Predictive Inference for Integrated Volatility*

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

April 2006
this revision: November 2007

Abstract

In recent years, numerous volatility-based derivative products have been engineered. This has led to interest in constructing conditional predictive densities and confidence intervals for integrated volatility. In this paper, we propose nonparametric kernel estimators of the aforementioned quantities. The kernel functions used in our analysis are based on different realized volatility measures, which are constructed using the *ex post* variation of asset prices. A set of sufficient conditions under which the estimators are asymptotically equivalent to their unfeasible counterparts, based on an unobservable volatility process, is provided. Additionally, asymptotic normality is established. Based on the interpretation of related asymptotic properties that we derive, we additionally comment on a number of issues, including: the effects of using inconsistent volatility estimators in the presence of jumps and microstructure noise; and the asymptotic trade-offs between using robust realized volatility measures versus non-robust measures with differing data frequencies and different noise magnitudes, for example. The efficacy of the proposed estimators is examined via Monte Carlo experimentation, and an empirical illustration based upon data from the New York Stock Exchange is provided.

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

JEL classification. C22, C53, C14.

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

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

‡Imperial College, Tanaka 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, 2005). One interesting class of assets for which predictive conditional densities are relevant is that based on the use of volatility. Indeed, since the early days in 1993, when the VIX, an index of implied volatility, was created for the Chicago Board of Trade, a plethora of volatility-based derivative products has been engineered, including variance and covariance swaps, overshooters, and up and downcrossers, for example (see, e.g., Carr and Lee, 2003). One of the reasons why volatility based products now form an important class of assets is the stylized fact that volatility is counter cyclical (see Schwert, 1989), suggesting the adoption of volatility exposure in order to reduce the riskiness of a portfolio.

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

The first and most widely used estimator of integrated volatility is realized volatility, concurrently proposed by Andersen, Bollerslev, Diebold and Labys (2001), and Barndorff-Nielsen and Shephard (2002).[1] Realized volatility consistently estimates the increments of quadratic variation, when the underlying asset follows a semimartingale process, a class of processes which is commonly employed in continuous time modeling. Important variants of realized volatility have subsequently

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

been proposed. These variants are largely motivated by the need to provide consistent estimators of integrated volatility in situations which are quite common in financial markets, such as when jumps occur in the asset price process, and when there are market frictions leading to market microstructure noise. Leading examples include power variation (Barndorff-Nielsen and Shephard, 2004) and different estimators that are robust to the presence of microstructure noise (see, e.g., Zhang, 2006, Aït-Sahalia, Mykland and Zhang, 2005,2006, Zhang, Mykland and Aït-Sahalia, 2005, Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2006a,b). The estimators due to the above authors remain consistent for integrated volatility, in the presence of jumps, and when observed prices are affected by microstructure noise. The cost of implementing these new robust estimators is a loss of efficiency, or a slower rate of convergence. Since all of the estimators discussed above are designed to measure the *ex post* variation of asset prices, in the remainder of the paper we will call them realized volatility measures.

In this paper, we develop a method for constructing conditional (predictive) densities and associated conditional (predictive) confidence intervals for daily volatility, given observed market information. Exploiting the usual factorization of joint densities, our density estimator is derived as the ratio between a (nonparametric) kernel estimator of the joint density of current and future volatility, and a kernel estimator of the marginal density of current (and past) volatility. Our conditional confidence interval estimator is based on the standard Nadaraya-Watson estimator. We also consider local linear estimators of both conditional densities and conditional confidence intervals, along the lines of Fan, Yao and Tong (1996) and Hall, Wolff and Yao (1999). We show that the proposed estimators are consistent and asymptotically normally distributed, under very mild assumptions on the underlying diffusion process. Our results require no parametric assumption on either the functional form of the estimated densities, or on the specification of the diffusion process driving the asset price. Nevertheless, we require the diffusive part of the log-price process to be Brownian. In this sense, our approach might be viewed as semiparametric.

The intuition for the approach taken in the sequel is as follows. Since integrated volatility is unobservable, we use the realized (volatility) measures discussed above as a key ingredient in the construction of kernel estimators. However, this introduces a technical difficulty, as each realized measure can be decomposed into integrated volatility, the object of interest, and an error term. Formally,

$$RM_{t,M} = IV_t + N_{t,M},$$

where $RM_{t,M}$ and $N_{t,M}$ denote a particular realized volatility measure and its corresponding error term, respectively. Here, IV_t denotes integrated volatility, and the subscripts t and M denote a particular day, t, and the number of intraday observations, M, used in the construction of the realized measure. Our estimators are therefore based on a variable which is subject to measurement error. For this reason, we provide sufficient conditions under which conditional density (and confidence interval) estimators based on (the unobservable) integrated volatility and ones based on a realized measure are asymptotically equivalent. Of note is that these conditions vary across different realized measures.

There is a well developed literature on kernel estimation in the presence of variables measured with error. The two common assumptions in this literature are that the error has a characteristic function bounded away from zero everywhere, and that the error is independent of the variable of interest. Since the error term does not vanish (even asymptotically), consistent estimators of the object of interest cannot be obtained via implementation of standard kernel based methods. Consistent density estimators and corresponding convergence rates are derived in Fan (1991) and Fan and Truong (1993), via use of deconvolution methods, for the case in which the density of the measurement error is known. The case for which the error density is unknown is treated by Li and Vuong (1998) for density estimation, and by Schennach (2004) for estimation of regression functions. Both of these papers rely on particular properties of the Fourier transform of the kernel function. Our set-up is different. In our case, the measurement error approaches zero as $M \to \infty$ (which is implied by the fact that realized measures are consistent estimators), and if M grows fast enough relative to T, then standard kernel estimators constructed using realized measures are asymptotically equivalent to those constructed using the unknown integrated volatility.

The idea of using a realized measure as a basis for predicting integrated volatility has been adopted in other papers. Andersen, Bollerslev, Diebold and Labys (2003), Andersen, Bollerslev and Meddahi (2004, 2005) deal with the problem of pointwise prediction of integrated volatility, using ARMA models based on the log of realized volatility; the latter authors also investigate the crucial issue of evaluating the loss of efficiency associated with the use of realized volatility measures, relative to optimal (unfeasible) forecasts (based on the entire volatility path). Andersen, Bollerslev and Meddahi (2006), Aït-Sahalia and Mancini (2006), and Ghysels and Sinko (2006) address the issue of forecasting volatility in the presence of microstructure effects.[2] The papers cited above deal with pointwise prediction of integrated volatility. To the best of our knowledge, Corradi, Distaso and Swanson (2006) was the first paper to focus on estimation of the conditional

[2] Realized volatility measures have also been used to estimate and test the specification of stochastic volatility models (see, e.g., Bollerslev and Zhou, 2002 and Corradi and Distaso, 2006).

density of integrated volatility, by establishing uniform rates of convergence for kernel estimators based on realized measures. However, when questions arise with regard to notions such as hedging derivatives based on volatility, the crucial question becomes how to asses the interval within which future daily volatility will fall, with a given level of confidence, and for given information on current and past values of realized volatility. Corradi, Distaso and Swanson (2006) could not answer this question, as uniform convergence does not provide enough information. This paper provides an answer to this sort of question by establishing asymptotic normality for Nadaraya-Watson and local polynomial estimators of conditional confidence intervals. This is a substantially more challenging task than simply establishing uniform convergence, as the measurement error enters into the indicator function, which is non-differentiable, and thus standard mean value expansion arguments no longer apply. Indeed, the case of conditional densities can be treated essentially as a corollary of the conditional interval case. Moreover, the current paper differs from Corradi, Distaso and Swanson (2006) in two other important respects. First, instead of restricting our attention to the class of eigenfunction stochastic volatility models (see Meddahi, 2001), we consider the general class of *cadlag* (right continuous with left limit) volatility processes. This makes the computation of the moment structure of the measurement error much more complicated. Indeed, we provide sufficient primitive conditions under which

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

where $b_M \to \infty$, as $M \to \infty$, at a rate varying across different realized measures. Second, we study the asymptotic behavior of our predictive interval and predictive density estimators when things go wrong. In particular, we analyze the case in which volatility violates the strong-mixing assumption and instead exhibits long-range dependence; and the case in which we use non-robust volatility measures in the presence of jumps or microstructure noise.

In order to evaluate the sharpness of our theoretical results, we carry out a Monte Carlo experiment in which pseudo true predictive intervals are used in conjunction with intervals based on various realized measures including realized volatility, bipower variation, two scale realized volatility, multi scale realized volatility, and realized kernels, in order to assess the finite sample behavior of our statistics, in the presence of jumps or microstructure noise. This is done for various daily sample sizes and for a variety of intradaily frequencies. As expected, robust realized volatility measures yield substantially more accurate predictive intervals than the other measures, when data are subject to microstructure noise, for relatively large values of M. However, for small values of M, realized volatility performs the best; and in the presence of jumps, bipower variation is superior, as expected. In general, our experiment underscores the relative trade-offs between T and M, under various different data generating assumptions. An empirical illustration based on New York Stock Exchange data underscores the importance of using microstructure robust measures when using data sampled at a high frequency.

The rest of the paper is organized as follows. In Section 2, we describe the model and the different realized volatility measures for which asymptotic results are derived. Section 3 outlines the conditional density and confidence interval estimators and establishes their asymptotic properties. Section 4 studies the applicability of the established asymptotic results to various well known realized measures, including realized volatility, power variation, two-scale and multiscale estimators and realized kernel estimators. We also study the behavior of confidence interval and predictive density estimators when we erroneously do not take into account the presence of jumps or microstructure effects. In Section 5, the results of a Monte-Carlo experiment designed to assess the extent to which our asymptotic limiting distribution results yield accurate finite sample approximations are discussed. Section 6 contains an empirical illustration based upon data from the New York Stock Exchange. Finally, some concluding remarks are given in Section 7. All proofs are contained in the Appendix.

2 Setup

Denote the log-price of a generic financial asset as Y_t, at a continuous time, t. Assume that the log-price process belongs to the class of semimartingale processes with jumps. Denote this by writing $Y \in \mathcal{BSMJ}$. Then:

$$\mathrm{d}Y_t = \mu_t \mathrm{d}t + \sigma_t \left(\sqrt{1-\rho^2}\mathrm{d}W_{1,t} + \rho \mathrm{d}W_{2,t} \right) + \mathrm{d}J_{1,t}, \tag{1}$$

with $\mathrm{E}(\mathrm{d}W_{1,t}\mathrm{d}W_{2,t}) = 0$ and $-1 < \rho < 1$. The drift component, μ_t, is a predictable process; the diffusion term, σ_t, is a cadlag process whose properties are specified below; and $\mathrm{d}J_{1,t} = \int_U c(u) N_1(\mathrm{d}t, \mathrm{d}u)$, where $N_1([t_1,t_2],\mathcal{A})$ is a Poisson measure which counts the number of jumps, in the time interval $[t_1,t_2]$, whose size belongs to the set $\mathcal{A}$, and $c(u)$ is a $i.i.d.$ random variable describing the size of the jump. Note that, over a finite time span, only a finite number of jumps can occur. In fact $N_1([t_1,t_2],\mathcal{A}) = v(\mathcal{A})(t_2-t_1)$, where the (Levy) intensity measure $v(\mathcal{A})$, which denotes the expected number of jumps per unit of time, of size belonging to $\mathcal{A}$, is a measure of finite variation.

The volatility process, σ_t^2, can be a diffusion process:

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

or a jump diffusion process:

$$\mathrm{d}\sigma_t^2 = b_1\left(\sigma_t^2\right)\mathrm{d}t + b_2\left(\sigma_t^2\right)\mathrm{d}W_{2,t} + \mathrm{d}J_{2,t}, \tag{3}$$

where $\mathrm{d}J_{2,t}$ is defined just as is $\mathrm{d}J_{1,t}$, except for the fact that the jump sizes are positive valued; or, finally, σ_t^2 can be a stochastic differential equation driven by an infinite activity Levy process:

$$\sigma_t^2 = b_1\left(\sigma_t^2\right)\mathrm{d}t + b_2\left(\sigma_t^2\right)\mathrm{d}W_{2,t} + \mathrm{d}\mathcal{Z}_t \tag{4}$$

where $\mathrm{d}\mathcal{Z}_t = \int_{u>1} c(u)N_3(\mathrm{d}t, \mathrm{d}u) + \int_{0\leq u\leq 1} c(u)(N_3(\mathrm{d}t, \mathrm{d}u) - \nu(\mathrm{d}t, \mathrm{d}u))$, with $N_3(\mathrm{d}t, \mathrm{d}u)$ a Poisson random measure. Here, the associated intensity measure ν is now a measure of infinite variation. The heuristic reason why we need to separately consider the case of $0 \leq u \leq 1$ and $u > 1$ is that there are infinitely many small jumps, so that the intensity measure can blow up in the neighborhood of zero.

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

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

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

$$\langle Y^c\rangle_t = \int_0^t \sigma_s^2 \mathrm{d}s \ \text{ and } \ \left\langle Y^d\right\rangle_t = \int_0^t \int_U c(u)N_1(\mathrm{d}s, \mathrm{d}u).$$

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

Thus far, we have considered a market that is free from frictions. However, there is a substantial literature in financial economics that documents the presence of market distortion or friction, and that has identified several possible causes thereof (see, e.g., O'Hara, 1997). We introduce market friction in the following way. Assume that transaction data available in financial markets are contaminated by measurement error, so that the observed process is given by:

$$X = Y + \epsilon.$$

Thus, we allow for the possibility that the observed transaction price can be decomposed into the "true" price and a "noise" term arising due to measurement error, the latter of which captures generic microstructure effects. In order to properly manage financial risk, one is interested in the contribution to quadratic variation of the Brownian component of Y. Now, in order to study integrated volatility using econometric tools, assume that there are a total of MT equi-spaced observations from the process X, consisting of M intradaily observations for T days. More precisely, a sample of data is given by:

$$X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}, \;\; t = 1, \ldots, T \;\text{ and }\; j = 1, \ldots, M, \tag{5}$$

where $\epsilon_{t+j/M}$ is a zero-mean weakly dependent process. Following Aït-Sahalia, Mykland and Zhang (2006), we let the error term be geometrically mixing, so that for each s there is a constant, $\rho < 1$, which satisfies:

$$\mathrm{cov}(\epsilon_{t+j/M}, \epsilon_{t+(j+|s|)/M}) = \mathrm{E}(\epsilon_{t+j/M}\epsilon_{t+(j+|s|)/M}) \simeq \rho^s, \tag{6}$$

where with the notation $\simeq$ we mean "approximately equal to". As mentioned in the introduction, when deriving kernel estimators for conditional (predictive) densities and confidence intervals of integrated volatility, we make no assumptions on the functional forms of the drift, diffusion, and jump components in (1). We will also make no parametric assumptions on the form of the density that characterizes integrated volatility. However, in this paper we consider a subset of the class of semimartingales, namely the Brownian subset. In this respect, our approach is inherently semiparametric.

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

$$IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s, \;\; t = 1, \ldots, T. \tag{7}$$

Since IV_t is not observable, different realized measures, based on the sample $X_{t+j/M}$, are used as proxies for IV_t. The realized measure, $RM_{t,M}$, is a noisy measure of true integrated volatility. Namely:

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

In the sequel, we first derive kernel estimators for conditional densities and conditional confidence intervals based on a generic realized volatility measure. We then provide sufficient conditions on the structure of the measurement error, $N_{t,M}$, ensuring that the distributions of the estimators based on realized measures, and the corresponding distribution associated with the "true" (but latent) daily volatility process, are asymptotically equivalent. Finally, we adapt the given primitive conditions on $N_{t,M}$ to several particular realized measures of integrated volatility. In particular, consider the following realized measures:

(i) *Realized Volatility* (due to Andersen, Bollerslev, Diebold and Labys, 2001, and Barndorff-Nielsen and Shephard, 2002):

$$RV_{t,M} = \sum_{j=1}^{M-1} \left(X_{t+(j+1)/M} - X_{t+j/M}\right)^2. \tag{8}$$

(ii) *Normalized Bipower and Tripower Variation* (due to Barndorff-Nielsen and Shephard, 2004):

$$BV_{t,M} = (\mu_1)^{-2} \sum_{j=2}^{M-1} \left|X_{t+(j+1)/M} - X_{t+j/M}\right| \left|X_{t+j/M} - X_{t+(j-1)/M}\right|, \tag{9}$$

$$TV_{t,M} = \left(\mu_{2/3}\right)^{-3} \sum_{j=1}^{M-3} \left|\Delta X_{(j+3)/M}\right|^{2/3} \left|\Delta X_{(j+2)/M}\right|^{2/3} \left|\Delta X_{(j+1)/M}\right|^{2/3} \tag{10}$$

where $\mu_k = \mathrm{E}\left|SN^k\right|$ and SN is a standard normal random variable.

(iii) *Two Scale Realized Volatility* (due to Zhang, Mykland and Aït-Sahalia, 2005):

$$\widehat{RV}_{t,l,M} = RV^{avg}_{t,l,M} - 2l\widehat{\nu}_{t,M}, \tag{11}$$

where

$$\widehat{\nu}_{t,M} = \frac{RV_{t,M}}{2M} = \frac{1}{2M} \sum_{j=1}^{M-1} \left(X_{t+\frac{j}{M}} - X_{t+\frac{j-1}{M}}\right)^2 \tag{12}$$

and

$$RV^{avg}_{t,l,M} = \frac{1}{B} \sum_{b=1}^{B} RV^b_{t,l} = \frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left(X_{t+\frac{jB+b}{M}} - X_{t+\frac{(j-1)B+b}{M}}\right)^2. \tag{13}$$

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

(iv) *Multi Scale Realized Volatility* (due to Zhang, 2006):

$$\begin{aligned}\widetilde{RV}_{t,e,M} &= \sum_{i=1}^{e} a_i \widetilde{RV}_{t,e_i,M} + \frac{RV_{t,M}}{M} \\ &= \sum_{i=1}^{e} a_i \frac{1}{e_i} \left(\sum_{j=1}^{M-e_i} \left(X_{t+\frac{j+e_i}{M}} - X_{t+\frac{j}{M}}\right)^2 \right) + \frac{RV_{t,M}}{M},\end{aligned} \tag{14}$$

where the weights a_i have to satisfy the following two restrictions

$$\sum_{i=1}^{e} a_i = 1 \text{ and } \sum_{i=1}^{e} \frac{a_i}{i} = 0.$$

Thus, $\widetilde{RV}_{t,e,M}$ is a linear weighted combination of e realized volatilities computed over e different frequencies e_i/M, with $i = 1, \ldots, e$. For $e_i = i$,

$$a_i = 12 \frac{i}{e^2} \frac{\left(\frac{i}{e} - \frac{1}{2} - \frac{1}{2e}\right)}{\left(1 - \frac{1}{e^2}\right)}.$$

(v) *Realized kernels* (due to Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2006a):

$$\mathcal{K}_{t,H,M} = \sum_{h=1}^{H} \kappa\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}^{X} + \gamma_{t,-,h}^{X}\right), \tag{15}$$

where $\kappa(0) = 1$ and $\kappa(1) = 0$, and

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

If $\kappa(x) = 1-x$ and $H = M^{2/3}$, then $\mathcal{K}_{t,H,M}$ and $\widehat{RV}_{t,l,M}$ are asymptotically equivalent. Also, if $\kappa(x) = 1-3x^2+2x^3$ and $H = M^{1/2}$ then $\mathcal{K}_{t,H,M}$ and $\widetilde{RV}_{t,e,M}$ are asymptotically equivalent.

