Predictive Inference for Integrated Volatility*

Valentina Corradi[†]
Queen Mary, University of London

Walter Distaso[‡]
Queen Mary, University of London

Norman R. Swanson[§]
Rutgers University

April 2006

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 the unobservable volatility process, is provided. Asymptotic normality is also established. The efficacy of the 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 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 and the 2005 EC^2 Conference in Instabul, for helpful comments and suggestions on earlier drafts of this paper. Corradi and Distaso gratefully acknowledge ESRC grant RES-000-23-0006, and Swanson acknowledges financial support from a Rutgers University Research Council grant.

[†]Queen Mary, University of London, Department of Economics, Mile End, London, E14NS, UK, email: `v.corradi@qmul.ac.uk`.

[‡]Queen Mary, University of London, Department of Economics, Mile End, London, E14NS, UK, email: `w.distaso@qmul.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, from 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). Important early examples of such volatility-based derivatives include short and long dated volatility options on various currencies such as the British pound and the Japanese Yen; and VOLAX futures, which are based upon the implied volatility of DAX index options. 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 very mild assumptions on the process driving the behaviour 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

[1] See also Barndorff-Nielsen and Shephard (2004a) for an extension and generalization to the multivariate case.

Brownian semimartingale process, a class of processes which is commonly employed in continuous time modeling. Important variants of realized volatility have subsequently been proposed. These variants are largely motivated by the need to provide consistent estimators of integrated volatility in situations which are quite common in financial markets, such as when jumps occur in the asset price process, and when there are market frictions leading to market microstructure noise. Leading examples include bipower variation (Barndorff-Nielsen and Shephard, 2004b) and different estimators that are robust to the presence of microstructure noise (see, e.g., Zhang, 2004, Aït-Sahalia, Mykland and Zhang, 2005,2006, Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2006 and Zhang, Mykland and Aït-Sahalia, 2005). 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 intra day 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. Broadly speaking, it suffices that the supremum over $t \leq T$ of the k-th moment of the measurement error approaches zero at rate $T^{1/2}b_M^{-k/2}$, for some $k > 2$, where $b_M \to \infty$, as $M \to \infty$. Of note is that the rate at which the moments of the measurement error approach zero varies across different realized measures. We also provide conditions on the relative rate of growth of b_M and T, under which the estimators constructed using various realized measures are asymptotically equivalent to their unfeasible counterparts.

It should be noted that 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. In this case, 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 (see e.g., Andersen, Bollerslev, Diebold and Labys, 2003, Andersen, Bollerslev and Meddahi, 2004, 2005, 2006, Aït-Sahalia and Mancini, 2006, Ghysels and Sinko, 2006, and

Corradi, Distaso and Swanson, 2005).[2] The first three papers deal with the problem of pointwise prediction of integrated volatility, using ARMA models based on the log of realized volatility. Andersen, Bollerslev and Meddahi (2004, 2005) 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) address the forecasting volatility issue in the presence of microstructure effects. Aït-Sahalia and Mancini (2006) compare the out of sample relative forecast ability of realized volatility and two scale realized volatility (which is robust to the presence of microstructure noise) in a variety of contexts. According to their findings, two scale realized volatility outperforms realized volatility not only in the presence of microstructure noise, as expected, but also in the presence of jumps and long memory in the return process. Ghysels and Sinko (2006) analyze the relative predictive ability of realized volatility and two scale realized volatility within the mixed data sampling (MIDAS) framework. The papers cited above deal with pointwise prediction of integrated volatility. Corradi, Distaso and Swanson (2005) focus on estimating the conditional density of integrated volatility. This paper differs from theirs in a number of respects. First, we also examine conditional density and confidence interval estimators based on local linear regression. Second, by focusing on pointwise convergence, we achieve much faster rates of convergence than in Corradi, Distaso and Swanson (2005), where conditions for uniform convergence are outlined. Third, and perhaps most importantly, we establish asymptotic normality, allowing for the construction of confidence bands around our estimators. Fourth, our results are valid in a much more general setting than those obtained in Corradi, Distaso and Swanson (2005). For example, the results in this paper are not restricted to the class of eigenfunction stochastic volatility models of Meddahi (2001). Additionally, in this paper we allow for dependence in the market 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 realized measures 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 different intradaily data 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 ex-

[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, 2005a).

pected. 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. Asymptotic theory is gathered in Section 4. 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. In Section 6, an empirical illustration based upon the use of data from the New York Stock Exchange is discussed. All the proofs are contained in the Appendix.

2 Setup

We denote the log-price of a generic financial asset as Y_t, at a continuous time, t. In the paper we will assume that the log-price process belongs to the class of Brownian semimartingale processes with jumps. We will denote this by writing $Y \in \mathcal{BSMJ}$.[3] Then:

$$Y_t = \int_0^t \mu_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s + \sum_{i=1}^{J_t} c_i. \tag{1}$$

The drift component, μ_s, is a predictable process; and the diffusion term, σ_s, is a continuous process whose properties are specified below. The jump component is modeled as a sum of nonzero i.i.d. random variables, c_i, which are independent of J_t, which is a finite activity counting process. Hence, in this paper we consider the case of a finite number of jumps over any fixed time span. The process described in (1) is very general. For example, both the case of a constant drift and the case of a mean reverting linear drift are nested within this framework.

From a risk management perspective, it is of interest to separate the discontinuous (due to jumps) part of Y, denoted by Y^d, from the continuous Brownian component, Y^c. Indeed, 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:

$$\langle Y^c \rangle_t = \int_0^t \sigma_s^2 \mathrm{d}s \quad \text{and} \quad \left\langle Y^d \right\rangle_t = \sum_{i=1}^{J_t} c_i^2.$$

[3]The notation and setup that we use is similar to that adopted in a series of papers by Barndorff-Nielsen and Shephard.

