	0.260	0.140	0.125
288	0.152	0.163	0.152	0.159	0.114	0.103
576	0.980	0.995	0.991	0.102	0.089	0.087
Nominal Size = 10%						
72	0.121	0.129	0.127	0.530	0.258	0.256
144	0.118	0.123	0.120	0.323	0.190	0.173
288	0.196	0.210	0.197	0.210	0.158	0.149
576	0.983	0.996	0.992	0.150	0.138	0.130
Noise has $\mathrm{N}(0, 0.007^2)$ *distribution*						
Nominal Size = 5%						
72	0.076	0.084	0.082	0.456	0.209	0.206
144	0.130	0.133	0.121	0.262	0.137	0.130
288	0.940	0.967	0.949	0.162	0.115	0.102
576	1.000	1.000	1.000	0.102	0.090	0.087
Nominal Size = 10%						
72	0.123	0.127	0.129	0.532	0.269	0.262
144	0.167	0.175	0.163	0.326	0.190	0.179
288	0.948	0.971	0.955	0.214	0.160	0.148
576	1.000	1.000	1.000	0.146	0.137	0.135
Noise has $\mathrm{N}(0, 0.014^2)$ *distribution*						
Nominal Size = 5%						
72	0.727	0.727	0.650	0.471	0.221	0.223
144	1.000	1.000	1.000	0.275	0.154	0.157
288	1.000	1.000	1.000	0.172	0.113	0.115
576	1.000	1.000	1.000	0.119	0.099	0.101
Nominal Size = 10%						
72	0.759	0.759	0.688	0.543	0.274	0.284
144	1.000	1.000	1.000	0.341	0.197	0.210
288	1.000	1.000	1.000	0.224	0.156	0.158
576	1.000	1.000	1.000	0.164	0.148	0.142
Panel B: Interval $= \widehat{\mu}_{IV} + 0.250\widehat{\sigma}_{IV}$						
Noise has $\mathrm{N}(0, 0.005^2)$ *distribution*						
Nominal Size = 5%						
72	0.069	0.069	0.070	0.616	0.229	0.217
144	0.076	0.078	0.079	0.316	0.129	0.110
288	0.114	0.133	0.123	0.153	0.088	0.077
576	0.970	0.988	0.981	0.085	0.075	0.073
Nominal Size = 10%						
72	0.116	0.117	0.114	0.698	0.295	0.285
144	0.133	0.133	0.133	0.393	0.185	0.160
288	0.162	0.187	0.169	0.212	0.135	0.126
576	0.977	0.990	0.985	0.137	0.123	0.120
Noise has $\mathrm{N}(0, 0.007^2)$ *distribution*						
Nominal Size = 5%						
72	0.073	0.070	0.066	0.617	0.230	0.220
144	0.100	0.105	0.089	0.316	0.131	0.114
288	0.934	0.956	0.939	0.148	0.089	0.086
576	1.000	1.000	1.000	0.087	0.074	0.076
Nominal Size = 10%						
72	0.123	0.124	0.114	0.702	0.300	0.290
144	0.142	0.152	0.134	0.392	0.185	0.167
288	0.949	0.964	0.951	0.207	0.136	0.130
576	1.000	1.000	1.000	0.132	0.121	0.121
Noise has $\mathrm{N}(0, 0.014^2)$ *distribution*						
Nominal Size = 5%						
72	0.789	0.787	0.725	0.638	0.244	0.255
144	1.000	1.000	1.000	0.336	0.135	0.146
288	1.000	1.000	1.000	0.164	0.093	0.101
576	1.000	1.000	1.000	0.093	0.072	0.075
Nominal Size = 10%						
72	0.834	0.834	0.784	0.719	0.312	0.323
144	1.000	1.000	1.000	0.415	0.194	0.199
288	1.000	1.000	1.000	0.224	0.142	0.151
576	1.000	1.000	1.000	0.142	0.125	0.127

[a] Notes: See notes to Table 1. All experiments are based on samples of 100 daily observations and 10,000 Monte Carlo iterations.

Table 4: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case III: Jumps in DGP[a]