Additionally, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006b) suggest the use of *subsampled realized kernels*, which are defined as

$$\mathcal{SK}_{t,H,M} = \sum_{h=1}^{H} \kappa\left(\frac{h-1}{H}\right)\frac{1}{S}\sum_{s=1}^{S}\left(\gamma_{s,t,h}^{X} - \gamma_{s,t,-h}^{X}\right),$$

with

$$\gamma_{t,s,h}^{X} = \sum_{j=H}^{M-H-1} \left(X_{t+(j+1+(s-1)/S)/M} - X_{t+(j+(s-1)/S)/M}\right)\left(X_{t+(j+1-h+(s-1)/S)/M} - X_{t+(j-h+(s-1)/S)/M}\right),$$

and discuss the relative advantages and disadvantages of subsampling, depending on the smoothness of the kernel function.

3 Asymptotics

Our objective is to construct a nonparametric estimator of the confidence interval of integrated volatility at time $T+\tau$, conditional on a given realized volatility measure that is observed at times $T, \ldots, T-(d-1)$. To simplify the discussion, and without loss of generality, we confine our attention to the case where $\tau = 1$. Extension to the case of $\tau > 1$ follows directly, for finite τ. Loosely speaking, our objective is to predict the density and the confidence interval of integrated volatility at time $T+1$, using the information contained in contemporaneous as well as $d-1$ lags of a given realized volatility measure.

Hereafter, let:

$$RM_{t,M}^{(d)} = (RM_{t,M}, \ldots, RM_{t-(d-1),M}).$$

Analogously, define:

$$IV_t^{(d)} = (IV_t, \ldots, IV_{t-(d-1)}).$$

In the sequel, we study Nadaraya-Watson estimators for conditional confidence intervals:

$$
\begin{aligned}
&\widehat{F}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(u_2|RM_{T,M}^{(d)}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(u_1|RM_{T,M}^{(d)}) \\
&= \frac{\frac{1}{T}\sum_{t=d}^{T-1} 1_{\{u_1 \le RM_{t+1,M} \le u_2\}} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}}{\frac{1}{T}\sum_{t=d}^{T-1} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}} \\
&= \frac{\frac{1}{T}\sum_{t=d}^{T-1} 1_{\{u_1 \le RM_{t+1,M} \le u_2\}} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}}{\widehat{f}_{RM_{T,M}^{(d)}}(RM_{T,M}^{(d)})};
\end{aligned} \tag{16}
$$

and for conditional densities:

$$
\widehat{f}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)}) = \frac{\frac{1}{T}\sum_{t=d}^{T-1} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d} K\left(\frac{RM_{t+1,M} - x}{\xi_2}\right)\frac{1}{\xi_2}}{\frac{1}{T}\sum_{t=d}^{T-1} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}}. \tag{17}
$$

Let K be a univariate kernel function, and let $\mathbf{K}$ be either be a $d-$dimensional kernel, or the product of d univariate kernel functions, such as:

$$
\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) = \prod_{i=1}^{d} Q\left(\frac{RM_{t-i+1,M} - RM_{T-i+1,M}}{\xi_1}\right),
$$

where Q may or may not be equal to K, and where ξ, ξ_1 and ξ_2 are bandwidth parameters.

We need the following assumptions.

Assumption A1: $IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s$ is a strictly stationary $\alpha-$mixing process with mixing coefficients satisfying $\sum_{j=1}^{\infty} j^{\lambda} \alpha_j^{1-2/\delta} < \infty$, with $\lambda > 1 - 2/\delta$.

Assumption A2:

(i) The kernel, $\mathbf{K}$, is a symmetric, nonnegative, continuous function with bounded support $[-\Delta, \Delta]^d$; and is at least twice differentiable on the interior of its support, satisfying:

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

(ii) Let $\mathbf{K}_i^{(j)}$ be the $j-th$ derivative of the kernel with respect to the $i-$th variable. Then, $\mathbf{K}_i^{(j)}(-\Delta) = \mathbf{K}_i^{(j)}(\Delta) = 0$, for $i = 1, \ldots, d$, $j = 1, \ldots, J$, $J \ge 1$.

Assumption A3:

(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 A4:

(i) $f_{IV_T^{(d)}}(\cdot)$ and, for any fixed x, $f_{IV_{T+1}|IV_T^{(d)}}(x|\cdot)$ are absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^d$, and $\omega-$times continuously differentiable on $\mathbb{R}^d$, with $\omega \geq 2$.

(ii) For any fixed x, u and $RM_{T,M}^{(d)}$, $f_{IV_t^{(d)}}(RM_{T,M}^{(d)}) > 0$, $f_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)}) > 0$, and $0 < F_{IV_{t+1}|IV_t^{(d)}}(u|RM_{T,M}^{(d)}) < 1$.

Assumption A5: There exists a sequence b_M, with $b_M \to \infty$, as $M \to \infty$, such that:

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

where $N_{t,M} = RM_{t,M} - IV_t$.

Assumption A1 requires the daily volatility process to be strong mixing. On the other hand, we do not impose memory conditions on Y_t, provided it is generated as in (1). Of note is that the mixing coefficients of the integrated and of the instantaneous volatility process are of the same order of magnitude. In fact, let $\mathcal{B}_{\sigma^2,t_1}^{t_2}$ and $\mathcal{B}_{IV,t_1}^{t_2}$ be the sigma-fields generated by $\left(\sigma_s^2, t_1 \leq s \leq t_2\right)$ and by $\left(\int_{s-1}^{s} \sigma_u^2 \mathrm{d}u, t_1 \leq s \leq t_2\right)$, respectively, and also define:

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

and

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

It is immediate to see that $\alpha_{IV}(m) \leq C\alpha_{\sigma^2}(m-1)$, for some constant C. Thus, it suffices that the instantaneous volatility process satisfies A1. For the non-jump case, geometric ergodicity, and strong mixing with mixing coefficients decaying at a geometric rate, follow from the drift condition in Meyn and Tweedie (1993, p.536):

$$\nabla f(\sigma^2) b_1(\sigma^2) + \frac{1}{2}\nabla^2 f(\sigma^2) b_2^2\left(\sigma^2\right) \leq -c_1 f(\sigma^2) + c_2,$$

where c_1, c_2 are positive constants and f is a norm-like function, with the property that as $\sigma^2 \to \infty$, $f(\sigma^2) \to \infty$. Thus, in terms of equation (2), taking $f(\sigma^2) = (\sigma^2)^2$, the drift condition is simply

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

which is satisfied, for example, when drift and variance terms in (2) grow at most at a linear rate, and if there is mean reversion (i.e. if $b_1(\sigma^2)$ is a decreasing function of σ^2). Masuda (2004, Section 3) provides conditions for geometric ergodicity of jump diffusions and in general for diffusions driven by Levy processes. For the case in which the jump size does not depend on the state variable, as is the case for equations (3) and (4), it suffices that $\int c(u)^2 v(u)\mathrm{d}u < \infty$, where $v(u)$ is the Levy intensity measure.

It should be pointed out that A1 does not rule out the possibility that volatility is a deterministic function of the price of the underlying asset, provided that the log-price follows a stationary mixing process.

A3 and A4 are somewhat standard assumptions in the literature on nonparametric density estimation. The reason why we require a kernel function with a bounded support is the following. Realized volatility measures are by construction (or can be made, in the case of multi-scale and realized kernel estimators) non negative. Thus, if we were to use a kernel function with unbounded support, the estimated density would suffer from a downward bias, as some weight would be given to negative observations. In this regard, it suffices to use a boundary corrected kernel function for those values of x and $RM_{T,M}^{(d)}$ that are "close" to the lower bound of the support. The use and the choice of boundary corrected kernels will be explained in Section 6.

Assumption A5 requires that the $k-$th moment of the measurement error decays to zero fast enough, as $M \to \infty$. Additionally, it implicitly requires that M grow fast enough relative to T. In Section 4, we shall provide primitive conditions, in terms of moments of σ_t^2 and μ_t, under which A5 is satisfied by the five realized measures defined in (8),(9),(11), (14) and (15).

For conditional intervals, we have the following result:

Theorem 1. *Let A1-A5 hold. If $\xi \to 0$, $T\xi^d \to \infty$, $T\xi^{d+4} \to 0$, and $T^{\frac{2k+6}{2k}} b_M^{-1} \xi^d \to 0$, then:*

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

where

$V(u_1, u_2)$

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

$$
\left. \times \left(1 - \left(\left(F_{IV_{T+1}|IV_T^{(d)}}(u_2|RM_{T,M}^{(d)}) - F_{IV_{T+1}|IV_T^{(d)}}(u_1|RM_{T,M}^{(d)}) \right) \right) \right) \right).
$$

Corollary 1. *Let A1-A5 hold. If* $\xi \to 0$, $T\xi^d \to \infty$, $T\xi^{d+4} \to 0$, *and* $T^{\frac{2k+6}{2k}} b_M^{-1} \xi^d \to 0$, *then:*

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

where

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

The key point in the proof of this theorem is to show the asymptotic equivalence between the estimator based on realized measures and that based on integrated volatility, that is to show that:

$$
\begin{aligned}
&\frac{1}{T} \sum_{t=d}^{T-1} \left(1_{\{u_1 \le RM_{t+1,M} \le u_2\}} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi} \right) \frac{1}{\xi^d} - 1_{\{u_1 \le IV_{t+1,M} \le u_2\}} \mathbf{K}\left(\frac{IV_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi} \right) \frac{1}{\xi^d} \right) \\
&= o_p\left(\sqrt{T\xi^d} \right).
\end{aligned}
$$

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 $\sqrt{T\xi^d}$.

Turning now to our predictive density estimator, we begin by considering the case in which the evaluation points, x and $RM_{T,M}$, are away from the boundary, so that no correction is necessary. The issue of boundary correction is treated in the empirical application (see Section 6).

Theorem 2. *Let A1-A5 hold. If* $\xi_1, \xi_2 \to 0$, $T\xi_1^d\xi_2 \to \infty$, $T\xi_1^{4+d}\xi_2 \to 0$, $T\xi_1^d\xi_2^5 \to 0$, *and* $T^{\frac{k+1}{k-1}} b_M^{-1} \xi_1^d \xi_2 \to 0$, *then:*

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

$$\xrightarrow{d} \mathrm{N}\left(0, \left(\frac{f_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})}{f_{IV_T^{(d)}}(RM_{T,M}^{(d)})}\int \mathbf{K}^2(u)\mathrm{d}u\int K^2(v)\mathrm{d}v\right)\right). \tag{18}$$

Corollary 2. *Let A1-A5 hold. If* $\xi_1, \xi_2 \to 0$, $T\xi_1^d\xi_2 \to \infty$, $T\xi_1^{4+d}\xi_2 \to 0$, $T\xi_1^d\xi_2^5 \to 0$, *and* $T^{\frac{k+1}{k-1}}b_M^{-1}\xi_1^d\xi_2 \to 0$, *then:*

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

3.1 Local Linear Estimator Results

Thus far, we have used standard Nadaraya-Watson kernel estimators to construct conditional confidence intervals and densities. A viable alternative is to use local linear estimators. One advantage of such estimators is that local polynomial estimators, for example, do not suffer of the boundary problem.

As a local linear estimator of conditional confidence intervals, define $\widehat{\boldsymbol{\alpha}}_{T,M}(u_1, u_2, RM_{T,M}^{(d)})$. In particular, let:

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

where

$$\begin{aligned}
&Z_{T,M}(\boldsymbol{\alpha}; u_1, u_2, RM_{T,M}^{(d)})\\
=&\frac{1}{T\xi^d}\sum_{t=d}^{T}\left(1_{\{u_1\le RM_{t+1}\le u_2\}} - \alpha_0 - \boldsymbol{\alpha}_1'\left(RM_{t,M}^{(d)} - RM_{T,M}^{(d)}\right)\right)^2 \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)
\end{aligned}$$

and $\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}_1')'$. Then:

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_{T,M}(u_1, u_2, RM_{T,M}^{(d)}) &= \begin{pmatrix} \widehat{\alpha}_{0,T,M}(u_1, u_2, RM_{T,M}^{(d)})\\ \widehat{\alpha}_{1,T,M}(u_1, u_2, RM_{T,M}^{(d)})\\ \vdots \\ \widehat{\alpha}_{d,T,M}(u_1, u_2, RM_{T,M}^{(d)})\end{pmatrix}\\
&= \left(\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)^{-1}\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\widetilde{\mathbf{y}}_{(M)}(u_1, u_2),
\end{aligned}$$

with

$$\mathbf{X}_{(M)}=\begin{pmatrix} 1 & RM_{d,M} - RM_{T,M} & \cdots & RM_{1,M} - RM_{T-(d-1),M}\\ 1 & RM_{d+1,M} - RM_{T,M} & \cdots & RM_{2,M} - RM_{T-(d-1),M}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & RM_{T-1,M} - RM_{T,M} & \cdots & RM_{T-(d-2),M} - RM_{T-(d-1),M}\end{pmatrix}, \tag{19}$$

$$
\mathbf{W}_{(M)}=\frac{1}{\xi^{d}}\begin{pmatrix} \mathbf{K}\left(\frac{RM_{d,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right) & 0 & \cdots & 0 \\ 0 & \mathbf{K}\left(\frac{RM_{d+1,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{K}\left(\frac{RM_{T,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right) \end{pmatrix}, \quad (20)
$$

$$
\widetilde{\mathbf{y}}_{(M)}(u_1,u_2)=\begin{pmatrix} 1_{\{u_1\le RM_{d+1}\le u_2\}} \\ 1_{\{u_1\le RM_{d+2}\le u_2\}} \\ \vdots \\ 1_{\{u_1\le RM_{T+1}\le u_2\}} \end{pmatrix}.
$$

The local linear estimator of the conditional confidence interval is given by $\widehat{\alpha}_{0,T,M}(u_1,u_2,RM_{T,M}^{(d)})$. The local linear estimator for conditional distributions outlined above has been recently used by Aït-Sahalia, Fan and Peng (2006), in the context of tests for the correct specification of diffusion models. Such an estimator is not ensured to lie between 0 and 1 in finite samples. More complex estimators based, for example, on logistic approximations, do lie between 0 and 1 for any sample size (see Hall, Wolff and Yao, 1999). However, they typically cannot be written in closed form.

For local linear conditional density estimation, following Fan, Yao and Tong (1996), consider $\widehat{\boldsymbol{\beta}}_{T,M}(x,RM_{T,M}^{(d)})$,where:

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

where

$$
\begin{aligned}
&S_{T,M}(\boldsymbol{\beta};x,RM_{T,M}^{(d)}) \\
=&\frac{1}{T\xi_1^d\xi_2}\sum_{t=d}^{T}\left(K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)-\beta_0-\boldsymbol{\beta}_1'\left(RM_{t,M}^{(d)}-RM_{T,M}^{(d)}\right)\right)^2\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right),
\end{aligned}
$$

$\boldsymbol{\beta}=(\beta_0,\boldsymbol{\beta}_1')'$, and K and $\mathbf{K}$ are defined as in (17). Therefore:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{T,M}(x,RM_{T,M}^{(d)}) &= \begin{pmatrix} \widehat{\beta}_{0,T,M}(x,RM_{T,M}^{(d)}) \\ \widehat{\beta}_{1,T,M}(x,RM_{T,M}^{(d)}) \\ \vdots \\ \widehat{\beta}_{d,T,M}(x,RM_{T,M}^{(d)}) \end{pmatrix} \\
&= \left(\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)^{-1}\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\mathbf{y}_{(M)},
\end{aligned}
$$

where $\mathbf{X}_{(M)}$ and $\mathbf{W}_{(M)}$ are defined as in (19) and (20), with ξ_1 instead of ξ, and with:

$$\mathbf{y}_{(M)}=\begin{pmatrix} \frac{1}{\xi_2}K\left(\frac{RM_{2,M}-x}{\xi_2}\right) \\ \frac{1}{\xi_2}K\left(\frac{RM_{3,M}-x}{\xi_2}\right) \\ \vdots \\ \frac{1}{\xi_2}K\left(\frac{RM_{T,M}-x}{\xi_2}\right) \end{pmatrix}.$$

We have the following result.

Theorem 3. *Let A1-A5 hold. Then:*

(i) If $\xi \to 0$, $T\xi^d \to \infty$, $T\xi^{d+4} \to 0$, *and* $T^{\frac{2k+6}{2k}} b_M^{-1} \xi^d \to 0$, *then:*

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

where $V(u_1, u_2)$ *is defined as in the statement of Theorem 1.*

(ii) If $\xi_1, \xi_2 \to 0$, $T\xi_1^d\xi_2 \to \infty$, $T\xi_1^{4+d}\xi_2 \to 0$, $T\xi_1^d\xi_2^5 \to 0$, *and* $T^{\frac{k+1}{k-1}} b_M^{-1} \xi_1^d\xi_2 \to 0$, *then:*

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

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

4 Applications

We now provide primitive conditions on the moments of the drift and variance terms which ensure that Assumption A5 is satisfied by the five realized measures outlined in Section 2. Let $RV_{t,M}$, $BV_{t,M}$, $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $\mathcal{K}_{t,H,M}$ be as defined in (8), (9), (11), (13) and (15), respectively. We have the following result.[3]

Lemma 1. *Let* Y_t, σ_t^2 *and* ϵ_t *be defined as in (1), (2) or (3) or (4), and (5)-(6) respectively. If* σ_t^2 *is* $\alpha-$*mixing with size* $-2r/(r-2)$, $r > 2$, $\mathrm{E}\left(\left(\sigma_t^2\right)^{2(k+r)}\right) < \infty$ *and* $\mathrm{E}\left(\left(\mu_t\right)^{2(k+r)}\right) < \infty$, *then there is a sequence* b_M, *where* $b_M \to \infty$ *as* $M \to \infty$, *such that:*

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

(i) If $J_{1,t} \equiv 0$ for all t (no jumps) and $\epsilon \equiv 0$ (no microstructure noise), then:

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

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

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

(iii) If $J_{1,t} \equiv 0$ for all t (no jumps), $\mathrm{E}\left(\epsilon_{t+j/M}\right)^{2k+\eta} < \infty$, $\eta > 0$, $\mathrm{E}\left(Y_{t+j/M}\epsilon_{t+j/M}\right) = 0$, and $l/M^{1/3} \to \pi$, and $0 < \pi < \infty$, then:

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

(iv) If $J_{1,t} \equiv 0$ for all t (no jumps), $\mathrm{E}\left(\epsilon_{t+j/M}\right)^{2k+\eta} < \infty$, $\eta > 0$, $\mathrm{E}\left(Y_{t+j/M}\epsilon_{t+j/M}\right) = 0$, and $e/M^{1/2} \to \pi$, $0 < \pi < \infty$, then:

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

(v) If $J_{1,t} \equiv 0$ for all t (no jumps), $\mathrm{E}\left(\epsilon_{t+j/M}\right)^{4+\eta} < \infty$, $\mathrm{E}\left(Y_{t+j/M}\epsilon_{t+j/M}\right) = 0$, then for $k = 2$:

$$\mathrm{E}\left(|\mathcal{K}_{t,H,M} - IV_t|^2\right) = O(b_M^{-1}) \text{ with } b_M = M^{1/3} \text{ if } H = M^{2/3} \text{ and } b_M = M^{1/2} \text{ if } H = M^{1/2},$$

and, for $k > 2$, $\mathrm{E}\left(|\mathcal{K}_{t,H,M} - IV_t|^k\right)$ diverges at rate $M^{k/2}$.