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 has 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:

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

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:

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

Finally, we assume independence between the true latent process, Y, and the error term ϵ.[4]

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

[4]This assumption is standard in the literature on the estimation of integrated volatility in the presence of noise (see, e.g., Aït-Sahalia, Mykland and Zhang, 2005, Zhang, Mykland and Aït-Sahalia, 2005, Bandi and Russell, 2004, 2005 and Barndorff-Nielsen, Hansen, Lunde and Shephard, 2006). The work by Hansen and Lunde (2006) suggests that removing this assumption does not have too damaging an impact when using one minute returns data. As in Aït-Sahalia, Mykland and Zhang (2006), when the microstructure noise is i.i.d., then our setup allows for the existence of some contemporaneous correlation between Y and ϵ. Of further note is that the specification tests for microstructure noise models by Awartani, Corradi and Distaso (2005) are also derived without assuming independence between Y and ϵ.

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:

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

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 assumed to be a noisy measure of true integrated volatility. Namely, assume that:

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

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 four particular realized measures of integrated volatility.

The four measures that we consider are:

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

$$RV_{t,M} = \sum_{j=1}^{M-1} \left(X_{t+(j+1)/M} - X_{t+j/M} \right)^2. \tag{3}$$

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

$$(\mu_1)^{-2} BV_{t,M} = (\mu_1)^{-2} \frac{M}{M-1} \sum_{j=2}^{M-1} \left| X_{t+(j+1)/M} - X_{t+j/M} \right| \left| X_{t+j/M} - X_{t+(j-1)/M} \right|, \tag{4}$$

where $\mu_1 = \mathrm{E}\,|Z| = 2^{1/2}\Gamma(1)/\Gamma(1/2)$ and Z is a standard normal random variable.

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

$$\widehat{RV}_{t,l,M} = RV_{t,l,M}^{avg} - 2l\widehat{\nu}_{t,M}, \tag{5}$$

where

$$\widehat{\nu}_{t,M} = \frac{RV_{t,M}}{2M} = \frac{1}{2M} \sum_{j=1}^{M-1} \left(X_{t+\frac{j}{M}} - X_{t+\frac{j-1}{M}} \right)^2 \tag{6}$$

and

$$RV_{t,l,M}^{avg} = \frac{1}{B}\sum_{b=1}^{B} RV_{t,l}^{b} = \frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(X_{t+\frac{jB+b}{M}} - X_{t+\frac{(j-1)B+b}{M}}\right)^2. \tag{7}$$

Here, $Bl \cong M$, $l = O(M^{1/3})$, l denotes the subsample size, and B denotes the number of subsamples. The logic underlying (5) is as follows. The first step is to construct B realized volatilities using l non overlapping subsamples. Then, one takes an average of these B realized volatilities and corrects this average by using an estimator of the bias term due to market microstructure noise, where the bias estimator is constructed using a finer grid of equispaced observations.

(iv) *Multi Scale Realized Volatility* (due to Zhang, 2004, Aït-Sahalia, Mykland and Zhang, 2006):

$$\begin{aligned}\widetilde{RV}_{t,e,M} &= \sum_{i=1}^{e} a_i \widetilde{RV}_{t,e_i,M} + \frac{RV_{t,M}}{M} \\ &= \sum_{i=1}^{e} a_i \frac{1}{e_i}\left(\sum_{j=1}^{M-e_i}\left(X_{t+\frac{j+e_i}{M}} - X_{t+\frac{j}{M}}\right)^2\right) + \frac{RV_{t,M}}{M},\end{aligned} \tag{8}$$

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 $i = 1, ..., e$. For $e_i = i$,

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

Barndorff-Nielsen, Hansen, Lunde and Shephard (2006) adopt a different approach in the estimation of integrated volatility. They use kernels to consistently estimate the contribution to quadratic variation of X due to the Brownian component of Y. When the asset price dynamics is described by a scaled Brownian motion they also suggest an optimal choice of e. Their estimator is asymptotically equivalent to $\widetilde{RV}_{t,e,M}$. Therefore, results derived for $\widetilde{RV}_{t,e,M}$ will also hold (at least asymptotically) for their kernel based realized volatility measure.

We now discuss our estimation methodology.

3 Conditional (Predictive) Density and Confidence Interval Estimation

As discussed to in the introduction, conditional densities of integrated volatility play a crucial role in the pricing of riskiness associated with the variability of volatility, and are instrumental when one wishes to construct model free measures of variance risk premia. Given this fact, our objective is to construct a nonparametric estimator of the density of integrated volatility at time $T+h$, conditional on given realized volatility measures that are 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 $h=1$. Extension to the case of $\tau-$step ahead prediction follows directly, for finite τ. Loosely speaking, our objective is to predict the density 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:

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

Following the typical approach, our conditional density estimator is defined as the ratio of joint and marginal density estimators, where the latter is assumed to be bounded away from zero. Thus, the kernel based conditional density estimator for a generic realized measure can be defined as:

$$\widehat{f}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)}) = \frac{\frac{1}{T}\sum_{t=d}^{T-1} \mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)\frac{1}{\xi_2}}{\frac{1}{T}\sum_{t=d}^{T-1} \mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}}, \tag{9}$$

where $\mathbf{K}$ can either be a $d-$dimensional kernel, or the product of d univariate kernel functions, such as:

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

where H and K are one dimensional kernels, which may or may not be the same. Note that we use the same bandwidth, ξ_1, for all of the conditioning variables, $RM_{t,M}, \ldots, RM_{t-d+1,M}$; and we use a different bandwidth, ξ_2, for the variable to be predicted, $RM_{t+1,M}$. The rationale behind this choice is the following. Since the underlying integrated volatility process is assumed to be strictly stationary (see A1 below), and as we are evaluating the density of the conditioning variables at the observed values $RM_{T,M}, \ldots, RM_{T-d+1,M}$, it is natural to use the same bandwidth for each conditioning variable. On the other hand, we want to compute the predictive density at some arbitrary point x, and therefore it may be sensible to use a different bandwidth for the dependent variable.[5]

Now, define the unfeasible conditional density estimator based on the unobservable integrated volatility process as follows:

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

Further, let $f_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})$ be the true conditional density evaluated at $RM_{T,M}^{(d)}$. Intuitively:

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

tends to zero (in probability) at a rate depending on the number of intraday observations M. On the other hand, it is immediate to see that:

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

approaches zero (in probability) at a rate depending only on T and on the bandwidth parameters.

In the sequel, we shall provide conditions on the relative rates of growth of M and T (as $M, T \to \infty$) as well as ξ_1 and ξ_2 (as $\xi_1, \xi_2 \to 0$), under which the limiting distribution of:

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

is the same as that of:

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

In other words, we provide conditions under which the effect of the measurement error associated with the realized volatility measure is asymptotically negligible. In this sense, we provide a tool for model free prediction of integrated volatility.

In financial risk management, it is often of interest to predict the probability that daily volatility will fall within a given interval, conditional on present and past values of a given realized measure. In our context, this corresponds to the construction of a model free estimator of:

$$\Pr\left((u_1 \leq IV_{T+1} \leq u_2)|RM_{T,M}^{(d)}\right).$$

[5] Of course, one may choose to let ξ_2 depend on the evaluation point x.

Provided that the daily volatility process is strictly stationary (see A1 below), it is immediate to see that:

$$\mathrm{E}\left(1_{\{u_1 \leq IV_{T+1} \leq u_2\}} | RM_{T,M}^{(d)}\right) = \Pr\left((u_1 \leq IV_{T+1} \leq u_2) | RM_{T,M}^{(d)}\right),$$

and, in order to measure the quantity above, we can therefore employ a standard Nadaraya-Watson estimator. The relevant statistic of interest is thus based on:

$$\begin{aligned}
&\widehat{F}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(u_2|RM_{T,M}^{(d)}) - \widehat{F}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(u_1|RM_{T,M}^{(d)}) \\
&= \frac{\frac{1}{T}\sum_{t=d}^{T-1} 1_{\{u_1 \leq RM_{t+1,M} \leq u_2\}} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}}{\frac{1}{T}\sum_{t=d}^{T-1} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}} \\
&= \frac{\frac{1}{T}\sum_{t=d}^{T-1} 1_{\{u_1 \leq RM_{t+1,M} \leq u_2\}} \mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}}{\widehat{f}_{RM_{T,M}^{(d)}}(RM_{T,M}^{(d)})}. \qquad (10)
\end{aligned}$$

As in the case of conditional densities, we also define the unfeasible conditional confidence interval, based on the unobservable integrated volatility IV_t. Namely, define:

$$\begin{aligned}
\widehat{F}_{IV_{T+1}|IV_T^{(d)}}(u_2|RM_{T,M}^{(d)}) - \widehat{F}_{IV_{T+1}|IV_T^{(d)}}(u_1|RM_{T,M}^{(d)}) &= \frac{\frac{1}{T}\sum_{t=d}^{T-1} 1_{\{u_1 \leq IV_{t+1} \leq u_2\}} \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}}{\frac{1}{T}\sum_{t=d}^{T-1} \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}} \\
&= \frac{\frac{1}{T}\sum_{t=d}^{T-1} 1_{\{u_1 \leq IV_{t+1} \leq u_2\}} \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}}{\widehat{f}_{IV_T^{(d)}}(RM_{T,M}^{(d)})}.
\end{aligned}$$

Further, the true conditional interval is:

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

Visual inspection of (9) and (10) immediately reveals that, in the case of conditional confidence interval estimation, the curse of dimensionality is reduced by one. However, the treatment of measurement error in this case is more complex, as the error enters into the indicator function, which is a nondifferentiable function. As a result, the conditions on the rate of growth of M, relative to T, that are required to ensure that measurement error is asymptitically negligible turn out to be more stringent, as we shall see in the next section.

4 Asymptotic Theory

We begin by stating the assumptions which will be used to derive our asymptotic results.

Assumption A1: $IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s$ is a strictly stationary $\alpha-$mixing process with size $-2r/(r-2)$, $r > 2$.

Assumption A2:

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

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

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

Assumption A3:

(i) The kernel K is a symmetric, nonnegative, continuous function with bounded support $[-\Delta, \Delta]^d$, 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:

$$\sup_{t \leq T} \mathrm{E}\left(|N_{t,M}|^k\right) = O\left(T^{1/2} b_M^{-k/2}\right), \ k \geq 2$$

where $N_{t,M} = RM_{t,M} - IV_t$, and $RM_{t,M}$ is a generic realized measure.

Assumption A1 requires the daily volatility process to be strong mixing. For example, this assumption holds for the case of the eigenfunction stochastic volatility models of Meddahi (2001). In this case, integrated volatility has an ARMA structure and is strong mixing, with mixing coefficients declining at a geometric rate. It should be pointed out that A2 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.[6]

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 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. Thus, 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 $M \to \infty$, uniformly in t; also, it implicitely requires that M grow fast enough relative to T. In Subsection 4.4, we shall provide primitive conditions, in terms of moments of σ_t^2 and μ_t, under which A5 is satisfied by the four realized measures defined in (3),(4),(5) and (8).