M	$RV_{t,M}$	$BV_{t,M}$	$TPV_{t,M}$	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,\tau,M}$	$RK_{t,H,M}$
Panel A: Interval $= \widehat{\mu}_{IV} + 0.125\widehat{\sigma}_{IV}$						
One i.i.d. $\mathrm{N}(0, 3 * 0.64 * \widehat{\mu}_{IV})$ *Jump Every 5 Days*						
Nominal Size = 5%						
72	0.239	0.197	0.194	0.622	0.401	0.396
144	0.213	0.161	0.153	0.438	0.318	0.293
288	0.196	0.136	0.131	0.317	0.261	0.242
576	0.189	0.126	0.122	0.243	0.237	0.211
Nominal Size = 10%						
72	0.301	0.259	0.253	0.695	0.476	0.468
144	0.273	0.212	0.204	0.509	0.385	0.363
288	0.254	0.186	0.183	0.390	0.323	0.305
576	0.245	0.175	0.170	0.304	0.300	0.270
One i.i.d. $\mathrm{N}(0, 2 * 0.64 * \widehat{\mu}_{IV})$ *Jump Every 2 Days*						
Nominal Size = 5%						
72	0.361	0.215	0.207	0.661	0.516	0.508
144	0.341	0.169	0.161	0.518	0.436	0.415
288	0.318	0.138	0.133	0.412	0.384	0.362
576	0.313	0.124	0.119	0.338	0.360	0.339
Nominal Size = 10%						
72	0.434	0.276	0.269	0.730	0.588	0.583
144	0.411	0.224	0.214	0.599	0.512	0.489
288	0.386	0.193	0.185	0.489	0.457	0.435
576	0.382	0.170	0.164	0.411	0.433	0.411
One i.i.d. $\mathrm{N}(0, 0.64 * \widehat{\mu}_{IV})$ *Jump Every Day*						
Nominal Size = 5%						
72	0.471	0.224	0.206	0.645	0.563	0.565
144	0.445	0.172	0.163	0.532	0.514	0.514
288	0.446	0.145	0.134	0.463	0.485	0.464
576	0.440	0.128	0.131	0.413	0.465	0.449
Nominal Size = 10%						
72	0.540	0.279	0.265	0.712	0.635	0.641
144	0.517	0.224	0.217	0.611	0.592	0.588
288	0.519	0.197	0.183	0.540	0.558	0.541
576	0.515	0.179	0.182	0.485	0.543	0.529
Panel B: Interval $= \widehat{\mu}_{IV} + 0.250\widehat{\sigma}_{IV}$						
One i.i.d. $\mathrm{N}(0, 3 * 0.64 * \widehat{\mu}_{IV})$ *Jump Every 5 Days*						
Nominal Size = 5%						
72	0.324	0.240	0.246	0.846	0.599	0.582
144	0.273	0.173	0.171	0.648	0.460	0.417
288	0.247	0.140	0.134	0.465	0.361	0.330
576	0.240	0.125	0.121	0.327	0.315	0.276
Nominal Size = 10%						
72	0.401	0.310	0.318	0.894	0.677	0.666
144	0.354	0.236	0.236	0.727	0.550	0.503
288	0.321	0.201	0.193	0.552	0.449	0.407
576	0.312	0.181	0.177	0.408	0.390	0.351
One i.i.d. $\mathrm{N}(0, 2 * 0.64 * \widehat{\mu}_{IV})$ *Jump Every 2 Days*						
Nominal Size = 5%						
72	0.474	0.241	0.225	0.857	0.688	0.687
144	0.438	0.166	0.153	0.704	0.586	0.560
288	0.413	0.131	0.121	0.557	0.508	0.479
576	0.392	0.107	0.101	0.449	0.465	0.436
Nominal Size = 10%						
72	0.562	0.316	0.296	0.902	0.764	0.760
144	0.527	0.227	0.212	0.772	0.671	0.645
288	0.500	0.185	0.175	0.642	0.599	0.561
576	0.482	0.160	0.155	0.532	0.556	0.520
One i.i.d. $\mathrm{N}(0, 0.64 * \widehat{\mu}_{IV})$ *Jump Every Day*						
Nominal Size = 5%						
72	0.673	0.273	0.244	0.858	0.781	0.788
144	0.641	0.188	0.171	0.749	0.722	0.718
288	0.633	0.151	0.135	0.651	0.678	0.659
576	0.623	0.125	0.118	0.590	0.657	0.643
Nominal Size = 10%						
72	0.746	0.351	0.317	0.903	0.841	0.843
144	0.717	0.251	0.235	0.813	0.793	0.791
288	0.713	0.208	0.192	0.731	0.751	0.738
576	0.703	0.182	0.172	0.671	0.732	0.723

[a] Notes: See notes to Table 1. All experiments are based on samples of 100 daily observations and 10,000 Monte Carlo iterations.

Table 5: 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	$\overline{RV}_T$	0.521	0.440
	$\overline{RV}_T$	0.578	0.426
	RS_T^-	0.543	0.461
	$\overline{RS}_T^-$	0.582	0.426
TPV	$\overline{TPV}_T$	0.501	0.318
	$\overline{TPV}_T$	0.562	0.378
	RS_T^-	0.539	0.360
	$\overline{RS}_T^-$	0.562	0.378
$\widehat{RV}$	$\overline{\widehat{RV}}_T$	0.503	0.476
	$\overline{\widehat{RV}}_T$	0.602	0.544
	RS_T^-	0.763	0.702
	$\overline{RS}_T^-$	0.659	0.602
$\widetilde{RV}$	$\overline{\widetilde{RV}}_T$	0.522	0.502
	$\overline{\widetilde{RV}}_T$	0.618	0.563
	RS_T^-	0.744	0.682
	$\overline{RS}_T^-$	0.659	0.603
RK	$\overline{RK}_T$	0.522	0.461
	$\overline{RK}_T$	0.577	0.522
	RS_T^-	0.757	0.701
	$\overline{RS}_T^-$	0.674	0.604

[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 6: 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	$\overline{RV}_T$	0.503	0.502
	$\overline{RV}_T$	0.578	0.579
	RS_T^-	0.601	0.604
	$\overline{RS}_T^-$	0.618	0.620
TPV	$\overline{TPV}_T$	0.683	0.541
	$\overline{TPV}_T$	0.720	0.660
	RS_T^-	0.681	0.578
	$\overline{RS}_T^-$	0.702	0.601
$\widehat{RV}$	$\overline{\widehat{RV}}_T$	0.541	0.519
	$\overline{\widehat{RV}}_T$	0.578	0.579
	RS_T^-	0.620	0.619
	$\overline{RS}_T^-$	0.661	0.662
$\widetilde{RV}$	$\overline{\widetilde{RV}}_T$	0.564	0.482
	$\overline{\widetilde{RV}}_T$	0.639	0.561
	RS_T^-	0.617	0.584
	$\overline{RS}_T^-$	0.615	0.578
RK	$\overline{RK}_T$	0.660	0.583
	$\overline{RK}_T$	0.704	0.601
	RS_T^-	0.681	0.655
	$\overline{RS}_T^-$	0.656	0.658

[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.