The statement in part (v) of the lemma applies also to subsampled realized kernels, though for $k > 2$, $\mathrm{E}\left(|\mathcal{SK}_{t,H,M} - IV_t|^k\right)$ diverges at rate $(M/S)^{k/2}$ instead of at rate $M^{k/2}$, where S is the number of subsamples.

From Lemma 1, it is immediate to see that Assumption A5 is satisfied for a generic k, by realized volatility and power variation measures with $b_M = M$, and by two-scale and multi-scale estimators with $b_M = M^{1/3}$ and $b_M = M^{1/2}$, respectively, provided that the drift and instantaneous variance have finite $2(k+r)-$th moments, and that $\mathrm{E}\left(\epsilon_{t+j/M}\right)^{2k+\eta} < \infty$.

It is somewhat surprising that, for $k > 2$, all $k-$th moments of the measurement errors associated with realized kernels diverge, as $M \to \infty$. Indeed, as neatly shown by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a), a judicious choice of the kernel function and of H ensures that the realized kernels are more efficient than two-scale and multi-scale estimators; in this sense $\mathrm{E}\left(|\mathcal{K}_{t,H,M} - IV_t|^2\right)$ can be made smaller than $\mathrm{E}\left(\left|\widetilde{RV}_{t,e,M} - IV_t\right|^2\right)$, for example. The question then is what is going wrong when we move from $k = 2$ to $k > 2$. Here, we just provide an heuristic argument. A detailed proof is provided in the Appendix. First, the divergence of higher moments is due to the pure noise component. For the sake of simplicity, consider the case of independent noise. Recalling (15), we concentrate on the term $\mathrm{E}\left(\sum_{h=2}^{H}\kappa\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}^{\epsilon}-\gamma_{t,-h}^{\epsilon}\right)\right)^{k}$, where $\gamma_{t,h}^{\epsilon}=\sum_{j=H}^{M-H-1}\left(\epsilon_{t+(j+1)/M}-\epsilon_{t+j/M}\right)\left(\epsilon_{t+(j+1-h)/M}-\epsilon_{t+(j-h)/M}\right)$. Now, for $k=2$, up to initial and end effects:

$$\mathrm{E}\left(\sum_{h=2}^{H}\kappa\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}^{\epsilon}+\gamma_{t,-h}^{\epsilon}\right)\right)^{2}=4\mathbf{u}'\mathrm{E}\left(\mathbf{\Gamma}_t\right)\mathbf{u},$$

where $\mathbf{u}'=\begin{pmatrix}\kappa(1/H) & \kappa(2/H) & \cdots & \kappa((H-1)/H)\end{pmatrix}$ and $\mathbf{\Gamma}_t=\left\{\gamma_{t,i}^{\epsilon}\gamma_{t,j}^{\epsilon}\right\}$, $i,j=2,\ldots,H$.

As $\mathrm{E}\left(\gamma_{t,i}^{\epsilon}\gamma_{t,j}^{\epsilon}\right)=0$ for $|i-j|>2$, and, up to some end effect, for $h>2$,

$$\mathrm{E}\left(\gamma_{t,h}^{\epsilon}\gamma_{t,h-2}^{\epsilon}+\gamma_{t,h}^{\epsilon}\gamma_{t,h-1}^{\epsilon}+\gamma_{t,h}^{\epsilon 2}+\gamma_{t,h}^{\epsilon}\gamma_{t,h+2}^{\epsilon}+\gamma_{t,h}^{\epsilon}\gamma_{t,h+2}^{\epsilon}\right)=(1-4+6-4+1)=0,$$

it follows for $j=3,...,H-2$, that the sum of elements of the j-th row of $\mathrm{E}\left(\mathbf{\Gamma}_t\right)$ is equal to zero. The contribution of the first and last two rows, weighted by the kernels, is of order nH^{-3} or nH^{-2}, depending on whether we use a smooth or kinked kernel. Now, consider the fourth moment. Again, up to initial and end effects:

$$\mathrm{E}\left(\sum_{h=2}^{H}\kappa\left(\frac{h-1}{H}\right)\left(\gamma_{t,h}^{\epsilon}+\gamma_{t,-h}^{\epsilon}\right)\right)^{4}=16\left(\mathbf{u}'\otimes\mathbf{u}'\right)\mathrm{E}\left(\mathbf{\Gamma}_t\otimes\mathbf{\Gamma}_t\right)\left(\mathbf{u}\otimes\mathbf{u}\right).$$

As shown in the Appendix, the sum of the elements in each row of $\mathrm{E}\left(\mathbf{\Gamma}_t\otimes\mathbf{\Gamma}_t\right)$ does not sum up to zero and is of order M^2. Hence, the divergence result stated in (v). The same type of reasoning applies to subsampled realized kernels, though the divergence occurs at a slower rate because of the averaging over subsamples. Indeed, as can be seen from the proof of (iii) and (iv) above, it is the joint effect of subsample averaging, returns computed over a sparse enough grid, and bias correction which ensures that the contribution of the pure noise component to higher moments approaches zero at an appropriate rate.

Finally, it should be mentioned that some form of correlation between noise and price can be allowed. For example, along the lines of Kalnina and Linton (2007), we could have modelled the noise as $\epsilon_{t+i/M}=v_{t+i/M}+u_{t+i/M}$, where $u_{t+i/M}$ is a mean zero geometrically mixing process and $v_{t+i/M}=\vartheta_M\left(W_{t+i/M}-W_{t+(i-1)/M}\right)$, where $\vartheta_M=O(M^{1/2})$, and W is the Brownian motion driving the price process. Also, for the case of realized kernels, following Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a), we could have modelled the endogenous part of the noise $v_{t+i/M}$ as $v_{i+t/M}=\sum_{h=0}^{\overline{H}}\beta_h\left(Y_{t+(i-h)/M}-Y_{t+(i-h-1)/M}\right)$, where $\{\beta_h\}_{h=1}^{\overline{H}}$ is a bounded sequence.

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

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

The case of jumps is relatively straightforward. Given (1), by Barndorff-Nielsen and Shephard (2004), note that:

$$BV_{t,M} - \left(IV_t + \int_{t-1}^{t} \int_U c(u) N_1(\mathrm{d}s, \mathrm{d}u) \right) = o_p(1).$$

Also, given that N_1 is a process with stationary and independent increments, if A1 holds for IV_t, it also holds for $IV_t + \int_{t-1}^{t} \int_U c(u) N_1(\mathrm{d}s, \mathrm{d}u)$. Hereafter, let $\langle Y \rangle_{t-1}^{t}$ be the quadratic variation of the return process over the interval $[t-1, t]$, so that

$$\langle Y \rangle_{t-1}^{t} = IV_t + \int_{t-1}^{t} \int_U c(u) N_1(\mathrm{d}s, \mathrm{d}u),$$

and let $\langle Y \rangle_{t-1}^{t,d} = \left(\langle Y \rangle_{t-2}^{t-1}, \cdots, \langle Y \rangle_{t-d}^{t-d-1} \right)$. The statements in Theorem 1 and 3(i) then hold with $F_{IV_{T+1}|IV_T^{(d)}}(u_1|RM_{T,M}^{(d)})$ replaced by $F_{\langle Y \rangle_T^{T+1}|\langle Y \rangle_{T-1}^{T,d}}(u_1|RM_{T,M}^{(d)})$, and the statements in Theorem 2 and 3(ii) hold with $f_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})$ replaced by $f_{\langle Y \rangle_T^{T+1}|\langle Y \rangle_{T-1}^{T,d}}(x|RM_{T,M}^{(d)})$. It is also immediate to see that the statement in part (i) of the Lemma holds with IV_t replaced by $\langle Y \rangle_{t-1}^{t}$. Thus, one can construct predictive density and confidence intervals which are valid for the "total" quadratic variation process, but not the integrated volatility process.

The case of microstructure noise contamination is more complex. We need to distinguish among three possible cases: (i) $\mathrm{var}(\epsilon_{t+i/M}) = o(M^{-1})$, (ii) $\mathrm{var}(\epsilon_{t+i/M}) = O(M^{-1})$ and (iii) $\mathrm{var}(\epsilon_{t+i/M}) = O(a_M)$, where as $M \to \infty$, $a_M \to \kappa$, with $0 \leq \kappa < \infty$, and $Ma_M \to \infty$. Note that the case of noise with variance independent of the sampling frequency is included in (iii), for $\kappa > 0$. In all other cases, the variance of the noise approaches 0, as $M \to \infty$, albeit at different speed. The choice between modelling the noise variance as fixed or approaching zero, as the number of intraday observation grows, is not an easy one. Empirically, microstructure noise is very small (for example, Bandi and Russell, 2007, report values ranging from 0.87e-07 to 2.1e-07). Based on this observation, Zhang, Mykland and Aït-Sahalia (2006) argue that the noise is indeed "too small" to be considered $O_p(1)$, and derive an Edgeworth correction for the two-scale estimator of Zhang, Mykland and Aït-Sahalia (2005), under the assumption that it approaches zero as $M \to \infty$.[4]

[4] Awartani, Corradi and Distaso (2007) suggest a test for the null hypothesis of a constant noise variance versus

For brevity and notational simplicity, we consider the case of independent noise. However, all statements carry through to the case in which $\mathrm{E}(\epsilon_{t+j/M}\epsilon_{t+(j+|s|)/M}) \simeq \rho^s$. It is immediate to see that in case (i), the effect of the microstructure noise is asymptotically negligible, and all asymptotic results carry over as in the noiseless case. It is also easy to see that case (ii) can be treated in the same way as the case of jumps. In fact, as $M \to \infty$,

$$RV_{t,M} \xrightarrow{p} IV_t + 2v,$$

where $v = \lim_{M\to\infty} M\mathrm{var}\left(\epsilon_{t+i/M}\right)$.

Thus, the statements in Theorem 1 and 3(i) hold with $F_{IV_{T+1}|IV_T^{(d)}}(u_1|RM_{T,M}^{(d)})$ replaced by $F_{IV_{T+1}+2v|IV_T^{(d)}+2v}(u_1|RM_{T,M}^{(d)})$, and the statements in Theorems 2 and 3(ii) with $f_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})$ replaced by $f_{IV_{T+1}+2v|IV_T^{(d)}+2v}(x|RM_{T,M}^{(d)})$. Moreover, the statement in part (i) of Lemma 1 holds with IV_t replaced by $IV_t + 2v$. Turning finally to case (iii), for sake of notational simplicity, but without loss of generality, assume that $d = 1$. Now consider predictive confidence intervals. Predictive densities can be treated along the same lines. We show that for all $u \in \mathcal{U}$, where $\mathcal{U}$ is a compact set of $\mathbb{R}_+$,

$$\widehat{F}_{RV_{T+1,M}|RV_{T,M}}(u|RV_{T,M}) = O_p(M^{-1/2}) \tag{21}$$

and for any $\varepsilon > 0$,

$$\lim_{T,M\to\infty} \Pr\left(\left(\widehat{F}_{RV_{T+1,M}|RM_{T,M}}(u|RV_{T,M}) - F_{IV_{T+1}|IV_T}(u|RV_{T,M})\right) < -\varepsilon\right) = 1, \tag{22}$$

so that $\sqrt{T\xi}\left(\widehat{F}_{RV_{T+1,M}|RM_{T,M}}(u_2|RV_{T,M}) - F_{IV_{T+1}|IV_T}(u_1|RV_{T,M})\right)$ diverges to minus infinity at rate $\sqrt{T\xi}$ as $M, T \to \infty$.

Turning to equation (21), and recalling equation (16), it suffices to show that:

$$\frac{1}{T\xi}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{RV_{t,M} - RV_{T,M}}{\xi}\right) = O_p\left(a_M^{-1}M^{-1/2}\right) \tag{23}$$

and

$$\frac{1}{T\xi}\sum_{t=d}^{T-1}1_{\{RV_{t+1,M}\leq u\}}\mathbf{K}\left(\frac{RV_{t,M} - RV_{T,M}}{\xi}\right) = O_p\left(a_M^{-1}M^{-1}\right). \tag{24}$$

Now, we can write:

$$\begin{aligned}&RV_{t,M} - RV_{T,M}\\&= \sum_{j=1}^{M-1}\left(\left(\left(Y_{t+(j+1)/M} - Y_{t+j/M}\right)^2 - \left(Y_{T-1+(j+1)/M} - Y_{T-1+j/M}\right)^2\right)\right.\end{aligned}$$

the alternative of a variance approaching zero, at a slower rate than $\sqrt{M}$. When applied to DJIA stocks, they find that the variance of the noise tends to zero, at frequencies of less than one minute.

$$+ a_M \sum_{j=1}^{M-1} \left(a_M^{-1} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right)^2 - a_M^{-1} \left(\epsilon_{T-1+(j+1)/M} - \epsilon_{T-1+j/M}\right)^2 \right)$$

$$+ 2a_M^{1/2} \sum_{j=1}^{M-1} \left(a_M^{-1/2} \left(Y_{T-1+(j+1)/M} - Y_{T-1+j/M}\right) \left(\epsilon_{T-1+(j+1)/M} - \epsilon_{T-1+j/M}\right) \right.$$

$$\left. - a_M^{-1/2} \left(Y_{t+(j+1)/M} - Y_{t+j/M}\right) \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right) \right) \right).$$

Assuming that the noise is uncorrelated with the stock price, and noting by the central limit theorem that

$$\frac{1}{\sqrt{M}} \sum_{j=1}^{M-1} \left(a_M^{-1} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right)^2 - a_M^{-1} \left(\epsilon_{T-1+(j+1)/M} - \epsilon_{T-1+j/M}\right)^2 \right) = O_p(1),$$

it follow that:

$$\frac{1}{T\xi} \sum_{t=1}^{T-1} K \left(\frac{RV_{t,M} - RV_{T,M}}{\xi} \right)$$

$$= \frac{1}{T\xi} \sum_{t=1}^{T-1} K \left(\frac{\frac{1}{\sqrt{M}} \sum_{j=1}^{M-1} \left(a_M^{-1} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right)^2 - a_M^{-1} \left(\epsilon_{T-1+(j+1)/M} - \epsilon_{T-1+j/M}\right)^2 \right)}{\xi a_M^{-1} M^{-1/2}} \right) + o_p(1)$$

$$= O_p\left(a_M^{-1} M^{-1/2}\right).$$

Similarly:

$$RV_{t,M} = \sum_{j=1}^{M-1} \left(Y_{t+(j+1)/M} - Y_{t+j/M}\right)^2 + a_M \sum_{j=1}^{M-1} a_M^{-1} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right)^2$$

$$- a_M^{1/2} \sum_{j=1}^{M-1} 2 \left(Y_{t+(j+1)/M} - Y_{t+j/M}\right) a_M^{-1/2} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right).$$

Then, noting that

$$\frac{1}{2M} \sum_{j=1}^{M-1} a_M^{-1} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right)^2 = O_p(1),$$

$$\frac{a_M^{1/2}}{M^{1/2}} \sum_{j=1}^{M-1} 2 \left(Y_{t+(j+1)/M} - Y_{t+j/M}\right) a_M^{-1/2} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right) = O_p(1),$$

the numerator in (16) can be written as:

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

$$= \frac{1}{T\xi} \sum_{t=1}^{T-1} 1_{\left\{\frac{1}{M} \sum_{j=1}^{M-1} a_M^{-1} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M}\right)^2 \leq u a_M^{-1} M^{-1}\right\}}$$

$$
\times K \left(\frac{\frac{1}{\sqrt{M}} \sum_{j=1}^{M-1} \left(a_M^{-1} \left(\epsilon_{t+(j+1)/M} - \epsilon_{t+j/M} \right)^2 - a_M^{-1} \left(\epsilon_{T-1+(j+1)/M} - \epsilon_{T-1+j/M} \right)^2 \right)}{\xi_1 a_M^{-1} M^{-1/2}} \right) + o_p(1)
$$
$$
= O_p \left(a_M^{-1} M^{-1} \right).
$$

The statement in (21) then follows.

Now, given that for $u > 0$, $F_{IV_{T+1}|IV_T}(u|RV_{T,M}) > 0$ and $T\xi M^{-1} \to 0$, because of the bandwidth conditions in Theorem 1,2 and 3, it follows that:

$$
\begin{aligned}
&\sqrt{T\xi} \left(\widehat{F}_{RV_{T+1,M}|RV_{T,M}}(u|RV_{T,M}) - F_{IV_{T+1}|RV_{T,M}}(u|RV_{T,M}) \right) \\
&= \sqrt{T\xi} \widehat{F}_{RV_{T+1,M}|RV_{T,M}}(u_2|RV_{T,M}) - \sqrt{T\xi} F_{IV_{T+1}|IV_T}(u|RV_{T,M}) \\
&= O_p \left(\sqrt{T\xi M^{-1}} \right) - O \left(\sqrt{T\xi} \right).
\end{aligned}
$$

This establishes (22). Hence, the scaled difference between the confidence intervals based on realized volatility and those based on integrated volatility diverges as $M, T \to \infty$. This is evidentiated by the Monte Carlo findings; in fact, our results show that that, as M grows the empirical size gets close to one.

4.2 Remarks

Remark 1. From Lemma 1, it follows that $b_M = M$ for the case of realized volatility and bipower variation, while $b_M = M^{1/3}$ for $RM_{t,M} = \widehat{RV}_{t,l,M}$, and $b_M = M^{1/2}$ for $RM_{t,M} = \widetilde{RV}_{t,e,M}$ and $RM_{t,M} = \mathcal{K}_{t,H,M}$. Thus, b_M grows with M at different rates across different realized volatility measures. More precisely, b_M grows as fast as M in the case of realized volatility and bipower variation, while it grows at a rate slower than M in the case of microstructure robust realized measures. Hence, for empirical implementation of our results, 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 M and a microstructure robust realized measure. This issue is investigated in the Monte Carlo section.

Remark 2. In general, we do not have a closed form expression for the "true" conditional confidence interval:

$$
F_{IV_{T+1}|IV_T^{(d)}}(u_2|RM_{T,M}^{(d)}) - F_{IV_{T+1}|IV_T^{(d)}}(u_1|RM_{T,M}^{(d)}).
$$

In fact, even if the data generating process for the instantaneous volatility process is known, this does not imply knowledge of the data generating process for the integrated volatility process. For example, if instantaneous volatility is modelled as a square root process, we know that integrated volatility is an ARMA(1,1) process (see, e.g., Meddahi, 2001). However, while we know the autoregressive parameter, we do not know the moving average component of the ARMA process. Additionally, we do not know the marginal distribution of the innovation term in the ARMA process. Of course, we can simulate the instantaneous volatility process, obtain the implied daily volatility process, and then construct the conditional confidence intervals using kernel estimators based on simulated integrated volatility. By keeping the number of simulations large enough, we can obtain conditional confidence intervals arbitrarily close to the true ones. Then, we can test whether the conditional confidence intervals implied by a given model of instantaneous volatility are correctly specified. In practice, simulated confidence intervals are constructed using estimated parameters based on samples of T observations.