4.1 Predictive Density Results

We are now in position to state our main results summarizing the asymptotic behavior of the kernel based conditional (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 1. *Let A1-A5 hold. Then, pointwise in x :*
(i) If $\xi_1, \xi_2 \to 0$ and $T\xi_1^d\xi_2 \to \infty$, then:

$$\widehat{f}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)}) - \widehat{f}_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)}) = O_P\left(T^{\frac{1}{2k-1}} b_M^{-1/2}\right).$$

(ii) If $\xi_1, \xi_2 \to 0$, $T\xi_1^d\xi_2 \to \infty$, and $T^{\frac{2k+1}{2k-1}} b_M^{-1} \xi_1^d \xi_2 \to 0$, then:

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

where k is defined as in Assumption 5.

[6]Hence, we do not need to choose between one-factor and stochastic volatility models. In this context, however, it should be noted that Corradi and Distaso (2005b) suggest a testing procedure for choosing between one-factor and stochastic volatility models, without making any assumption on the parametric specification of either the drift or variance term.

As the unfeasible estimator based on the unobservable integrated volatility process for IV_t converges to the true conditional density at rate $\sqrt{T\xi_1^d\xi_2}$, the limiting distributions of the feasible and unfeasible estimators are asymptotically equivalent, provided that $T^{\frac{2k+1}{2k-1}}b_M^{-1}\xi_1^d\xi_2 \to 0$. In this sense, the rate at which b_M has to grow (relative to T) is lower the higher is the number of conditioning variables and the faster the bandwidth parameters go to zero.

We can now establish the limiting distribution of our conditional density estimator.

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* $T^{\frac{2k+1}{2k-1}}b_M^{-1}\xi_1^d\xi_2 \to 0$, *then:*

$$\begin{aligned}&\sqrt{T\xi_1^d\xi_2}\left(\widehat{f}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)}) - f_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})\right)\\&\xrightarrow{d} \mathrm{N}\left(0, \left(\frac{f_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})}{f_{IV_T^{(d)}}(RM_{T,M}^{(d)})}\int \mathbf{K}^2(u)\mathrm{d}u\int K^2(v)\mathrm{d}v\right)\right).\end{aligned} \tag{11}$$

The limit theory provided above is unfeasible, since the variance in (11) needs to be estimated. A feasible limit theory is given in the following corollary.

Corollary 1. *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* $T^{\frac{2k+1}{2k-1}}b_M^{-1}\xi_1^d\xi_2 \to 0$, *then:*

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

As the underlying conditional density is at least twice differentiable, the bias term goes to zero at a rate not slower than $\min(\xi_1^2, \xi_2^2)$. Thus, if $T\xi_1^{4+d}\xi_2 \to 0$ and $T\xi_1^d\xi_2^5 \to 0$, the bias term associated with the estimator will be asymptotically negligible. If instead, $T\xi_1^{4+d}\xi_2$ and $T\xi_1^d\xi_2^5$ converge to a fixed constant, then in principle one could bias-correct the estimated density. However, bias correction is only feasible when the degree of differentiability of the density is known.

4.2 Predictive Confidence Interval Results

We now turn to the asymptotic behavior of the conditional confidence interval estimator defined in (10).

Theorem 3. *Let A1-A5 hold. Then:*
(i) If $\xi \to 0$ *and* $T\xi^d \to \infty$, *then:*

$$\begin{aligned}
&\left(\widehat{F}_{RM_{T+1,M}|RM^{(d)}_{T,M}}(u_2|RM^{(d)}_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM^{(d)}_{T,M}}(u_1|RM^{(d)}_{T,M})\right)\\
&- \left(\widehat{F}_{IV_{T+1}|IV^{(d)}_T}(u_2|RM^{(d)}_{T,M}) - \widehat{F}_{IV_{T+1}|IV^{(d)}_T}(u_1|RM^{(d)}_{T,M})\right)\\
&= O_P\left(T^{\frac{3}{2k-1}} b_M^{-1/2}\right) + O_P\left(T^{-\frac{1}{2}+\frac{3}{4k-2}\left(\frac{1}{4}+\frac{1}{2r}\right)} b_M^{-\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2r}\right)}\right),
\end{aligned}$$

where $k \geq 2$ *(see A5) and* $r > 2$ *(see A1).*
(ii) If $\xi \to 0$, $T\xi^d \to \infty$, *and* $T^{\frac{2k+5}{2k-1}} b_M^{-1} \xi^d \to 0$, *then:*

$$\begin{aligned}
&\sqrt{T\xi^d}\left(\left(\widehat{F}_{RM_{T+1,M}|RM^{(d)}_{T,M}}(u_2|RM^{(d)}_{T,M}) - \widehat{F}_{RM_{T+1,M}|RM^{(d)}_{T,M}}(u_1|RM^{(d)}_{T,M})\right)\right.\\
&\left.- \left(F_{IV_{T+1}|IV^{(d)}_T}(u_2|RM^{(d)}_{T,M}) - F_{IV_{T+1}|IV^{(d)}_T}(u_1|RM^{(d)}_{T,M})\right)\right)\\
&= \sqrt{T\xi^d}\left(\left(\widehat{F}_{IV_{T+1}|IV^{(d)}_T}(u_2|RM^{(d)}_{T,M}) - \widehat{F}_{IV_{T+1}|IV^{(d)}_T}(u_1|RM^{(d)}_{T,M})\right)\right.\\
&\left.- \left(F_{IV_{T+1}|IV^{(d)}_T}(u_2|RM^{(d)}_{T,M}) - F_{IV_{T+1}|IV^{(d)}_T}(u_1|RM^{(d)}_{T,M})\right)\right) + o_P(1).
\end{aligned}$$