Remark 3. From a practical point of view, the asymptotic normality results stated in Theorems 1 and 2 and 3 are useful, as these results facilitate the construction of confidence bands around estimated conditional densities and confidence intervals. The sort of empirical problem for which these results may be useful is the following. Suppose that we want to predict the probability that integrated volatility will take a value between IV_l and IV_u, say, given that we observe the current (and past) values for a chosen realized measure. Then, as b_M and $T \to \infty$, and if b_M grows fast enough relative to T,

$$\Pr\left((IV_l \leq IV_{T+1} \leq IV_u)|IV_T^{(d)} = RM_{T,M}^{(d)}\right)$$

will fall in the interval

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

with probability $1-\alpha$, where $\widehat{V}(l,u)$ is defined in Corollary 1 and $z_{\alpha/2}$ denotes the $\alpha/2$ quantile of a standard normal. Analogous confidence bands can be constructed for conditional densities at different evaluation points.

Remark 4. In empirical work, volatility is often modelled and predicted with ARMA models that are constructed using logs of realized volatility. For example, Andersen, Bollerslev, Diebold and Labys (2001, 2003) use the log of realized volatility for modelling and predicting stock returns and exchange rate volatility. According to these authors, one reason for using logs is that while the distribution of realized volatility is highly skewed to the right, the distribution of logged realized volatility is much closer to normal. It is immediate to see that a Taylor expansion of $\log(RM_{t,M})$ around IV_t gives:

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

where $N_{t,M} = IV_t - RM_{t,M}$. Provided that IV_t is bounded away from zero, under the conditions in Lemma 1, it follows that $\mathrm{E}\left(|\log(RM_{t,M}) - \log(IV_t)|^k\right) = O(b_M^{-k/2})$. Therefore, the statements in the theorems above hold in the case where we are interested in predictive densities and confidence intervals for the log of integrated volatility, conditional on the log of current and past realized volatility measures.

Remark 5. All of our asymptotic results have been obtained under the assumption that the volatility process is short memory. Indeed, this is required as one of the conditions for the measurement error to approach zero at a fast enough rate and for the application of the central limit theorem for kernel estimators. Andersen, Bollerslev, Diebold and Labys (2001, 2003) suggest modelling realized volatility using an ARFIMA model with a differencing coefficient equal to 0.4. Coutin and Renault (2003) show that by allowing for long-memory in the volatility process one can explain some puzzles, such as steep volatility smiles in long-term options and co-movements between implied and realized volatility. In essence, they propose a new model in which long-memory volatility is obtained via fractional integration of a geometrically mixing volatility process (such as square root volatility or log-normal volatility, as in Comte and Renault, 1998). The resulting volatility, and the integrated volatility process is characterized by the same autocovariance structure of an ARFIMA(p, d_0, q) where the differencing parameter, d_0, is the same as that used in the fractional integration of short memory spot volatility. Thus, from e.g. Taqqu (1975), it follows that

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

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

$$\xi^d \sum_{k=0}^{T} \mathrm{E}\left((IV_0 - \mathrm{E}(IV_0))(IV_{0+k} - \mathrm{E}(IV_0))\right) \to 0, \text{ as } T \to \infty. \tag{25}$$

Outside of the case of (functions of) Gaussian processes, Claeskens and Hall's conditions imply, but are not implied by (25). Thus, for ARFIMA processes we need that $\xi^d T^{2a} L(T) \to 0$: the stronger is the memory (i.e. the higher is d_0), the faster the bandwidth should go to zero. For example, for

$d = 1$ and $\xi = O(T^{-1/5})$, we need $d_0 < 1/10$. Now, empirical evidence suggests that d_0 is often between 0.3 and 0.4. For $d_0 = 0.4$, the long-variance condition is satisfied if $\xi \to 0$ at a rate faster than $T^{4/5}$, which implies a rate of convergence for the kernel density estimator slower than $T^{1/10}$. Therefore, unless d_0 is very small, the rate of convergence for the estimator becomes rather slow. It is also necessary to investigate the effect of long-memory on the rate of decay of the moments of the measurement error. Indeed, Lemma 1 requires the volatility process to be strong mixing. Nevertheless, if we weaken Assumption A5 by simply requiring that $\mathrm{E}\left(\left|N_{t,M}^k\right|^k\right) = O(T^{1/2}b_M^{-k/2})$, we can relax the memory requirement on the instantaneous volatility process, though we still need that $\mathrm{E}\left(\left(\sigma_t^2\right)^{2k+\eta}\right) < \infty$, for $\eta > 0$.

5 Monte Carlo Results

In this section, we assess the finite sample behavior of the conditional confidence interval estimator defined in (16) and studied in Theorem 1 and Corollary 2, using the five realized measures outlined in Section 2. Namely, for a variety of different experimental designs we construct:

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

Our objective is thus to assess the empirical level properties of $G_{T,M}(u_1,u_2)$. Data are generated according to the following DGP:

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

where $W_{1,t}$ and $W_{2,t}$ are two independent Brownian motion.

Experiments are carried out by first generating S paths of length $d+1$, with $S > \max(T)$, using a discrete interval $(1/N)$ between successive observations, with $N > \max(M)$, where $\max(T)$ and $\max(M)$ are the largest daily and intradaily sample sizes used in our experiments. These initial data are used to construct a "pseudo true" confidence interval for integrated volatility. Then, Monte Carlo iterations are executed via subsequent generation paths of length T. In each iteration, the first dN observations are kept fixed, and are taken from the first d days of data used in the construction of the pseudo true interval.[5] All data are generated using the Milstein approximation

[5] In this way, the conditioning set, $RM_{T,M}^{(d)}$, is held fixed across all iterations, and is the same as that used in the construction of the pseudo true interval.

scheme.

Simulated data are then sampled at frequency $1/M$, for various values of M, where $M < N$, and samples of T days are constructed. Thus, each simulated sample contains MT observations. Finally, the five realized volatility measures defined in (8), (9), (11), (14) and (15) are constructed, and are in turn used to form $G_{T,M}(u_1, u_2)$.

In the implementation of this experiment, we set $m = 0.05$, $\eta = 1$, $\psi = 2.5$, $\upsilon = 1$.[6] In our basic case (denoted by Case I in Table 1), we simply set $X_{t+j/M} = Y_{t+j/M}$ for $t = 1, \ldots, T$, and $j = 1, \ldots, M$. In Case II (see Table 2), daily data are generated by adding microstructure noise, that is $X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}$, $t = 1, \ldots, T$, and $j = 1, \ldots, M$, where $\epsilon_{t+j/M} \sim i.i.d.$ $\mathrm{N}(0, \nu)$, with $\nu = \{(3/2 * 1152)^{-1}, (3 * 1152)^{-1}, (6 * 1152)^{-1}\}$. In our experiments, this corresponds to integrated volatility noise/signal ratios of approximately 1/1000, 1/2000, and 1/4000. Finally, in Case III (see Table 3) jumps are added by including an $i.i.d.$ $\mathrm{N}(0, 0.64 a_{jump} \widehat{\mu}_{IV_t})$ shock to the process for $Y_{t+j/M}$, where a_{jump} is set equal to $\{3, 1, 0.5\}$, and $\widehat{\mu}_{IV_t}$ is the mean of our pseudo true IV_t values. In this case, it is assumed that jumps arrive randomly with equal probability at any point in time, once within each 5 day interval when $a_{jump} = 3$, once within each 2 day interval when $a_{jump} = 1$, and once within each 1 day interval when $a_{jump} = 0.5$, on average.

We set $S = 3000$, and $d = 1$. In addition, we set the interval $[u1, u2] = [\widehat{\mu}_{IV_t} - \beta\widehat{\sigma}_{IV_t}, \widehat{\mu}_{IV_t} + \beta\widehat{\sigma}_{IV_t}]$ where $\widehat{\mu}_{IV_t}$ is defined above, $\widehat{\sigma}_{IV_t}$ is the standard error of the pseudo true data, and $\beta = \{0.125, 0.25, 0.5\}$. The associated confidence intervals based on these combinations of $[u1, u2]$ are 0.128, 0.181, 0.341. We consider different values of T and M, i.e. $T = \{100, 300, 500\}$ and $M = \{72, 144, 288, 576\}$. For space reasons, we only reports the results for $\beta = 0.125$ and $\beta = 0.25$; also we consider only $T = 100$ and $T = 300$ for the case of no noise and no jumps, and $T = 100$ for the case in which either noise or jumps are present. Complete results are available upon request. All results are based upon 500 Monte Carlo iterations.

In Tables 1-3, rejection frequencies are reported, using two-sided 5% and 10% nominal level tests. The first column contains results for $RV_{t,M}$, the second for $BV_{t,M}$, the third for $\widehat{RV}_{t,l,M}$, the fourth for $\widetilde{RV}_{t,e,M}$, and the fifth for $\mathcal{K}_{t,H,M}$. For the implementation of $\mathcal{K}_{t,H,M}$ we have used 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 their feasible method; in fact, as shown by Barndorff-Nielsen, Hansen, Lunde and Shephard (2006a), this kernel choice gives raise to the smallest asymptotic variance. Results for

[6] Numerous additional Monte Carlo designs were tried, including ones where $\psi = 1.25$ and $\psi = 0.625$, and ones where $W_{1,t}$ and $W_{2,t}$ are correlated, for example. Conclusions drawn from experiments using these alternative parameterizations were qualitatively the same as those reported below. Complete tabulated results are available upon request.

different values of M are reported in different rows of the tables. A number of clear conclusions emerge upon examination of the results.

Turning first to Table 1, where there is neither microstructure noise nor jumps, note that $RV_{t,M}$ and $BV_{t,M}$ perform approximately equally well, although $RV_{t,M}$ performs marginally better in a number of instances, as might be expected. In particular, $RV_{t,M}$ and $BV_{t,M}$ have empirical sizes close to the nominal 5% and 10% levels in various cases, and there is a substantial improvement when both M and T increase. Indeed, in many cases the nominal size is achieved, or very nearly so, a finding which might be viewed as rather surprising given the small values of M and T used in our experiment. Of note is that, roughly speaking, our findings are qualitatively the same, regardless of the width of the confidence interval for which the test statistics are constructed. In particular, the two confidence interval widths in Panels A-B of Table 1 yield similar empirical findings, with marginal improvement as the interval, $[u1, u2]$, increases in width. As expected, $RV_{t,M}$ and $BV_{t,M}$ yield more accurate confidence intervals than the other three robust measures. In particular, note that rejection frequencies at the nominal 10% level for $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$, and $\mathcal{K}_{t,H,M}$ are often 0.20-0.30 when $M = 72$ and 144, whereas rates for $RV_{t,M}$ and $BV_{t,M}$ are generally quite close to 0.10. Indeed, empirical performance of $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$, and $\mathcal{K}_{t,H,M}$ is quite poor for very small values of M (rejection frequencies of 0.50-0.80 are not unusual in such cases), and performance often worsens as T increases, for fixed M. Nevertheless, it should be stressed that the robust measures clearly yield empirical rejection frequencies that improve quite quickly as M increases, for fixed T. Moreover, the performance of $\mathcal{K}_{t,H,M}$ actually improves as T increases, for fixed M, in cases other than when M is very small. Additionally, $\mathcal{K}_{t,H,M}$ performs better than the other microstructure noise robust measures in virtually all cases reported in the table, although the relative difference in performance is clearly shrinking rapidly as M increases. In summary, there is clearly a need for reasonably large values of M when implementing the microstructure robust realized measures in our context.

We now turn to Tables 2, where microstructure noise is added to the simulated efficient price. It is immediate to see that $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $\mathcal{K}_{t,H,M}$ are superior to the non robust realized measures, for large values of M, as expected. For example, consider Panel B in Table 2. The rejection frequencies at the nominal 10% level for $RV_{t,M}$ range from 0.207 up to 1, when $M = 576$, depending upon the signal to noise ratio. On the other hand, the rejection frequencies for $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $\mathcal{K}_{t,H,M}$ range from 0.78-0.267, which indicate a marked improvement when using robust measures, as long as M is large, even though T is only 100. Indeed, when the noise to signal ratio is 1/1000 or 1/2000, the rejection frequency for the non-robust measures is very close to unity in all cases. Of course, for M too small, there is nothing to gain by using the robust measures. Indeed, for $M = 72$, $RV_{t,M}$ rejection frequencies are much closer to the nominal level than $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $\mathcal{K}_{t,H,M}$ rejection frequencies. This is as expected, given Lemma 1. In particular, recall from Remark 1 in Section 4.2 that b_M grows as fast as M in the case of realized volatility and bipower variation, while it grows at a rate slower than M in the case of microstructure noise robust realized measures. We thus noted in Remark 1 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 M and a microstructure robust realized measure. Interestingly, we see in our experiments that the best performing of our robust measures at small values of M (i.e. $\mathcal{K}_{t,H,M}$) outperforms $RV_{t,M}$ in many cases for values of M as small as 144, which suggests that the relative gains associated with using robust measures are achieved very quickly as M increases.

We now turn to Table 3, in which jumps are present. We note that $BV_{t,M}$ outperforms all microstructure noise robust measures, as expected. Note also that within each panel in Table 3, there are three different jump parameterizations reported on. As might be expected, when the number of jumps increases, the relative performance of $BV_{t,M}$ increases. For example, when there is a jump every day, then $BV_{t,M}$ outperforms all other measures, regardless of the value of M. However, when jumps are much less frequent, $RV_{t,M}$ rejection frequencies are sometimes marginally closer to nominal levels.

In summary, the above experiment suggests that asymptotic theory established in Section 3 yields reasonably sharp finite sample distributional approximations, even for small values of T and M, such as $T = 300$ and $M = 576$. Additionally, all realized measures perform as expected, and the robust measures perform as well as might be expected for moderately small values of M (i.e. $M = 576$), and very small values of T (i.e. $T = 100$ daily observations). Finally, in the context of microstructure noise, the trade-off between using robust measures with large values of M versus non-robust measures with small values of M that is predicted by Lemma 1 is clearly apparent in our experimental results.

6 Empirical Illustration: Daily Volatility Predictive Densities for Intel

In this section we construct and examine predictions of the conditional distribution of daily integrated volatility for Intel using two different samples of data, and using the realized measures discussed in Section 2. The rest of this section is broken into three subsections, including: a discussion of the data; a discussion of boundary corrected kernels; and a discussion of our empirical findings.

6.1 Data Description

Data were retrieved from the Trade and Quotation (TAQ) database at the New York Stock Exchange (NYSE), and we base our analysis on two different sample sizes. The first one extends from January 2 to May 27, 1998; the second from January 2 to May 22, 2002. Both sample sizes consist of a total of 100 trading days. The reason of the choice of two different sample periods is to analyze the effect of the decimalization of the tick size (the tick size was reduced from a sixteenth of a dollar to one cent on January 29, 2001). From the original data set, which includes prices recorded for every trade, we extracted 10 second and 5 minute interval data. 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, Andersen, Bollerslev and Lang, 1999 and Ebens, 1999). Hence, our choice of the two frequencies allows us to evaluate the effect of microstructure noise on the estimated predictive densities.

The price figures for each 10 seconds and 5 minutes intervals are determined using the last tick method, which was first proposed by Wasserfallen and Zimmermann (1985). From the calculated series we have obtained 10 second and 5 minute intradaily returns as the difference between successive log prices. A full trading day consists of 2340 (resp. 78) intraday returns calculated over an interval of ten seconds (resp. five minutes).

6.2 Boundary Corrected Kernels

Since variances are by construction positive, the densities that we want to predict will have support on the positive real line. Furthermore, it is well known that conventional kernel functions do not produce consistent estimates when the evaluation points are close to the boundaries of the support.

In the literature, different approaches have been proposed to resolve this problem. We have used the boundary corrected kernel function of Müller (1991), using a locally variable bandwidth. Apart from their optimality properties (in terms of minimizing the integrated mean square error), a nice and convenient feature of boundary corrected kernel functions is that they simplify to conventional ones when the evaluation point is not close to the boundary. For ease of exposition, we will highlight how the method works in the case of univariate densities, but extensions are straightforward.

Consider a density estimator based on the standard quartic kernel:

$$\widehat{f}(x)=\frac{1}{n\xi}\sum_{i=1}^{n}K\left(\frac{x-X_i}{\xi}\right),$$

where

$$K(u)=\frac{15}{16}\left(1-u^2\right)^2 1_{\{|u|\leq 1\}}.$$

Denote $q=\min(x/\xi,1)$. The boundary modified kernel estimator has the form:

$$\widehat{f}_q(x)=\frac{1}{n\xi_q}\sum_{i=1}^{n}K_q\left(\frac{x-X_i}{\xi_q}\right),$$

where

$$K_q(u)=\frac{30\left(1+u^2\right)(q-u)^2}{(1+q)^5}\left(1+7\left(\frac{1-q}{1+q}\right)^2+14\frac{(1-q)\,u}{(1+q)^2}\right)1_{\{-1\leq u\leq q\}}$$

and

$$\xi_q=b(q)\,\xi=(2-q)\,\xi.$$

Notice that $K_1(u)=K(u)$. Hence, the resulting limiting distributions are the same.[7]

6.3 Empirical Results

Using the two series of returns at different frequencies, predictive densities and 10% confidence intervals were calculated for each of the five considered realized measures. The Results are reported for the case of $d=1$ (that is, we have conditioned on the current value of integrated volatility), using the boundary modified quartic kernel function and 1000 evaluation points.

Selected results are presented in Figures 1-4. The graphs reveal some interesting facts. First, the graphs for the realized volatility and bipower variation are quite similar (see Figures 1 and 2).[8] This seems to imply that jumps occur occasionally in the price process, and therefore do not affect a procedure which is based on samples containing a large number of daily observations.

Second, and not surprisingly, the graphs displaying results for two scale, multi scale realized volatility and realized kernel are somewhat similar (see Figure 1, for example). At the 5 minute frequency, the range of the realized kernel estimator is, however, wider than that of the two and multi scale estimator and is closer to that of the non-robust estimators. This disparity disappears at the higher frequency.

[7] The only difference is that, in order to calculate the variance of our conditional density estimator close to the boundary at zero, we will need to compute integrals which depend on q.

[8] This similarity occurs for all values of M and for both subsamples. Therefore, we report results only for one of the two realized measures in some of the figures, for the sake of brevity.

Third, the effect of market microstructure noise emerges clearly in our empirical illustration. In fact, by looking at the range of the densities of realized volatility and bipower variation for the two different frequencies, the distorting impact of microstructure noise is made quite clear. As predicted by theory (see Aït-Sahalia, Mykland and Zhang, 2005), and confirmed by the simulation results in Section 5, when the time interval between successive observations becomes small, then the signal to noise ratio of the data decreases, and realized volatility and bipower variation tend to explode, instead of converging to the increments of quadratic variation. This result is apparent upon inspection of the predictive densities, as the ranges of the densities of the two estimators, estimated with higher frequency data, are considerably wider than the corresponding ones obtained with lower frequency data. Furthermore, the microstructure robust realized volatility measures are clearly more stable, and increasing the frequency at which the data are sampled does not seem to induce any appreciable distortion in density estimators based on these robust measures (see the lower 3 plots in Figure 3). If anything, the range becomes shorter as the frequency increases.

Fourth, tick decimalization has had a marked impact in reducing the effect of market microstructure noise. This can be seen by comparing the change of the range of the densities using realized volatility and bipower variation (when moving from $M = 78$ to $M = 2340$), in the two years considered. For example, comparing the upper 2 plots in Figure 1 with the same in Figure 2, one can see that the range of the density using bipower variation increases sixfold increase in 1998, when moving from $M = 78$ to $M = 2340$. There is only a twofold increase in range when an analogous comparison is made using the 2002 data (see the upper two plots in Figure 3). However, microstructure noise still seems to have an important effect on the estimation of financial volatility.

A final interesting feature that can be observed upon inspection of Figures 1-3 is the multimodality of the densities, which is probably due to volatility clustering effects.

Figure 4 reports various predictive densities and 10% confidence intervals based on the log of integrated volatility, calculated using a standard quartic kernel function with 1000 evaluation points. Similar to results reported in the literature (see, e.g., Andersen, Bollerslev, Diebold and Labys, 2001, 2003) logging our realized volatility measures appears to induce the densities to be closer to "normal" (see the plots in Figure 4), with the same microstructure related distortion effect noted earlier for the non-robust measures.

7 Concluding Remarks

In recent years, numerous volatility-based derivative products have been engineered. This has led to interest in constructing conditional predictive densities and confidence intervals for integrated volatility. In this paper, we establish asymptotic normality for Nadaraya-Watson and local polynomial estimators of conditional confidence intervals and conditional densities, based on the use realized measures. This is accomplished following two key steps. First, given an assumption on the rate of decay to zero of the moments of the measurement error, we provide sufficient conditions on the relative rate of growth of M, number of intraday observations, and T, the number of days, under which estimators based on realized measures are asymptotically equivalent to their unfeasible counterparts based on integrated volatility. Second, we provide primitive conditions, under which realized volatility, power variation and several microstructure robust measures satisfy the assumption on the rate of decay of the measurement error. Our results apply to a general class of *cadlag* volatility processes. The finite sample behavior of the suggested procedures is analyzed via a Monte Carlo experiment, which highlights how the relative performance of five important realized measures varies with the ratio M/T, and with the "size" of the microstructure noise component. Finally, an empirical applications based on New York Stock Exchange data is provided. We find strong evidence of microstructure noise contamination, but we also find that the magnitude of the noise substantially decreased after the decimalization in January 2001.

Appendix

Proof of Theorem 1:

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

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

Now,

$$\begin{aligned}
&\sqrt{T\xi^d}\left(\widehat{F}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(u|RM_{T,M}^{(d)}) - \widehat{F}_{IV_{T+1}|IV_T^{(d)}}(u|RM_{T,M}^{(d)})\right) \\
&= \frac{1}{\widehat{f}_{IV_T^{(d)}}(RM_{T,M}^{(d)})} && (26)\\
&\times\left(\frac{1}{\sqrt{T\xi^d}}\sum_{t=d}^{T-1}\left(1_{\{RM_{t+1,M}\le u\}}\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right) - 1_{\{IV_{t+1}\le u\}}\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right)\right) \\
&+\left(\frac{\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right) - \mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right)}{\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)}\right) \\
&\times\left(\frac{1}{\sqrt{T\xi^d}}\sum_{t=d}^{T-1}\left(1_{\{RM_{t+1,M}\le u\}}\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right) - 1_{\{IV_{t+1}\le u\}}\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right)\right). && (27)
\end{aligned}$$

We begin by considering the term in (26). Given A1-A2, by Theorem 2.22 in Fan and Yao (2005),

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

and, by A4(ii), $f_{IV_T^{(d)}}(RM_{t,M}^{(d)}) > 0$. It suffices to consider the numerator. Now, it is immediate to see that:

$$\begin{aligned}
&\frac{1}{\sqrt{T\xi^d}}\sum_{t=d}^{T-1}\left(1_{\{RM_{t+1,M}\le u\}}\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right) - 1_{\{IV_{t+1}\le u\}}\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right) \\
&= \frac{1}{\sqrt{T\xi^d}}\sum_{t=d}^{T-1}1_{\{IV_{t+1}\le u\}}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right) - \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right) && (28)\\
&+ \frac{1}{\sqrt{T\xi^d}}\sum_{t=d}^{T-1}\left(1_{\{IV_{t+1}\le u\}} - 1_{\{RM_{t+1,M}\le u\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right) && (29)
\end{aligned}$$

$$+\frac{1}{\sqrt{T\xi^d}}\sum_{t=d}^{T-1}\left(\left(1_{\{IV_{t+1}\leq u\}}-1_{\{RM_{t+1,M}\leq u\}}\right)\right.$$

$$\left.\times\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)-\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right)\right). \tag{30}$$

With regard to the term in (28), note that:

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

where $\mathbf{K}_i^{(1)}$ denotes the first derivative of $\mathbf{K}$ with respect to its i-th argument, and $\widetilde{RM}_{t,M}^{(d)}\in\left(RM_{t,M}^{(d)},IV_t^{(d)}\right)$. Let:

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

Now,

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=d}^{T-1}\sum_{i=0}^{d-1}R_{t,i,M}1_{\{IV_{t+1}\leq u\}}N_{t-i,M}\right|&\leq\frac{1}{T}\sum_{t=d}^{T-1}\sum_{i=0}^{d-1}\left|R_{t,i,M}\right|\left|N_{t-i,M}\right|\\&\leq d\sup_{t\leq T}\sup_{i=1,\ldots,d-1}\left|N_{t-i,M}\right|\mathrm{E}\left|R_{t,i,M}\right|\\&+d\sup_{t\leq T}\sup_{i=1,\ldots,d-1}\left|N_{t-i,M}\right|\frac{1}{T}\sum_{t=d}^{T-1}\sum_{i=0}^{d-1}\left(\left|R_{t,i,M}\right|-\mathrm{E}\left|R_{t,i,M}\right|\right)\\&=d\sup_{t\leq T}\sup_{i=1,\ldots,d-1}\left|N_{t-i,M}\right|\left(O(1)+o_p(1)\right),\end{aligned}$$

as, by usual change of variable and integration by part argument, $\mathrm{E}\left|(R_{t,i,M})^k\right|=O\left(\left(\frac{1}{\xi^{d+1}}\right)^{k-1}\right)$, and thus $\mathrm{E}\left|(R_{t,i,M})\right|$ is bounded, and, by the central limit theorem,

$$\frac{1}{T}\sum_{t=d}^{T-1}\sum_{i=0}^{d-1}\left(\left|R_{t,i,M}\right|-\mathrm{E}\left|R_{t,i,M}\right|\right)=O\left(\left(T\xi^{d+1}\right)^{-1/2}\right).$$

Now, given A5,

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

Thus, $\sup_{t\leq T}|N_{t,M}| = O_p\left(T^{\frac{1}{k-1}} b_M^{-1/2}\right)$, and the term in (28) is $O_p\left(T^{\frac{k+1}{2(k-1)}} b_M^{-1/2}\xi^{d/2}\right)$.

In order to treat the term (29), note that:

$$\left|\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{IV_{t+1}\leq u\}} - 1_{\{RM_{t+1,M}\leq u\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right|$$
$$\leq \frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{u-\sup_{t\leq T}|N_{t+1,M}|\leq IV_{t+1}\leq u+\sup_{t\leq T}|N_{t+1,M}|\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right). \qquad (31)$$

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

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

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

$$\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{u-\sup_{t\leq T}|N_{t+1,M}|\leq IV_{t+1}\leq u+\sup_{t\leq T}|N_{t+1,M}|\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)$$
$$\leq \frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{u-c\varepsilon b_M^{-1/2}T^{3/2k}\leq IV_{t+1}\leq u+c\varepsilon b_M^{-1/2}T^{3/2k}\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right). \qquad (32)$$

Hereafter, let $c\varepsilon b_M^{-1/2}T^{\frac{3}{2k}} = d_{T,M}$. Then, using (31) and (32), for all $\omega\in\Omega_{T,M}$:

$$\left|\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{IV_{t+1}\leq u\}} - 1_{\{RM_{t+1,M}\leq u\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right|$$
$$\leq \frac{1}{T\xi^d}\sum_{t=d}^{T-1}1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)$$
$$= \frac{1}{T}\sum_{t=d}^{T-1}\mathrm{E}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\frac{1}{\xi^d}\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right)$$
$$+ \frac{1}{T}\sum_{t=d}^{T-1}\left(\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\frac{1}{\xi^d}\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right.$$
$$\left.-\mathrm{E}\left(\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\frac{1}{\xi^d}\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right)\right)\right)$$
$$= I_{T,M} + II_{T,M}.$$

We begin by considering $I_{T,M}$. As $RM_{T,M}^{(d)}$ is used as the evaluation point, without loss of generality let $RM_{T,M}^{(d)} = \mathbf{x}$, where $\mathbf{x} = (x_1, \ldots, x_d)$. Now, given stationarity, let $\mathbf{y}_0 = (y_{01,} \ldots, y_{0d})$, and let $\mathbf{y} = \frac{\mathbf{y}_0 - \mathbf{x}}{\xi}$. We then have that:

$$\begin{aligned}
&\mathrm{E}\left(\left(1_{\left\{u-d_{T,M} \leq IV_{t+1} \leq u+d_{T,M}\right\}}\right) \frac{1}{\xi^d} \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right) \\
&= \frac{1}{\xi^d} \int_{\mathbb{R}^d} \int_{u-d_{T,M}}^{u+d_{T,M}} \mathbf{K}\left(\frac{\mathbf{y}_0 - \mathbf{x}}{\xi}\right) f\left(\mathbf{y}_0, z\right) \mathrm{d}\mathbf{y}_0 \mathrm{d}z \\
&= \int_{\mathbb{R}^d} \int_{u-d_{T,M}}^{u+d_{T,M}} \mathbf{K}\left(\mathbf{y}\right) f\left(\mathbf{x}+\xi\mathbf{y}, z\right) \mathrm{d}\mathbf{y}\ \mathrm{d}z \\
&= \int_{\mathbb{R}^d} \mathbf{K}\left(\mathbf{y}\right) \mathrm{d}\mathbf{y} \int_{u-d_{T,M}}^{u+d_{T,M}} f\left(\mathbf{x}, z\right)\ \mathrm{d}z (1 + O(\xi)) \\
&= O(d_{T,M}) + O(d_{T,M}\xi).
\end{aligned}$$

Thus, $I_{T,M} = O(d_{T,M})$. Now,

$$\begin{aligned}
&\operatorname{var}\left(II_{T,M}\right) \\
&= \mathrm{E}\left(\frac{1}{T}\sum_{t=d}^{T-1}\left(1_{\left\{u-d_{T,M} \leq IV_{t+1} \leq u+d_{T,M}\right\}}\right) \frac{1}{\xi^{2d}} \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right)^2 + O(d_{T,M}^2).
\end{aligned}$$

Given stationarity,

$$\begin{aligned}
&\mathrm{E}\left(\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\left\{u-d_{T,M} \leq IV_{t+1} \leq u+d_{T,M}\right\}}\right) \frac{1}{\xi^d} \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right)^2 \\
&= \frac{1}{T}\mathrm{E}\left(1_{\left\{u-d_{T,M} \leq IV_1 \leq u+d_{T,M}\right\}} \frac{1}{\xi^{2d}} \mathbf{K}\left(\frac{IV_0^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)^2\right) \\
&+ \frac{2}{T}\sum_{j=1}^{T}\left(\mathrm{E}\left(1_{\left\{u-d_{T,M} \leq IV_1 \leq u+d_{T,M}\right\}} 1_{\left\{u-d_{T,M} \leq IV_{1+j} \leq u+d_{T,M}\right\}}\right.\right. \\
&\left.\left.\times \frac{1}{\xi^{2d}} \mathbf{K}\left(\frac{IV_0^{(d)} - RM_{T,M}^{(d)}}{\xi}\right) \mathbf{K}\left(\frac{IV_{0+j}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\right)\right)
\end{aligned}$$

Now,

$$\begin{aligned}
&\frac{1}{T}\mathrm{E}\left(1_{\left\{u-d_{T,M} \leq IV_1 \leq u+d_{T,M}\right\}} \frac{1}{\xi^{2d}} \mathbf{K}\left(\frac{IV_0^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)^2\right) \\
&= \frac{1}{T\xi^{2d}} \int_{\mathbb{R}^d} \int_{u-d_{T,M}}^{u+d_{T,M}} \mathbf{K}\left(\frac{\mathbf{y}_0 - \mathbf{x}}{\xi}\right)^2 f\left(\mathbf{y}_0, z\right) \mathrm{d}\mathbf{y}_0\ \mathrm{d}z
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T\xi^d}\int_{\mathbb{R}^d}\mathbf{K}(\mathbf{y})^2\,\mathrm{d}\mathbf{y}\int_{u-d_{T,M}}^{u+d_{T,M}} f(\mathbf{x},z)\ \mathrm{d}z(1+O(\xi))\\
&= O\left(T^{-1}\xi^{-d}d_{T,M}\right),
\end{aligned}
$$

and, letting $\mathbf{y} = \frac{\mathbf{y}_0-\mathbf{x}}{\xi}$, $\mathbf{y}^{(j)} = \frac{\mathbf{y}_{0+j}-\mathbf{x}}{\xi}$, it follows that:

$$
\begin{aligned}
&\frac{2}{T}\sum_{j=1}^{T}\Bigg(\mathrm{E}\Bigg(1_{\{u-d_{T,M}\le IV_1\le u+d_{T,M}\}}1_{\{u-d_{T,M}\le IV_{1+j}\le u+d_{T,M}\}}\\
&\times\frac{1}{\xi^{2d}}\mathbf{K}\left(\frac{IV_0^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\mathbf{K}\left(\frac{IV_{0+j}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\Bigg)\Bigg)\\
&= \frac{1}{\xi^{2d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{u-d_{T,M}}^{u+d_{T,M}}\int_{u-d_{T,M}}^{u+d_{T,M}}\mathbf{K}\left(\frac{\mathbf{y}_0-\mathbf{x}}{\xi}\right)\mathbf{K}\left(\frac{\mathbf{y}_{0+j}-\mathbf{x}}{\xi}\right)f(\mathbf{y}_0,\mathbf{y}_{0+j},z,z_j)\,\mathrm{d}\mathbf{y}_0\ \mathrm{d}\mathbf{y}_{0+j}\mathrm{d}z\mathrm{d}z_j\\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{u-d_{T,M}}^{u+d_{T,M}}\int_{u-d_{T,M}}^{u+d_{T,M}}\mathbf{K}(\mathbf{y})\,\mathbf{K}\left(\mathbf{y}^{(j)}\right)f\left(\mathbf{x}+\xi\mathbf{y},\mathbf{x}+\xi\mathbf{y}^{(j)},z,z_j\right)\mathrm{d}\mathbf{y}\ \mathrm{d}\mathbf{y}_0^{(j)}\mathrm{d}z\mathrm{d}z_j\\
&= \left(\int_{\mathbb{R}^d}\mathbf{K}(\mathbf{y})\,\mathrm{d}\mathbf{y}\right)^2\int_{u-d_{T,M}}^{u+d_{T,M}}\int_{u-d_{T,M}}^{u+d_{T,M}} f(\mathbf{x},\mathbf{x},z,z_j)\,\mathrm{d}z\mathrm{d}z_j(1+O(\xi))\\
&= O\left(d_{T,M}^2\right)+O\left(d_{T,M}^2\xi\right)=O\left(d_{T,M}^2\right).
\end{aligned}
$$

Thus, $I_{T,M}+II_{T,M} = O_p(d_{T,M})+O_p\left(T^{-1/2}\xi^{-d/2}d_{T,M}^{1/2}\right)$, and thus the term in (29) is $O_p(T^{1/2}\xi^{d/2}d_{T,M})$. By noting that (30) is of smaller probability order than (28) and (29), and recalling that $d_{T,M} = c\varepsilon b_M^{-1/2}T^{\frac{3}{2k}}$, it follows that (26) is $O_p(T^{\frac{3+k}{2k}}\xi^{d/2}b_M^{-1/2})$. It is immediate to see that (27) is of smaller probability order than (26). The statement in the theorem follows. ■

Proof of Theorem 2:

By Theorem 2.22 in Fan and Yao (2005),

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

Thus, it suffices to show that:

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

Now,

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

$$
= \frac{\sqrt{T\xi_1^d\xi_2}\left(\frac{1}{T\xi_1^d\xi_2}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)-\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\right)\right)}{\widehat{f}_{IV_T^{(d)}}(x|RM_{T,M}^{(d)})}
$$

$$
+\left(\frac{1}{\widehat{f}_{RM_{T,M}^{(d)}}(RM_{T,M}^{(d)})}-\frac{1}{\widehat{f}_{IV_T}(RM_{T,M}^{(d)})}\right) \tag{33}
$$

$$
\left(\frac{1}{\sqrt{T\xi_1^d\xi_2}}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)-\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\right)\right).
$$

The first term on the right hand side of (**??**) can be treated the same way as the term in (28) in the proof of Theorem 1, and therefore it is $O_p\left(T^{\frac{k+1}{2(k-1)}}b_M^{-1/2}\xi_1^{d/2}\xi_2^{1/2}\right)$. Furthermore, the second term on the right hand side of (**??**) is of smaller probability order than the first. The statement in the theorem then follows. ■

Proof of Theorem 3:

(i) As in the proof of Theorem 1, for notational simplicity let $u_1 = 0$ and $u_2 = u$. Now,

$$
\widehat{\boldsymbol{\alpha}}_T(u, IV_T) = \begin{pmatrix} \widehat{\alpha}_{0,T}(u, IV_T) \\ \widehat{\alpha}_{1,T}(u, IV_T) \\ \vdots \\ \widehat{\alpha}_{d,T}(u, IV_T) \end{pmatrix} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\widetilde{\mathbf{y}}(u_1, u_2),
$$

where $\mathbf{X}, \mathbf{W}, \widetilde{\mathbf{y}}(u)$ are defined as $\mathbf{X}_{(M)}, \mathbf{W}_{(M)}, \widetilde{\mathbf{y}}_{(M)}(u)$, except that the realized measure series are replaced with integrated volatility. From Remark 4, in Hall, Wolff and Yao (1999):

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

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

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

Note that:

$$
\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_{T,M}(u, RM_{T,M}^{(d)}) - \widehat{\boldsymbol{\alpha}}_T(u, IV_T) \\
&= \left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}T^{-1}\left(\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\widetilde{\mathbf{y}}_{(M)}(u) - \mathbf{X}'\mathbf{W}\widetilde{\mathbf{y}}(u)\right) \\
&+ \left(\left(T^{-1}\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)^{-1} - \left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\right)T^{-1}\mathbf{X}'\mathbf{W}\widetilde{\mathbf{y}}(u) \\
&+ \left(\left(T^{-1}\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)^{-1} - \left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\right) \\
&\times \left(T^{-1}\mathbf{X}_{(M)}'\mathbf{W}_{(M)}\widetilde{\mathbf{y}}_{(M)}(u) - T^{-1}\mathbf{X}'\mathbf{W}\widetilde{\mathbf{y}}(\mathbf{u})\right).
\end{aligned} \tag{34}
$$

Now:

$$
\begin{aligned}
&T^{-1}\left(\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{y}_{(M)}-\mathbf{X}'\mathbf{W}\mathbf{y}\right)\\
&=\left(\begin{array}{c}
\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM^{(d)}_{t,M}-RM^{(d)}_{T,M}}{\xi}\right)1_{\{RM_{t+1,M}\leq u\}}-\mathbf{K}\left(\frac{IV^{(d)}_{t}-RM^{(d)}_{T,M}}{\xi}\right)1_{\{IV_{t+1}\leq u\}}\right)\\
\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM^{(d)}_{t,M}-RM^{(d)}_{T,M}}{\xi}\right)1_{\{RM_{t+1,M}\leq u\}}\left(RM_{t,M}-RM_{T,M}\right)\right.\\
\left.-\mathbf{K}\left(\frac{IV^{(d)}_{t}-RM^{(d)}_{T,M}}{\xi}\right)K\left(\frac{IV_{t+1}-x}{\xi}\right)1_{\{IV_{t+1}\leq u\}}\left(IV_t-RM_{T,M}\right)\right)\\
\vdots\\
\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM^{(d)}_{t,M}-RM^{(d)}_{T,M}}{\xi}\right)1_{\{RM_{t+1,M}\leq u\}}\left(RM_{t-(d-1),M}-RM_{T-(d-1),M}\right)\right.\\
\left.-\mathbf{K}\left(\frac{IV^{(d)}_{t}-RM^{(d)}_{T,M}}{\xi}\right)1_{\{IV_{t+1}\leq u\}}\left(IV_{t-(d-1)}-RM_{T-(d-1),M}\right)\right)
\end{array}\right)\\
&=O_p(T^{\frac{3}{2k}}b_M^{-1/2}),
\end{aligned}
$$

by the same argument as that used in the proof of Theorem 1.