From part (i) of the theorem, it follows that, in the case of conditional confidence interval estimation, the measurement error is asymptotically negligible if b_M goes to zero at a rate faster than $T^{\frac{6}{2k-1}}$. This contrasts with the statement of Theorem 1, part (i), where it is required that $b_M \to \infty$ at a rate faster than $T^{\frac{2}{2k-1}}$. This is due to the fact that in the conditional confidence interval case, the measurement error component enters in the indicator function, which is a non differentiable function. Furthermore, $N_{t,M}$ is not independent of IV_t.[7] Therefore, in our proofs we can no longer employ intermediate value expansions, but we can only use the fact that:

$$\sup_{t \leq T} N_{t,M} = O_P(T^{\frac{3}{2k-1}} b_M^{-1/2}).$$

The next theorem and corollary state the unfeasible and feasible limiting distributions of our conditional confidence interval estimators.

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

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

[7] Meddahi (2002, Proposition 4.3) shows that in the case of realized volatility, and in the absence of leverage, the measurement error is uncorrelated with the integrated volatility process.

where

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

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

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

Comparing the statements in Theorems 2 and 4 with the associated corollaries, we see that in the latter case the asymptotic equivalence of the limiting distributions of kernel estimators constructed using the unobservable integrated volatility and a generic realized volatility measure requires that b_M grows faster than $T^{\frac{2k+5}{2k-1}}\xi^d$. This condition is more demanding than in Theorem 2, where it is required that $T^{\frac{2k+1}{2k-1}} b_M^{-1}\xi_1^d\xi_2 \to 0$.

4.3 Local Linear Estimators Results

So far, we have resorted to standard Nadaraya-Watson and kernel estimators for conditional confidence intervals and densities. Nevertheless, a viable alternative is to use local linear estimators. The objective of this subsection is to provide the relevant asymptotic theory for local linear estimators.

We begin by considering conditional density estimators. Following Fan, Yao and Tong (1996), define $\widehat{\boldsymbol{\beta}}_{T,M}(x, RM_{T,M}^{(d)})$ as:

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

where

$$
S_{T,M}(\boldsymbol{\beta}; x, RM_{T,M}^{(d)})
$$

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

$\boldsymbol{\beta}=(\beta_0,\boldsymbol{\beta}_1')'$, K and $\mathbf{K}$ are defined as in (9). Therefore:

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

where $\mathbf{X}_{(M)}$ is a matrix of dimension $(T-d)\times(d+1)$, $\mathbf{W}_{(M)}$ is a diagonal $(T-d)\times(T-d)$ matrix, and $\mathbf{Y}_{(M)}$ is a $(T-d)\times 1$ vector. In particular:

$$\mathbf{X}_{(M)}=\begin{pmatrix}1 & RM_{d,M}-RM_{T,M} & \cdots & RM_{1,M}-RM_{T-(d-1),M}\\ 1 & RM_{d+1,M}-RM_{T,M} & \cdots & RM_{2,M}-RM_{T-(d-1),M}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & RM_{T-1,M}-RM_{T,M} & \cdots & RM_{T-(d-2),M}-RM_{T-(d-1),M}\end{pmatrix},\tag{12}$$

$$\mathbf{W}_{(M)}=\frac{1}{\xi_1^d}\begin{pmatrix}\mathbf{K}\left(\frac{RM_{d,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right) & 0 & \cdots & 0\\ 0 & \mathbf{K}\left(\frac{RM_{d+1,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \mathbf{K}\left(\frac{RM_{T,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\end{pmatrix}\tag{13}$$

and

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

The local linear estimator of the conditional density is given by $\widehat{\beta}_{0,T,M}(x,RM_{T,M}^{(d)})$. Hereafter, let:

$$\widehat{f}^{(ll)}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)})=\widehat{\beta}_{0,T,M}(x,RM_{T,M}^{(d)})$$

and let $\widehat{f}^{(ll)}_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})$ be its unfeasible counterpart. We have the following result.

Theorem 5. *Let A1-A5 hold. Then, pointwise in x:*

(i) If $\xi_1,\xi_2\to 0$, $T\xi_1^d\xi_2\to\infty$, and $T^{\frac{2k+1}{2k-1}}b_M^{-1}\xi_1^d\xi_2\to 0$, then:

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

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

(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* $T^{\frac{2k+1}{2k-1}}b_M^{-1}\xi_1^d\xi_2\to 0$, *then:*

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

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

We now consider local linear estimators of conditional confidence intervals. Define $\widehat{\boldsymbol{\alpha}}_{T,M}(u_1,u_2,RM_{T,M}^{(d)})$ as:

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

where

$$Z_{T,M}(\boldsymbol{\alpha};u_1,u_2,RM_{T,M}^{(d)})$$
$$=\frac{1}{T\xi_1^d\xi_2}\sum_{t=d}^{T}\left(1_{\{u_1\le RM_{t+1}\le u_2\}}-\alpha_0-\boldsymbol{\alpha}_1'\left(RM_{t,M}^{(d)}-RM_{T,M}^{(d)}\right)\right)^2\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)$$

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

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

where $\mathbf{X}_{(M)}$ and $\mathbf{W}_{(M)}$ are defined in (12) and (13) (using $\xi_1=\xi$), and:

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

The local linear estimator of the conditional density is given by $\widehat{\alpha}_{0,T,M}(u_1,u_2,RM_{T,M}^{(d)})$. Hereafter, let:

$$\widehat{F}^{(ll)}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(u_2|RM_{T,M}^{(d)})-\widehat{F}^{(ll)}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(u_1|RM_{T,M}^{(d)})=\widehat{\alpha}_{0,T,M}(u_1,u_2,RM_{T,M}^{(d)}),$$

and let:

$$\widehat{F}^{(ll)}_{IV_{T+1}|IV_T^{(d)}}(u_2|RM^{(d)}_{T,M}) - \widehat{F}^{(ll)}_{IV_{T+1}|IV_T^{(d)}}(u_1|RM^{(d)}_{T,M})$$

denote the corresponding estimator obtained using the unobservable IV_t. The local estimator for conditional distributions outlined above has been recently used by Aït-Sahalia, Fan and Peng (2005), 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 approximation do instead lie between 0 and 1 for any sample size (see Hall, Wolff and Yao, 1999); however they typically cannot be written in closed form. Finally we have the following result.

Theorem 6. *Let A1-A5 hold. Then:*

(i) If $\xi \to 0$, $T\xi^d \to \infty$, *and* $T^{\frac{2k+5}{2k-1}} b_M^{-1} \xi^d \to 0$, *then:*

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

(ii) If $\xi \to 0$, $T\xi^d \to \infty$, $T\xi^{4+d} \to 0$, *and* $T^{\frac{2k+5}{2k-1}} b_M^{-1} \xi^d \to 0$, *then:*

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

where $V(u_1,u_2)$ *is defined as in the statement of Theorem 4.*

4.4 Primitive Conditions for Assumption A5

We now provide primitive conditions on the moments of the drift and variance terms which ensure that Assumption A5 is satisfied by the four realized measures outlined in Section 2.

Lemma 1. *If* $\mathrm{E}\left(\left(\sigma_t^2\right)^{2k+\eta}\right) < \infty$ *and* $\mathrm{E}\left(\left(\mu_t\right)^{2k+\eta}\right) < \infty$, *for some* $\eta > 0$, *then there is a sequence* b_M, *where* $b_M \to \infty$ *as* $M \to \infty$, *such that for all finite* k :

(i) If $J_t \equiv 0$ *for all* t *(no jumps) and* $\epsilon \equiv 0$ *(no microstructure noise), then:*

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

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

$$\sup_{t\leq T} \mathrm{E}\left(\left|BV_{t,M} - \mu_1^2 IV_t\right|^k\right) = O(T^{1/2} b_M^{-k/2}), \text{ with } b_M = M,$$

where $\mu_1 = \mathrm{E}(|Z|) = 2^{1/2}\Gamma(1)/\Gamma(1/2)$, and Z is a standard normal variate.

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

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

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

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

where $RV_{t,M}$, $BV_{t,M}$, $\widehat{RV}_{t,l,M}$, and $\widetilde{RV}_{t,e,M}$ are defined in (3), (4), (5) and (7), respectively.

From the proof of the Lemma above, it is immediate to see that in the case of no leverage and/or bounded drift and variance, the following sharper rates hold. Namely:

$$\sup_{t\leq T} \mathrm{E}\left(\left|RV_{t,M} - IV_t\right|^k\right) = O(M^{-k/2}),$$

$$\sup_{t\leq T} \mathrm{E}\left(\left|BV_{t,M} - \mu_1^2 IV_t\right|^k\right) = O(M^{-k/2}),$$

$$\sup_{t\leq T} \mathrm{E}\left(\left|\widehat{RV}_{t,l,M} - IV_t\right|^k\right) = O(M^{-k/6}),$$

$$\sup_{t\leq T} \mathrm{E}\left(\left|\widetilde{RV}_{t,e,M} - IV_t\right|^k\right) = O(M^{-k/4}).$$

Thus, in the case of no leverage and/or bounded drift and variance term, Assumption A5 can be restated as A5':

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

Now, with regard to the conditional density estimators, by replacing A5 with A5', Theorem 1(i) holds with a measurement error term of order $O_P\left(b_M^{-1/2}\right)$, instead of $O_P\left(b_M^{-1/2} T^{\frac{1}{2k-1}}\right)$. Additionally, the statements in Theorem 1(ii), Theorem 2 and Corollary 1 require that $T b_M^{-1} \xi_1^d \xi_2 \to 0$, instead of $T^{\frac{2k+1}{2k-1}} b_M^{-1} \xi_1^d \xi_2 \to 0$. With regard to the conditional interval estimators, under A5', the statement in Theorem 3(i) holds with a measurement error term of order

$$O_P\left(T^{\frac{1}{k-1}} b_M^{-1/2}\right) + O_P\left(T^{-\frac{1}{2}+\frac{1}{2(k-1)}\left(\frac{1}{4}+\frac{1}{2r}\right)} b_M^{-\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2r}\right)}\right),$$

instead of

$$O_P\left(T^{\frac{3}{2(k-1)}}b_M^{-1/2}\right)+O_P\left(T^{-\frac{1}{2}+\frac{3}{4k-2}\left(\frac{1}{4}+\frac{1}{2r}\right)}b_M^{-\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2r}\right)}\right).$$

Additionally, the statements in Theorem 3(ii), Theorem 4 and Corollary 2 require that $T^{\frac{k+1}{k-1}}b_M^{-1}\xi^d \to 0$, instead of $T^{\frac{2k+5}{2k-1}}b_M^{-1}\xi^d \to 0$.