The $(i,j)-$th element of $\left(T^{-1}\left(\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)-\left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)\right)$, for $1<i,j\leq d$, is given by:

$$
\begin{aligned}
&\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM^{(d)}_{t,M}-RM^{(d)}_{T,M}}{\xi}\right)\left(RM_{t-(j-1),M}-RM_{T-(j-1),M}\right)\left(RM_{t-(i-1),M}-RM_{T-(i-1),M}\right)\right.\\
&\left.-\mathbf{K}\left(\frac{IV^{(d)}_{t}-RM^{(d)}_{T,M}}{\xi_1}\right)\left(IV_{t-(j-1)}-RM_{T-(j-1),M}\right)\left(IV_{t-(i-1)}-RM_{T-(i-1),M}\right)\right)\\
&=O_p(T^{\frac{1}{k-1}}b_M^{-1/2}),
\end{aligned}
$$

by the same argument used in the proof of Theorem 1. So, provided the two matrices are uniformly positive definite,

$$
\left(\left(T^{-1}\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)^{-1}-\left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\right)=O_p(T^{\frac{1}{k-1}}b_M^{-1/2})=o_p(T^{\frac{3}{2k}}b_M^{-1/2}).
$$

Thus, $\sqrt{T\xi^d}\left(\widehat{\boldsymbol{\alpha}}_T(u,IV_T)-\widehat{\boldsymbol{\alpha}}_{T,M}(u,RM^{(d)}_{T,M})\right)=O_p\left(T^{\frac{2k+6}{2k}}\xi^d b_M^{-1}\right)=o_p(1)$.

(ii) Define:

$$
\widehat{\boldsymbol{\beta}}_T(x,IV_T)=\begin{pmatrix}\widehat{\beta}_{0,T}(x,IV_T)\\ \widehat{\beta}_{1,T}(x,IV_T)\\ \vdots\\ \widehat{\beta}_{d,T}(x,IV_T)\end{pmatrix}=\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{y},
$$

where $\mathbf{X},\mathbf{W},\mathbf{y}$ are defined as $\mathbf{X}_{(M)},\mathbf{W}_{(M)},\mathbf{y}_{(M)}$, except that the realized measure series are replaced with integrated volatility.

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

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

Thus, it suffices to show that:

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

This follows by the same argument as that used in (i), simply by replacing $1_{\{RM_{t+1,M}\leq u\}}$ and $1_{\{IV_{t+1}\leq u\}}$ with $K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)$ and $K\left(\frac{IV_{t+1}-x}{\xi_2}\right)$, respectively. ■

Proof of Lemma 1:

(i) We begin by considering the case of zero drift. As a straightforward application of Ito's lemma, note that:

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

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

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

$$\times\left(\int_{t+j_3/M}^{t+(j_3+1)/M}\left(\int_{t+j_3/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\left(\int_{t+j_4/M}^{t+(j_4+1)/M}\left(\int_{t+j_4/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\Bigg].$$

Consider the case in which, $j_1 \neq j_2 \neq j_3 \neq j_4$, and without loss of generality suppose that $j_4 > j_1, j_2, j_3$. Also let $\mathcal{F}_{t+j_4/M} = \sigma\left(\sigma_s^2, W_s, s \leq t + j_4/M\right)$,

$$\begin{aligned}
&\mathrm{E}\left(\left(\sqrt{M}N_{t+1,M}^{(1)}\right)^4\right)\\
&= M^2\sum_{j_1}\sum_{j_2}\sum_{j_3}\sum_{j_4}\mathrm{E}\Big[\sigma^2_{t+j_1/M}\sigma^2_{t+j_2/M}\sigma^2_{t+j_3/M}\sigma^2_{t+j_4/M}\\
&\times\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\\
&\times\left(\int_{t+j_3/M}^{t+(j_3+1)/M}\left(\int_{t+j_3/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)\mathrm{E}\left(\int_{t+j_4/M}^{t+(j_4+1)/M}\left(\int_{t+j_4/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\,\middle|\mathcal{F}_{t+j_4/M}\right)\Bigg] = 0.
\end{aligned}\tag{35}$$

Now, suppose that $j_3 = j_4$, and $j_3 \neq j_2 \neq j_1$. If $j_3 < j_1$ and/or $j_3 < j_2$, then $\mathrm{E}\left(\left(\sqrt{M}N_{t+1,M}^{(1)}\right)^4\right) = 0$, by the same argument as that used in (**??**). Moreover, if $j_3 > j_1, j_2$, then:

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

$$\leq \sum_{j_1}\sum_{j_2}\left(\left(\mathrm{E}\left(\sigma^8_{t+j_1/M}\sigma^8_{t+j_2/M}\right)\right)^{1/2}\left(\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^4\right.\right.\right.$$

$$\left.\left.\left.\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^4\right)\right)^{1/2}\right)^{1/2} \mathrm{E}\left(\sigma^{4(2+r)}_{t+j_3/M}\right)^{1/2r}\sum_{j_3}\alpha^{1/2-1/2r}_{|j3-\max\{j_1,j_2\}|}$$

$$= O(1),$$

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

Now, suppose that $j_1 = j_3$ and $j_2 = j_4$, $j_3 \neq j_4$. Then:

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

$$= M^2\sum_{j_1}\sum_{j_2}\mathrm{E}\left[\sigma^4_{t+j_1/M}\sigma^4_{t+j_2/M}\right.$$

$$\left.\times\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^2\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^2\right]$$

$$\leq M^2\sum_{j_1}\sum_{j_2}\left[\left(\mathrm{E}\left(\sigma^8_{t+j_1/M}\sigma^8_{t+j_2/M}\right)\right)^{1/2}\right.$$

$$\left.\left(\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^4\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^4\right)\right)^{1/2}\right]$$

$$= O(1).$$

Turning now to the case where $j_2 = j_3 = j_4$, note that if $j_1 > j_2$, then $\mathrm{E}\left(\left(\sqrt{M}N^{(1)}_{t+1,M}\right)^4\right) = 0$, by the same argument as that used in (**??**). Moreover, if $j_1 < j_2$, then:

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

$$= M^2\sum_{j_1}\sum_{j_2}\mathrm{E}\left[\sigma^2_{t+j_1/M}\sigma^6_{t+j_2/M}\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right.$$

$$\left.\left(\mathrm{E}\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^3|\mathcal{F}_{t+j_2.M}\right)\right]$$

$$= 0.$$

Finally, let $j_1 = j_2 = j_3 = j_4$. Then:

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

$$= M^2\sum_{j_1}\mathrm{E}\left[\sigma^8_{t+j_1/M}\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^4\right]$$

$$
\begin{aligned}
&\leq M^2 \sum_{j_1} \left(\mathrm{E}\left[\left(\sigma^{16}_{t+j_1/M} \right) \right] \right)^{1/2} \left(\mathrm{E}\left(\int_{t+j_1/M}^{t+(j_1+1)/M} \left(\int_{t+j_1/M}^{s} \mathrm{d}W_u \right) \mathrm{d}W_s \right)^8 \right)^{1/2} \\
&= O(1) M^2 O_p(M^{-4}) = O(M^{-2}) = o(1).
\end{aligned}
$$

Because of the Hölder continuity of a diffusion, $\mathrm{E}\left(\left(\sqrt{M} N^{(i)}_{t+1,M} \right)^4 \right)$ cannot be of larger order of magnitude than $\mathrm{E}\left(\left(\sqrt{M} N^{(1)}_{t+1,M} \right)^4 \right)$, for $i = 2, 3, 4$.

Considering now the drift term, note that its contribution to the measurement error on an interval of length $1/M$ is given by:

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

which is of smaller order than $\sqrt{M} \sum_{j=0}^{M-1} \left(\int_{t+j/M}^{t+(j+1)/M} \left(\int_{t+j/M}^{s} \sigma_u \mathrm{d}W_u \right) \sigma_s \mathrm{d}W_s \right)$.

(ii) We consider the case of $N_{t,M} = TV_{t,M} - IV_t$. By Barndorff-Nielsen, Shephard and Winkel (2006), we can ignore the contribution of the jump component. In the case of $N_{t,M} = BV_{t,M} - IV_t$, one can ignore the contribution of the pure jump component, but one must take into account the cross term. This is done in Lemma 5 in Corradi and Distaso (2006), and the same argument applies in the present context. Let $\Delta W_{t+(j+1)/M} = W_{t+(j+1)/M} - W_{t+j/M}$ and let

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

Further, let $\mathrm{E}_{t+j/M}$ be the expectation conditional on $\sigma\left(X_{t+i/M},\ 0 < t + i/M \leq t + j/M \right)$. Now,

$$
\begin{aligned}
\sqrt{M} N_{M,t} &= \sqrt{M} \sum_{j=1}^{M-3} \left(TV_{t+(j+1)/M}(X) - \int_{t+j/M}^{t+(j+1)/M} \sigma^2_s \mathrm{d}s \right) \\
&= \sqrt{M} \sum_{j=1}^{M-3} \left(TV_{t+(j+1)/M}(X) - \mathrm{E}_{j/M}\left(TV_{t+(j+1)/M}(X) \right) \right) \\
&+ \sqrt{M} \sum_{j=1}^{M-3} \left(\mathrm{E}_{t+j/M}\left(TV_{t+(j+1)/M}(X) \right) - \frac{1}{M} \sigma^2_{t+j/M} \mathrm{E}_{t+j/M}\left(TV_{t+(j+1)/M}(W) \right) \right) \\
&+ \sqrt{M} \sum_{j=1}^{M-3} \left(\frac{1}{M} \sigma^2_{t+j/M} \mathrm{E}_{t+j/M}\left(TV_{t+(j+1)/M}(W) \right) - \int_{t+j/M}^{t+(j+1)/M} \sigma^2_s \mathrm{d}s \right)
\end{aligned}
$$

$$= I + II + III.$$

From Theorem 5.1 and Lemma 5.2 in Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006), it follows that II and III are $o_p(1)$, and that

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

Again, consider the case of $k = 4$. Suppose that $j_1 \neq j_2 \neq j_3 \neq j_4$, and without loss of generality suppose that $j_4 > j_1, j_2, j_3$. Also let $\mathcal{F}_{t+(j_4-2)/M} = \sigma\left(\sigma^2_s, W_s, s \leq t + j_4/M\right)$. Then:

$$\begin{aligned}
&\mathrm{E}\left(\left(\sqrt{M} N_{t+1,M} \right)^4 \right) \\
&= M^2 \sum_{j_1=3}^{M-3} \sum_{j_2=3}^{M-3} \sum_{j_3=3}^{M-3} \sum_{j_4=3}^{M-3} \mathrm{E}\left[\sigma^2_{t+j_1/M} \sigma^2_{t+j_2/M} \sigma^2_{t+j_3/M} \sigma^2_{t+j_4/M} \right. \\
&\times \left(TV_{t+(j_1+1)/M}(W) - \frac{1}{M} \right) \left(TV_{t+(j_2+1)/M}(W) - \frac{1}{M} \right) \\
&\times \left. \left(TV_{t+(j_3+1)/M}(W) - \frac{1}{M} \right) \mathrm{E}\left(\left(TV_{t+(j_4+1)/M}(W) - \frac{1}{M} \right) |\mathcal{F}_{t+(j_4-2)/M} \right) \right] = 0.
\end{aligned}$$

Finally, by the same argument as that used in part (i) note that: (a) if $j_3 = j_4$ and $j_3 \neq j_2 \neq j_1$, with $j_3 > j_1, j_2$, or if (b) $j_1 = j_2$ and $j_3 = j_4$, with $j_1 \neq j_3$, then $\mathrm{E}\left(\left(\sqrt{M} N^{(1)}_{t+1,M} \right)^4 \right) = O(1)$. In all the other cases, $\mathrm{E}\left(\left(\sqrt{M} N_{t+1,M} \right)^4 \right) = o(1)$ or $\mathrm{E}\left(\left(\sqrt{M} N_{t+1,M} \right)^4 \right) = 0$.

(iii) We can expand the expectation argument as follows:

$$\begin{aligned}
&\mathrm{E}\left(\left(\widehat{RV}_{t,l,M} - IV_t \right)^k \right) \\
&= \mathrm{E}\left(\left(\left(\frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left(X_{t+(jB+b)/M} - X_{t+((j-1)B+b)/M} \right)^2 \right. \right. \right. \\
&\left. \left. \left. - \frac{l}{M} \sum_{j=1}^{M} \left(X_{t+j/M} - X_{t+((j-1)/M)} \right)^2 \right) - IV_t \right)^k \right) \\
&\approx \mathrm{E}\left(\left(\frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left(Y_{t+(jB+b)/M} - Y_{t+((j-1)B+b)/M} \right)^2 - IV_t \right)^k \right) && (36) \\
&+ \mathrm{E}\left(\left(\frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left(Y_{t+(jB+b)/M} - Y_{t+((j-1)B+b)/M} \right) \left(\epsilon_{t+(jB+b)/M} - \epsilon_{t+((j-1)B+b)/M} \right) \right)^k \right) && (37) \\
&+ \mathrm{E}\left(\left(\frac{1}{B} \sum_{b=1}^{B} \sum_{j=1}^{l-1} \left(\left(\epsilon_{t+(jB+b)/M} - \epsilon_{t+((j-1)B+b)/M} \right)^2 - 2\nu \right) \right)^k \right) && (38)
\end{aligned}$$

$$+\,\mathrm{E}\left(\left(\frac{l}{M}\sum_{j=1}^{M}\left(\left(\epsilon_{t+j/M}-\epsilon_{t+(j-1)/M}\right)^2-2\nu\right)\right)^k\right), \tag{39}$$

where with $\approx$ we mean "of the same order of magnitude".

Provided that $\mathrm{E}(\epsilon_{t+j/M}^{2k})<\infty$, the term in (39) is $O(l^k/M^{k/2})$, when the microstructure noise is *i.i.d.* For the geometrically mixing microstructure error case, (39) is still $O(l^k/M^{k/2})$, provided that $\mathrm{E}(\epsilon_{t+j/M}^{2k+\eta})<\infty$, with $\eta>0$. This follows by the same argument as that used by Yoshihara (1975, Lemma 1), for the case of $k=2$.

Also, the term in (38) is $O(l^{k/2}/B^{k/2})$. Thus, (39) and (38) are $O(b_M^{-k/2})$, with $b_M=M^{1/3}$, provided that $B=M^{2/3}$ and $l=M^{1/3}$. Turning to (37), note that given the independence between the noise and the price, it's of order $O(B^{-k/2})$.

We are left with (36). Note that:

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

$$\approx\mathrm{E}\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(Y_{t+(jB+b)/M}-Y_{t+((j-1)B+b)/M}\right)^2-\sum_{j=1}^{M}\left(Y_{t+j/M}-Y_{t+(j-1)/M}\right)^2\right)^k\right) \tag{40}$$

$$+\,\mathrm{E}\left(\left(\sum_{j=1}^{M}\left(Y_{t+j/M}-Y_{t+(j-1)/M}\right)^2-IV_t\right)^k\right). \tag{41}$$

From the proof of part (i), it follows that the term in (41) is $O(M^{-k/2})$. With regard to (40), from the proof of Theorem 2 in Zhang, Mykland and Aït-Sahalia (2005), note that:

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

where the last term captures the end effects. Without loss of generality assume the drift is zero. Note that,

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

$$
\begin{aligned}
&= \sum_{j=1}^{M-1} \sigma_{t+j/M} \int_{t+j/M}^{t+(j+1)/M} \mathrm{d}W_s \sum_{i=1}^{B\wedge j} \left(1-\frac{j}{B}\right) \sigma_{t+(j-i)/M} \int_{t+(j-i)/M}^{t+(j-i+1)/M} \mathrm{d}W_u \\
&+ \sum_{j=1}^{M-1} \sigma_{t+j/M} \int_{t+j/M}^{t+(j+1)/M} \mathrm{d}W_s \sum_{i=1}^{B\wedge j} \left(1-\frac{j}{B}\right) \int_{t+(j-i)/M}^{t+(j-i+1)/M} \left(\sigma_u - \sigma_{t+(j-i)/M}\right) \mathrm{d}W_u \\
&+ \sum_{j=1}^{M-1} \int_{t+j/M}^{t+(j+1)/M} \left(\sigma_{t+j/M} - \sigma_s\right) \mathrm{d}W_s \sum_{i=1}^{B\wedge j} \left(1-\frac{j}{B}\right) \int_{t+(j-i)/M}^{t+(j-i+1)/M} \sigma_{t+(j-i)/M} \mathrm{d}W_u \\
&+ \sum_{j=1}^{M-1} \int_{t+j/M}^{t+(j+1)/M} \left(\sigma_{t+j/M} - \sigma_s\right) \mathrm{d}W_s \sum_{i=1}^{B\wedge j} \left(1-\frac{j}{B}\right) \int_{t+(j-i)/M}^{t+(j-i+1)/M} \left(\sigma_u - \sigma_{t+(j-i)/M}\right) \mathrm{d}W_u.
\end{aligned}
$$

Moreover, note that

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

Thus, recalling that $l = M^{1/3}$,

$$
\begin{aligned}
&\mathrm{E}\left(l^2 \left(\sum_{j=B+1}^{M-1} \left(Y_{t+(j+1)/M} - Y_{t+j/M}\right) \sum_{i=1}^{B\wedge j} \left(1-\frac{j}{B}\right) \left(Y_{t+(j-i+1)/M} - Y_{t+(j-i)/M}\right)\right)^4\right) \\
&= O(1),
\end{aligned}
$$

by the same argument as that used in Part (i). The statement then follows.

(iv) Recalling that $\sum_{i=1}^{e} \frac{a_i}{i} = 0$, along the lines of Zhang (2006), it follows that:

$$
\begin{aligned}
&\widetilde{RV}_{t,e,M} - IV_t \\
&= \left(\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \left(Y_{t+(j+i)/M} - Y_{t+j/M}\right)^2 - IV_t\right) && (42) \\
&- 2\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \epsilon_{t+(j+i)/M}\epsilon_{t+j/M} && (43) \\
&+ 2\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \left(Y_{t+(j+i)/M} - Y_{t+j/M}\right)\left(\epsilon_{t+(j+i)/M} - \epsilon_{t+j/M}\right) && (44) \\
&- \left(\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{i} \left(\epsilon_{t+j/M}^2 - \nu\right)\right) && (45) \\
&- \left(\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=M-i}^{M} \left(\epsilon_{t+j/M}^2 - \nu\right)\right) && (46) \\
&+ 2\left(\widehat{\nu}_{t,M} - \nu\right),
\end{aligned}
$$

where $\widehat{\nu}_{t,M}$ is defined in (12) and $\mathrm{E}\left(\epsilon^2_{t+j/M}\right) = \nu$. Hence, $\mathrm{E}\left((\widehat{\nu}_{t,M} - \nu)^k\right) = O(M^{-k/2})$. Note also that $a_i \simeq i^2/e^3$. Therefore,

$$\begin{aligned}
&\mathrm{E}\left(\left(\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{i} \left(\epsilon^2_{t+j/M} - \nu\right)\right)^k\right) \\
&\approx \left(\sum_{i=1}^{e} \frac{i^{3/2}}{e^3}\right)^k \mathrm{E}\left(\left(\frac{1}{\sqrt{i}} \sum_{j=1}^{i} \left(\epsilon_{t+j/M} - \nu\right)\right)^k\right) = O(e^{-k/2}),
\end{aligned}$$

so that the expectations of the k-th moments of (45) and (46) are $O(e^{-k/2})$.

Now, with regard to the term in (43), note that:

$$\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \epsilon_{t+(j+i)/M} \epsilon_{t+j/M} \simeq \frac{1}{e^2} \sum_{i=1}^{e} \sum_{j=1}^{M-1} \epsilon_{t+(j+i)/M} \epsilon_{t+j/M}.$$

Hence, when the microstructure noise is *i.i.d.*

$$\mathrm{E}\left(\left(\frac{1}{e^2} \sum_{i=1}^{e} \sum_{j=1}^{M-1} \epsilon_{t+(j+i)/M} \epsilon_{t+j/M}\right)^2\right) = O\left(M/e^3\right) = O(b_M^{-1}),$$

and

$$\mathrm{E}\left(\left(\frac{1}{e^2} \sum_{i=1}^{e} \sum_{j=1}^{M-1} \epsilon_{t+(j+i)/M} \epsilon_{t+j/M}\right)^4\right) = O\left(M^2/e^6\right) = O(b_M^{-2}).$$

In general,

$$\mathrm{E}\left(\left(\frac{1}{e^2} \sum_{i=1}^{e} \sum_{j=1}^{M-1} \epsilon_{t+(j+i)/M} \epsilon_{t+j/M}\right)^k\right) = O\left(M^{k/2}/e^{3k/2}\right) = O(b_M^{-k/2}),$$

for $b_M = M^{1/2}$ and $e = M^{1/2}$.

The same rate holds in the case of geometrically mixing errors, provided that $\mathrm{E}(\epsilon^{2k+\eta}_{t+j/M}) < \infty$, with $\eta > 0$, by the same argument as that used in Yoshihara (1975, Lemma 1).

Turning now to (44), note that given independence between the error and price, it's of order $O\left(e^{-k/2}\right)$.

With regard to the term in (42), the least favorable case (i.e. the case where the rate of convergence to integrated volatility is slowest) occurs when $i = e$, and in that case, by the same argument as the one used in the proof of part (i):

$$\mathrm{E}\left(\left(\sum_{i=1}^{e} \frac{a_i}{i} \sum_{j=1}^{M-i} \left(Y_{t+(j+i)/M} - Y_{t+j/M}\right)^2 - IV_t\right)^k\right) \tag{47}$$

$$= O\left(e^{-k/2}\right) = O\left(M^{-k/4}\right) = O\left(b_M^{-k/2}\right),$$

for $e = M^{1/2}$ and $b_M = M^{1/2}$.

(v) The case of $k = 2$ follows from Theorem 2 and Proposition 3 in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a). We now consider the case of $k = 4$. We begin by considering the contribution of the pure noise component to the 4th moment.

Following Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a), we can write:

$$\begin{aligned}\gamma_{t,h}^{\epsilon} &= 2\sum_{i=H+1}^{M-H-1} \epsilon_{t+i/M}\epsilon_{t+(i-h)/M} - \sum_{i=H+1}^{M-H-1} \epsilon_{t+i/M}\epsilon_{t+(i-h-1)/M} - \sum_{i=H+1}^{M-H-1} \epsilon_{t+i/M}\epsilon_{t+(i-h+1)/M} + G_{t,h}^{\epsilon} \\ &= \widetilde{\gamma}_{t,h}^{\epsilon} + G_{t,h}^{\epsilon}\end{aligned}$$

and

$$\begin{aligned}\gamma_{t,-h}^{\epsilon} &= 2\sum_{i=H+1}^{M-H-1} \epsilon_{t+i/M}\epsilon_{t+(i-h)/M} - \sum_{i=H+1}^{M-H-1} \epsilon_{t+i/M}\epsilon_{t+(i-h-1)/M} - \sum_{i=H+1}^{M-H-1} \epsilon_{t+i/M}\epsilon_{t+(i-h+1)/M} + G_{t,-h}^{\epsilon} \\ &= \widetilde{\gamma}_{t,-h}^{\epsilon} + G_{-t,h}^{\epsilon},\end{aligned}$$

where $G_{t,h}^{\epsilon} - G_{t,-h}^{\epsilon} = O_p(1)$, with G_h^{ϵ} and G_{-h}^{ϵ} capturing the effects of the first and last observations. Now, consider the fourth moment of the pure noise component. Namely, consider:

$$\begin{aligned}&\mathrm{E}\left(\sum_{h=2}^{H} \kappa\left(\frac{h-1}{H}\right)\left(\widetilde{\gamma}_{t,h}^{\epsilon} + \widetilde{\gamma}_{t,-h}^{\epsilon}\right)\right)^4 \\ &= 16\left(\mathbf{u}' \otimes \mathbf{u}'\right) E(\mathbf{\Gamma}_t \otimes \mathbf{\Gamma}_t)\left(\mathbf{u} \otimes \mathbf{u}\right),\end{aligned}$$

where $\mathbf{u}' = \begin{pmatrix} \kappa(1/h) & \kappa(2/h) & . & . & . & \kappa((H-1)/h) \end{pmatrix}$, $\mathbf{\Gamma}_t = \{\widetilde{\gamma}_{t,i}^{\epsilon}\widetilde{\gamma}_{t,j}^{\epsilon}\}$ for $i, j = 2, ..., H$, and $\mathbf{\Gamma}_t \otimes \mathbf{\Gamma}_t$ is $(H-1)^2 \times (H-1)^2$. Hereafter, for notational simplicity, and as the noise is identically distributed, we suppress the subscript t. Thus

$$\begin{aligned}&\mathrm{E}(\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \\ &= \mathrm{E}\begin{pmatrix} \widetilde{\gamma}_2^{\epsilon 2}\mathbf{\Gamma} & \widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_3^{\epsilon}\mathbf{\Gamma} & \widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_4^{\epsilon}\mathbf{\Gamma} & \cdots & \widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_{H-1}^{\epsilon}\mathbf{\Gamma} & \widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_H^{\epsilon}\mathbf{\Gamma} \\ \widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_3^{\epsilon}\mathbf{\Gamma} & \widetilde{\gamma}_3^{\epsilon 2}\mathbf{\Gamma} & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \widetilde{\gamma}_4^{\epsilon 2}\mathbf{\Gamma} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_{H-1}^{\epsilon}\mathbf{\Gamma} & \vdots & \vdots & \ddots & \vdots & \widetilde{\gamma}_{H-1}^{\epsilon}\widetilde{\gamma}_H^{\epsilon}\mathbf{\Gamma} \\ \widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_H^{\epsilon}\mathbf{\Gamma} & \widetilde{\gamma}_3^{\epsilon}\widetilde{\gamma}_H^{\epsilon}\mathbf{\Gamma} & \widetilde{\gamma}_3^{\epsilon}\widetilde{\gamma}_H^{\epsilon}\mathbf{\Gamma} & \cdots & \widetilde{\gamma}_{H-1}^{\epsilon}\widetilde{\gamma}_H^{\epsilon}\mathbf{\Gamma} & \widetilde{\gamma}_H^{\epsilon 2}\mathbf{\Gamma} \end{pmatrix}.\end{aligned}$$

Just as in the proof of Theorem 2 in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a), a necessary condition for the fourth moment of the measurement error to vanish is that, except for initial and end effects, the elements in each row of $\mathrm{E}(\mathbf{\Gamma}\otimes\mathbf{\Gamma})$ sum to zero. While this is the case for the rows of $E(\mathbf{\Gamma})$, and so for the second moments, it is no the case for the rows of $\mathrm{E}(\mathbf{\Gamma}\otimes\mathbf{\Gamma})$.

Straightforward, but tedious calculations give the following.

First, $\mathrm{E}\left(\widetilde{\gamma}_i^{\epsilon}\widetilde{\gamma}_j^{\epsilon}\widetilde{\gamma}_k^{\epsilon}\widetilde{\gamma}_l^{\epsilon}\right)=0$ for $i\neq j\neq k\neq l$. Also, for all $h\geq 2$,

$$\mathrm{E}(\widetilde{\gamma}_i^{\epsilon 4})=72\left((M-2H)\left(\mathrm{E}(\epsilon_{j/M}^4)\right)^2+(M-2H)^2\left(\mathrm{E}(\epsilon_{j/M}^2)\right)^4\right),$$

$$\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon}\widetilde{\gamma}_{h+1}^{\epsilon}\right)=8(M-2H)\left(\mathrm{E}(\epsilon_{j/M}^4)\right)^2+52\left((M-2H)^2\left(\mathrm{E}(\epsilon_{j/M}^2)\right)^4\right),$$

and

$$\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon}\widetilde{\gamma}_{h+2}^{\epsilon}\right)=(M-2H)\left(\mathrm{E}(\epsilon_{j/M}^4)\right)^2+36\left((M-2H)^2\left(\mathrm{E}(\epsilon_{j/M}^2)\right)^4\right);$$

and for $s>2$,

$$\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon 2}\widetilde{\gamma}_{h+s}^{\epsilon 2}\right)=36\left((M-2H)^2\left(E(\epsilon_{j/M}^2)\right)^4\right).$$

Also,

$$\begin{aligned}&\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon 3}\widetilde{\gamma}_{h+1}^{\epsilon}\right)\\&=-10(M-2H)\left(\mathrm{E}(\epsilon_{j/M}^4)\right)^2-12\left((M-2H)^2\left(\mathrm{E}(\epsilon_{j/M}^2)\right)^4\right),\end{aligned}$$

$$\begin{aligned}&\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon 3}\widetilde{\gamma}_{h+2}^{\epsilon}\right)\\&=-1(M-2H)\left(\mathrm{E}(\epsilon_{j/M}^4)\right)^2+4\left((M-2H)^2\left(\mathrm{E}(\epsilon_{j/M}^2)\right)^4\right),\end{aligned}$$

and $\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon 3}\widetilde{\gamma}_{h+s}^{\epsilon}\right)=0$ for $s>2$. Finally,

$$\begin{aligned}&\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon 2}\widetilde{\gamma}_{h+1}^{\epsilon}\widetilde{\gamma}_{h+2}^{\epsilon}\right)\\&=-2(M-2H)\left(\mathrm{E}(\epsilon_{j/M}^4)\right)^2-16\left((M-2H)^2\left(\mathrm{E}(\epsilon_{j/M}^2)\right)^4\right)\end{aligned}$$

and $\mathrm{E}\left(\widetilde{\gamma}_h^{\epsilon 2}\widetilde{\gamma}_{h+s}^{\epsilon}\widetilde{\gamma}_{h+s'}^{\epsilon}\right)=0$ if $\max\{s,s'\}>2$.

Consider the $((H-1)i+j)$th row of $\mathrm{E}(\mathbf{\Gamma}\otimes\mathbf{\Gamma})$, with $i=0,\dots,H-2$ and $j=1,\dots,H-1$. Its

$(H-1)^2$ elements are given by:

$$
\mathrm{E}\left(\begin{array}{c}
\widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_{2+i}^{\epsilon}\left\{\begin{array}{c}\widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\\ \widetilde{\gamma}_3^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\\ \vdots \\ \widetilde{\gamma}_H^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\end{array}\right.\\
\widetilde{\gamma}_3^{\epsilon}\widetilde{\gamma}_{2+i}^{\epsilon}\left\{\begin{array}{c}\widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\\ \widetilde{\gamma}_3^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\\ \vdots \\ \widetilde{\gamma}_H^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\end{array}\right.\\
\vdots\\
\widetilde{\gamma}_H^{\epsilon}\widetilde{\gamma}_{2+i}^{\epsilon}\left\{\begin{array}{c}\widetilde{\gamma}_2^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\\ \widetilde{\gamma}_3^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\\ \cdot \\ \widetilde{\gamma}_H^{\epsilon}\widetilde{\gamma}_{j+1}^{\epsilon}\end{array}\right.
\end{array}\right).
$$

Given the expressions for the moments computed above, it follows immediately that the elements do not sum up to zero, but instead diverge at rate M^2. The statement in (v) then follows. ■

References

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

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

Aït-Sahalia, Y., J. Fan and H. Peng (2006). Nonparametric Transition-Based Tests for Diffusions. Manuscript, Princeton University.

Aït-Sahalia, Y., and L. Mancini (2006). Out of Sample Forecasts of Quadratic Variation, Manuscript, Princeton University.

Andersen, T.G., T. Bollerslev, P. F. Christoffersen and F.X. Diebold (2005). Practical Volatility and Correlation Modeling for Financial Market Risk Management. To appear 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 (2007). Assessing Market Microstructure Effects via Realized Volatility Measures, with an Application to the Dow Jones Industrial Average Stocks. Working Paper. Tanaka Business School-Imperial College.

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

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

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2006a). Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. Manuscript, University of Oxford.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2006b), Subsampling Realized Kernels. Manuscript, University of Oxford.

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.

Bollerslev, T. and H. Zhou (2002). Estimating Stochastic Volatility Diffusion Using Conditional Moments of integrated volatility. *Journal of Econometrics*, 109, 33-65.

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

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

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

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

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

Corradi, V., W. Distaso and N.R. Swanson (2006). Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures. *Journal of Econometrics,* forthcoming.

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

Ebens, H. (1999). Realized Stock Volatility. Working Paper, Johns Hopkins University.

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

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

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

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

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

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

Kalnina, I. and O. Linton (2007). Estimating Quadratic Variation Consistently in the Presence of Correlated Measurement Error. Working Paper, LSE.

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

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

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

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

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

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

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

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

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

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

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

Yoshihara, K.I. (1975). Weak Convergence of Multidimensional Empirical Processes for Strong Mixing Sequences of Stochastic Vectors. *Zeitschrift fur Wahrscheinlichkitstheorie und Verwandte Gebiete,*, 33, 133-137.

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

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

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

Table 1: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case I: No Microstructure Noise or Jumps in DGP

M	$RV_{t,M}$	$BV_{t,M}$	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,e,M}$	$\mathcal{K}_{t,H,M}$
Panel A: Interval $= \widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$					
Sample Size = 100 Daily Realized Measure Observations					
Nominal Size = 5%					
72	0.096	0.106	0.206	0.200	0.130
144	0.126	0.110	0.223	0.163	0.143
288	0.096	0.083	0.153	0.126	0.133
576	0.080	0.063	0.140	0.080	0.113
Nominal Size = 10%					
72	0.130	0.133	0.243	0.216	0.170
144	0.156	0.140	0.250	0.193	0.166
288	0.120	0.110	0.180	0.166	0.166
576	0.106	0.096	0.153	0.106	0.156
Sample Size = 300 Daily Realized Measure Observations					
Nominal Size = 5%					
72	0.116	0.136	0.430	0.380	0.240
144	0.120	0.123	0.300	0.283	0.173
288	0.113	0.083	0.200	0.163	0.130
576	0.066	0.063	0.146	0.190	0.100
Nominal Size = 10%					
72	0.156	0.160	0.483	0.463	0.293
144	0.173	0.160	0.360	0.330	0.226
288	0.143	0.096	0.236	0.230	0.153
576	0.096	0.090	0.190	0.216	0.140
Panel B: Interval $= \widehat{\mu}_{IV_t} + 0.250\widehat{\sigma}_{IV_t}$					
Sample Size = 100 Daily Realized Measure Observations					
Nominal Size = 5%					
72	0.110	0.113	0.440	0.370	0.223
144	0.106	0.103	0.213	0.273	0.160
288	0.106	0.103	0.216	0.196	0.136
576	0.076	0.113	0.176	0.150	0.093
Nominal Size = 10%					
72	0.136	0.156	0.510	0.413	0.273
144	0.153	0.143	0.266	0.363	0.213
288	0.146	0.146	0.266	0.256	0.170
576	0.130	0.163	0.223	0.193	0.110
Sample Size = 300 Daily Realized Measure Observations					
Nominal Size = 5%					
72	0.090	0.103	0.676	0.480	0.273
144	0.070	0.076	0.426	0.363	0.123
288	0.073	0.083	0.220	0.190	0.063
576	0.063	0.083	0.163	0.190	0.070
Nominal Size = 10%					
72	0.150	0.160	0.720	0.530	0.363
144	0.130	0.136	0.506	0.436	0.166
288	0.120	0.136	0.276	0.263	0.106
576	0.086	0.126	0.210	0.230	0.103

* Notes: Entries denote rejection frequencies based on the construction of $G_{T,M}(u_1, u_2)$. In particular, values of $G_{T,M}(u_1, u_2)$ are compared with 5% and 10% nominal size critical values of the standard normal distribution. We use "pseudo true" IV values in place of actual IV values when constructing $G_{T,M}(u_1, u_2)$, as discussed above. Results are reported for various realized measures (including $RV_{t,M}$, $BV_{t,M}$, $\widehat{RV}_{t,l,M}$, $\widetilde{RV}_{t,e,M}$ and $\mathcal{K}_{t,H,M}$) for various different values of M, and two various daily sample sizes. The interval over which the statistics are cvalculated is $[u1, u2] = [\widehat{\mu}_{IV_t} - \beta\widehat{\sigma}_{IV_t}, \widehat{\mu}_{IV_t} + \beta\widehat{\sigma}_{IV_t}]$ where $\widehat{\mu}_{IV_t}$ and $\widehat{\sigma}_{IV_t}$ are the mean and standard error of the pseudo true data, and $\beta = \{0.125and0.250\}$. See Section 5 for further details.