Finally, within the class of eigenvalues stochastic volatility models of Meddahi (2001), and in the case of constant drift, it follows from Corradi and Distaso (2005a, Propositions 1-3) that for $k = 1, 2:$

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

4.5 Remarks

Remark 1. From Lemma 1, it follows that $b_M = M$ for the case of realized volatility and bipower variation, while $b_M = M^{1/3}$ and $b_M = M^{1/2}$ for $RM_{t,M} = \widehat{RV}_{t,l,M}$ and $RM_{t,M} = \widetilde{RV}_{t,e,M}$, respectively. 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 2 and 4 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 2 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 return 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:

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

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 $\sup_{t \leq T} \mathrm{E}\left(|\log(RM_{t,M}) - \log(IV_t)|^k\right) = O(T^{1/2}b_M^{-k/2})$. 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.

5 Monte Carlo Results

In this section, our objective is to assess the finite sample efficacy of the limiting distribution result given for the conditional (predictive) interval estimator in Corollary 2. Namely, we will 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 will then be to assess the finite sample properties of the interval estimator by examining the empirical level properties of $G_{T,M}(u_1,u_2)$ for a given data generating process (DGP). In particular, data are generated according to the following DGP:

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

where $W_{1,t}$ and $W_{2,t}$ may be correlated with correlation coefficient ρ. Using a Milstein scheme, S paths of length $d+1$ are generated, with S much larger than T, and with a very small discrete interval $(1/N)$ between successive observations. These data are used to construct a "pseudo true" confidence interval for integrated volatility.

Then, Monte Carlo iterations are carried out by generating paths of length T days. 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.[8] Simulated data are then sampled at frequency $1/M$, for various values of M, where $M < N$. Thereafter, the four realized volatility measures outlined in Section 2 are constructed, using each path of simulated asset data, and are in turn used to construct $G_{T,M}(u_1,u_2)$.

In the implementation of this experiment, we set $m = 0.05$, $\eta = 1$, $\kappa = 2.5$, $v = 1$, and $\rho = 0$.[9] In our basic case (denoted by Case I in Table 1), we simply set $X_{t+j/M} = Y_{t+j/M}$ (i.e. we assume that there is no microstructure noise), for $t = 1,\ldots,T$, and $j = 1,\ldots,M$. Additionally, we assume in this case that $J_t \equiv 0$ for all t (no jumps). Under Case II (see Table 2), daily data are generated by adding microstructure noise. Accordingly, $X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}$,

[8] 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.

[9] This parameterization is artificial, in the sense that it is not calibrated from any particular dataset. Rejection frequency results reported below may worsen when the magnitude of the mean reversion parameter is decreased, for example.

$t = 1, \ldots, T$, and $j = 1, \ldots, M$, as discussed above, where we set $\epsilon_{t+j/M} \sim i.i.d.\ \mathrm{N}(0, \nu)$, and where $\nu = \{(3 * 1152)^{-1}, (4 * 1152)^{-1}, (5 * 1152)^{-1}\}$. 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 $\{3, 5, 10\}$, 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, on average. We set $S = 3000$, and $d = 1$. Additionally, 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 $\beta = \{0.125, 0.25, 0.50\}$. The associated confidence intervals based on these combinations of $[u1, u2]$ are 0.128, 0.181, and 0.341, for the three different β values, respectively. We consider daily samples of $T = \{100, 300, 500\}$ observations, and we set $M = \{72, 144, 288, 576\}$. Experimental results are reported for estimated intervals constructed using a quartic kernel. Additionally, whenever the message of the experiment is not lost, tabulated results are only reported for $T = 100$. Finally, all results are based upon 500 Monte Carlo iterations.

In Tables 1-3, rejection frequencies are reported, using two-sided 10% nominal level tests. The first column of rejection frequencies contains results for $RV_{t,M}$, the second for $BV_{t,M}$, the third for $\widehat{RV}_{t,l,M}$ and the fourth for $\widetilde{RV}_{t,e,M}$. Results for different values of M are reported in different rows of the tables. A number of clear conclusions emerge upon examination of the results.

Turning first to Table 1, where there is no microstructure noise or jumps in the DGP, note that $RV_{t,M}$ and $BV_{t,M}$ perform approximately equally well. Both of these measures have empirical sizes close to the nominal 10% level in various cases, and there is a substantial improvement as both M and T increase. Indeed, in many cases the nominal size is achieved, or very nearly so, a finding which might be viewed as rather surprising given the small values of M and T used in our 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 three confidence interval widths reports in Panels A-C 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 two subsampled measures. In particular, note that empirical rejection frequencies for $\widehat{RV}_{t,l,M}$ and $\widetilde{RV}_{t,e,M}$ are often 0.20-0.30 when $M = 576$, whereas rates for $RV_{t,M}$ and $BV_{t,M}$ are often 0.10-0.20. Furthermore, empirical performance of $\widehat{RV}_{t,l,M}$ and $\widetilde{RV}_{t,e,M}$ is very 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. 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 and 3, where microstructure noise and jumps are added to the DGP. We report results only for the case where $T = 100$.[10] Our conclusions based on these tables are as follows. Under microstructure noise (Table 2), $\widehat{RV}_{t,l,M}$ and $\widetilde{RV}_{t,e,M}$ are both superior to the non-microstructure noise robust realized measures, particularly for large values of M, as expected. For example, consider Panel C in Table 2. The rejection frequency for $RV_{t,M}$ ranges from 0.310-0.763 for $M = 576$, whereas the analogous entries for $\widetilde{RV}_{t,e,M}$ range from 0.190-0.196, which indicates a marked improvement when using the robust measure, as long as M is large, and even though T is only 100. Of course, for M too small, there is nothing to gain by using the robust measures. Indeed, for $M = 48$, $RV_{t,M}$ rejection frequencies are much closer to the nominal level than $\widehat{RV}_{t,l,M}$ and $\widetilde{RV}_{t,e,M}$ rejection frequencies. Additionally, $BV_{t,M}$ outperforms all other measures under jumps (see Table 3). Note also that within each panel in Table 3, there are three different jump intensities reported on. As expected, when jump intensity increases, while holding jump frequency, T, and M fixed, the performance of $BV_{t,M}$ remains stable, while that of the other three measures worsens.