Table 2: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case II: Microstructure Noise in DGP

M	$RV_{t,M}$	$BV_{t,M}$	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,e,M}$	$\mathcal{K}_{t,H,M}$
Panel A: Interval = $\widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$					
Noise/Signal Ratio = 1/1000					
Nominal Size = 5%					
72	0.263	0.233	0.483	0.409	0.296
144	0.363	0.284	0.424	0.325	0.206
288	0.894	0.890	0.327	0.304	0.192
576	1.000	1.000	0.251	0.261	0.198
Nominal Size = 10%					
72	0.284	0.234	0.491	0.424	0.325
144	0.397	0.325	0.442	0.355	0.239
288	0.899	0.899	0.340	0.323	0.204
576	1.000	1.000	0.273	0.285	0.231
Noise/Signal Ratio = 1/ 2000					
Nominal Size = 5%					
72	0.264	0.323	0.491	0.424	0.342
144	0.221	0.189	0.422	0.329	0.211
288	0.419	0.420	0.295	0.355	0.256
576	0.918	0.952	0.222	0.263	0.210
Nominal Size = 10%					
72	0.284	0.372	0.502	0.433	0.402
144	0.273	0.228	0.445	0.368	0.235
288	0.440	0.456	0.327	0.374	0.256
576	0.919	0.955	0.256	0.304	0.232
Noise/Signal Ratio = 1/4000					
Nominal Size = 5%					
72	0.227	0.277	0.514	0.366	0.366
144	0.247	0.148	0.405	0.316	0.257
288	0.217	0.277	0.297	0.336	0.237
576	0.376	0.376	0.217	0.287	0.217
Nominal Size = 10%					
72	0.237	0.306	0.514	0.396	0.405
144	0.277	0.168	0.435	0.366	0.306
288	0.237	0.277	0.297	0.346	0.237
576	0.396	0.376	0.237	0.326	0.237
Panel B: Interval = $\widehat{\mu}_{IV_t} + 0.250\widehat{\sigma}_{IV_t}$					
Noise/Signal Ratio = 1/1000					
Nominal Size = 5%					
72	0.158	0.128	0.475	0.376	0.316
144	0.237	0.188	0.356	0.287	0.128
288	0.811	0.851	0.247	0.267	0.108
576	1.000	1.000	0.128	0.217	0.178
Nominal Size = 10%					
72	0.198	0.168	0.524	0.425	0.356
144	0.267	0.217	0.435	0.356	0.207
288	0.821	0.861	0.306	0.297	0.128
576	1.000	1.000	0.178	0.257	0.247
Noise/Signal Ratio = 1/2000					
Nominal Size = 5%					
72	0.198	0.128	0.445	0.366	0.336
144	0.108	0.049	0.356	0.336	0.138
288	0.198	0.188	0.257	0.257	0.099
576	0.881	0.871	0.148	0.227	0.128
Nominal Size = 10%					
72	0.247	0.178	0.504	0.415	0.386
144	0.158	0.069	0.425	0.366	0.217
288	0.227	0.207	0.306	0.297	0.148
576	0.881	0.871	0.207	0.267	0.247
Noise/Signal Ratio = 1/4000					
Nominal Size = 5%					
72	0.148	0.108	0.485	0.386	0.306
144	0.089	0.118	0.366	0.316	0.158
288	0.099	0.128	0.247	0.257	0.128
576	0.188	0.188	0.158	0.207	0.108
Nominal Size = 10%					
72	0.237	0.148	0.514	0.465	0.396
144	0.118	0.138	0.386	0.366	0.217
288	0.148	0.158	0.287	0.316	0.178
576	0.207	0.198	0.178	0.247	0.178

* Notes: See notes to Table 1. All experiments are based on samples of 100 daily observations.

Table 3: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case III: Jumps in DGP

M	$RV_{t,M}$	$BV_{t,M}$	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,e,M}$	$\mathcal{K}_{t,H,M}$
Panel A: Interval $= \widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$					
One i.i.d. $\mathrm{N}(0, 3 * 0.64 * \widehat{\mu}_{IV_t})$ *Jump Every 5 Days*					
Nominal Size = 5%					
72	0.297	0.316	0.643	0.623	0.504
144	0.356	0.277	0.485	0.495	0.415
288	0.287	0.316	0.356	0.495	0.326
576	0.198	0.198	0.435	0.435	0.217
Nominal Size = 10%					
72	0.465	0.356	0.712	0.643	0.603
144	0.425	0.346	0.554	0.554	0.465
288	0.326	0.376	0.415	0.564	0.366
576	0.247	0.257	0.514	0.485	0.267
One i.i.d. $\mathrm{N}(0, 1 * 0.64 * \widehat{\mu}_{IV_t})$ *Jump Every 2 Days*					
Nominal Size = 5%					
72	0.356	0.316	0.643	0.504	0.475
144	0.326	0.188	0.524	0.534	0.376
288	0.326	0.207	0.544	0.524	0.336
576	0.287	0.267	0.415	0.504	0.247
Nominal Size = 10%					
72	0.405	0.386	0.663	0.613	0.564
144	0.425	0.227	0.584	0.574	0.465
288	0.386	0.257	0.594	0.613	0.396
576	0.366	0.336	0.445	0.544	0.346
One i.i.d. $\mathrm{N}(0, 0.5 * 0.64 * \widehat{\mu}_{IV_t})$ *Jump Every 1 Days*					
Nominal Size = 5%					
72	0.396	0.237	0.495	0.524	0.504
144	0.366	0.148	0.386	0.554	0.376
288	0.306	0.138	0.366	0.495	0.316
576	0.227	0.198	0.366	0.405	0.306
Nominal Size = 10%					
72	0.455	0.297	0.603	0.613	0.594
144	0.465	0.217	0.445	0.643	0.465
288	0.336	0.148	0.445	0.584	0.405
576	0.277	0.237	0.445	0.485	0.346
Panel B: Interval $= \widehat{\mu}_{IV_t} + 0.250\widehat{\sigma}_{IV_t}$					
One i.i.d. $\mathrm{N}(0, 3 * 0.64 * \widehat{\mu}_{IV_t})$ *Jump Every 5 Days*					
Nominal Size = 5%					
72	0.386	0.267	0.603	0.534	0.455
144	0.326	0.267	0.514	0.425	0.396
288	0.247	0.356	0.356	0.465	0.326
576	0.297	0.287	0.435	0.366	0.306
Nominal Size = 10%					
72	0.396	0.297	0.613	0.564	0.485
144	0.366	0.287	0.554	0.475	0.405
288	0.297	0.366	0.376	0.485	0.356
576	0.316	0.297	0.465	0.415	0.356
One i.i.d. $\mathrm{N}(0, 1 * 0.64 * \widehat{\mu}_{IV_t})$ *Jump Every 2 Days*					
Nominal Size = 5%					
72	0.346	0.386	0.653	0.445	0.396
144	0.297	0.237	0.534	0.514	0.376
288	0.326	0.227	0.504	0.534	0.326
576	0.247	0.257	0.405	0.326	0.297
Nominal Size = 10%					
72	0.386	0.435	0.663	0.475	0.455
144	0.366	0.247	0.574	0.554	0.415
288	0.386	0.237	0.514	0.564	0.356
576	0.277	0.277	0.435	0.366	0.386
One i.i.d. $\mathrm{N}(0, 0.5 * 0.64 * \widehat{\mu}_{IV_t})$ *Jump Every 1 Days*					
Nominal Size = 5%					
72	0.425	0.306	0.514	0.465	0.544
144	0.356	0.257	0.445	0.475	0.405
288	0.346	0.207	0.386	0.534	0.336
576	0.316	0.237	0.336	0.405	0.386
Nominal Size = 10%					
72	0.524	0.366	0.524	0.504	0.613
144	0.376	0.297	0.475	0.574	0.485
288	0.425	0.267	0.425	0.584	0.415
576	0.356	0.287	0.366	0.504	0.455

* Notes: See notes to Table 2.

Figure 1: Predictive Densities for Intel Integrated Volatility Based on Various Realized Measures

One-Step Ahead Based Upon Data Until May 28, 1998: M=78, T=100

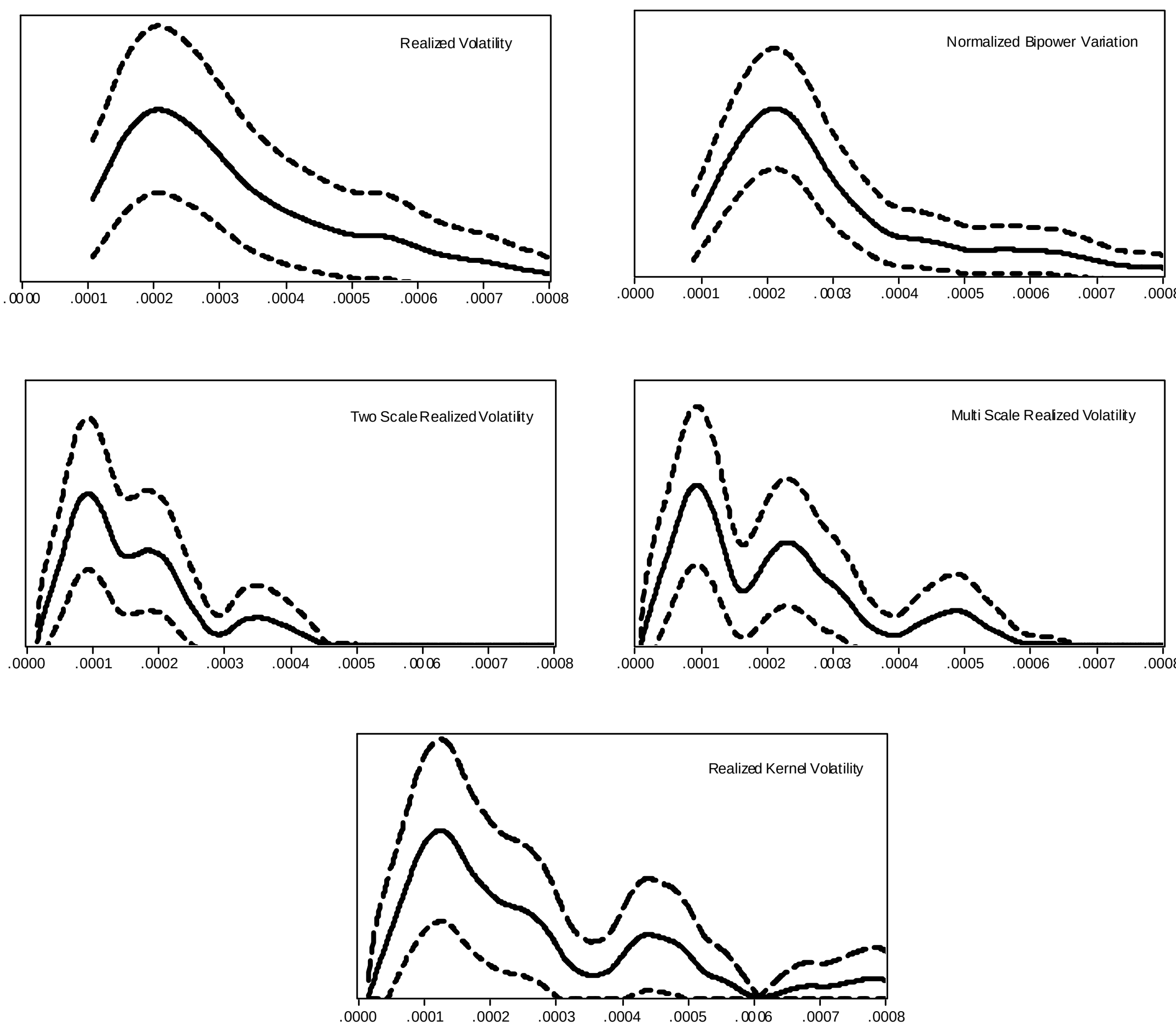

Figure 2: Predictive Densities for Intel Integrated Volatility Based on Various Realized Measures

One-Step Ahead Based Upon Data Until May 28, 1998: M=2340, T=100

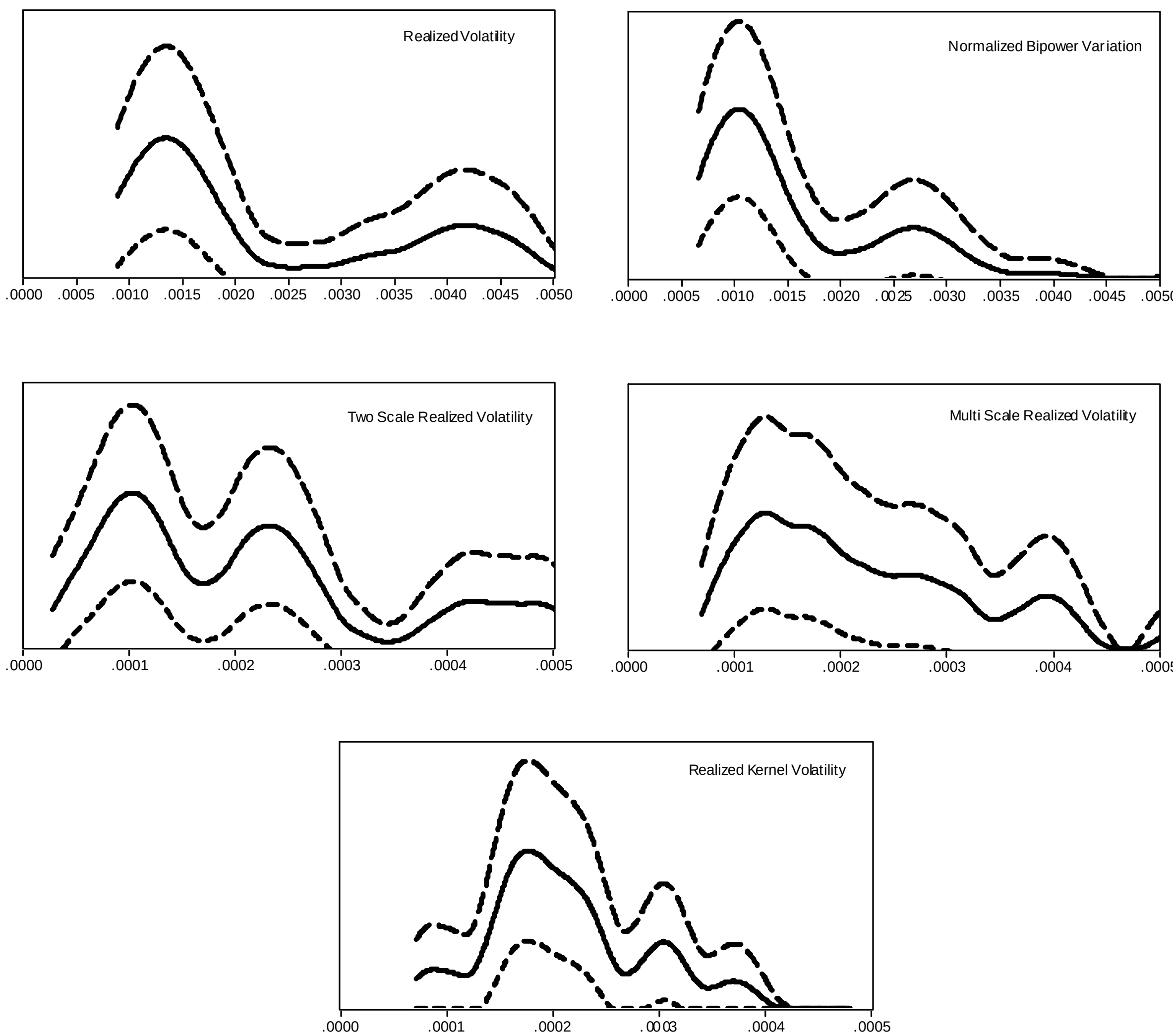

Figure 3: Predictive Densities for Intel Integrated Volatility Based on Various Realized Measures

One-Step Ahead Based Upon Data Until May 23, 2002: Various M, T=100

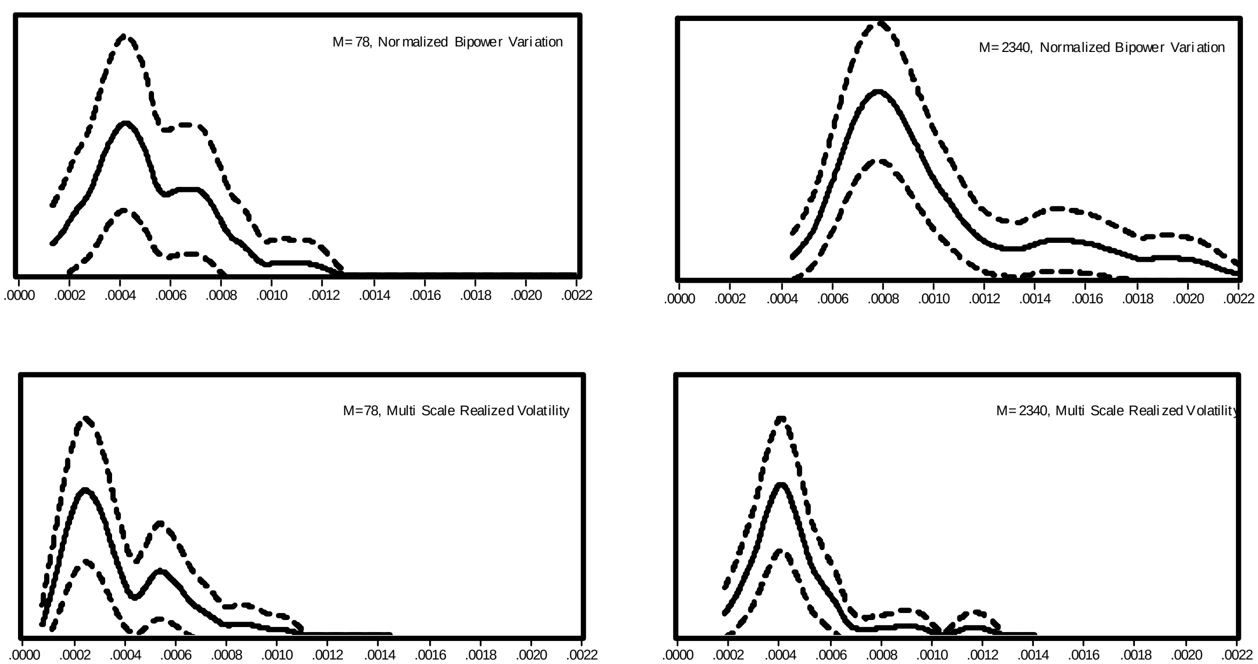

Figure 4: Predictive Densities for Intel Logged Integrated Volatility Based on Various Logged Realized Measures

One-Step Ahead Based Upon Data Various Dates: M=2340, T=100