In summary, the above experiment suggests that our asymptotic theory 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 our asymptotic theory 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.

[10] In order to extrapolate the reported findings to $T = 300$ and $T = 500$, one need only note that the magnitudes of rejection frequencies associated with increasing T are of the same order of magnitude as those reported in Table 1. Tabulated results illustrating this are available upon request from the authors.

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 is generally accepted as the highest frequency at which the effect of microstructure biases are not too distorting (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.[11]

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 four considered realized measures. 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).[12] 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 and multi scale realized volatility are somewhat similar (see Figure 1, for example).

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

[11] 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.

[12] 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.

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 esitmators based on these robust measures (see the lower 2 plots in Figure 3).

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.

Appendix

Proof of Theorem 1:

(i) Note that:[13]

$$\widehat{f}_{RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)}) = \frac{\frac{1}{T}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)\frac{1}{\xi_2}}{\widehat{f}_{IV_T^{(d)}}(x|RM_{T,M}^{(d)})} \tag{14}$$

$$+\left(\frac{1}{\widehat{f}_{RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)})}-\frac{1}{\widehat{f}_{IV_T}(x|RM_{T,M}^{(d)})}\right)$$
$$\times\frac{1}{T}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)\frac{1}{\xi_2}. \tag{15}$$

Given A2-A3, begin by expanding the numerator of the right hand side of (14) as follows:

$$\frac{1}{T}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)\frac{1}{\xi_2}$$
$$=\frac{1}{T}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\frac{1}{\xi_2}$$
$$+\frac{1}{T}\sum_{t=d}^{T-1}\sum_{i=0}^{d-1}K_i^1\left(\frac{\widetilde{RM}_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^{d+1}}K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\frac{1}{\xi_2}N_{t-i,M} \tag{16}$$
$$+\frac{1}{T}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^d}K^1\left(\frac{\widetilde{RM}_{t+1,M}-x}{\xi_2}\right)\frac{1}{\xi_2^2}N_{t,M}, \tag{17}$$

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

We need to find the orders of probability of the terms in (16) and (17), which represent the contribution of measurement error. We begin by considering (16):

$$\frac{1}{T}\sum_{t=d}^{T-1}\sum_{i=0}^{d-1}K_i^{(1)}\left(\frac{\widetilde{RM}_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^{d+1}}K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\frac{1}{\xi_2}N_{t-i,M}.$$

Let

$$R_{t,i,M}=K_i^{(1)}\left(\frac{\widetilde{RM}_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi_1}\right)\frac{1}{\xi_1^{d+1}}K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\frac{1}{\xi_2}.$$

[13] All summations should be rescaled by $\frac{1}{T-d}$, however for notational brevity we rescale by $\frac{1}{T}$.

Now, recalling A5:

$$\Pr\left(T^{-\frac{1}{2k-1}}b_M^{1/2}\left|\frac{1}{T}\sum_{t=d}^{T-1}\sum_{i=0}^{d-1}R_{t,i,M}N_{t-i,M}\right|>\varepsilon\right)\leq\sum_{i=0}^{d-1}\Pr\left(T^{-\frac{1}{2k-1}}b_M^{1/2}\left|\frac{1}{T}\sum_{t=d}^{T-1}R_{t,i,M}N_{t-i,M}\right|>\varepsilon\right)$$
$$\leq\frac{d}{\varepsilon^k}T^{-\frac{k}{2k-1}}b_M^{k/2}\sup_{i\leq d}\sup_{t\leq T}\mathrm{E}\left(|R_{t,i,M}|^k|N_{t-i,M}|^k\right)$$
$$=\frac{d}{\varepsilon^k}T^{-\frac{k}{2k-1}}b_M^{k/2}O\left(T^{1/2}b_M^{-k/2}\right)\to 0,$$

so that (16) is $O_P(T^{\frac{1}{2k-1}}b_M^{-1/2})$. The term in (17) can be treated in an analogous way, and so it is $O_P(T^{\frac{1}{2k-1}}b_M^{-1/2})$.

Finally, we evaluate (15). As $\xi_1^d,\xi_2\to 0$ and $T\xi_1^d\xi_2\to\infty$,

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

since it satisfies a law of large numbers, while:

$$\left(\frac{1}{\widehat{f}_{RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)})}-\frac{1}{\widehat{f}_{IV_T}(x|RM_{T,M}^{(d)})}\right)=\left(\frac{\widehat{f}_{IV_T}(x|RM_{T,M}^{(d)})-\widehat{f}_{RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)})}{\widehat{f}_{RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)})\widehat{f}_{IV_T}(x|RM_{T,M}^{(d)})}\right)$$
$$=O_P(T^{-\frac{1}{2k-1}}b_M^{-1/2}),$$

by the same argument as the one used above. The statement in the theorem then follows.

(ii) Immediate by Part (i). ■

Proof of Theorem 2:

From Theorem 4.1 in Robinson (1983). ■

Proof of Theorem 3:

(i) We start by rearranging our object of interest:

$$\begin{aligned}&\widehat{F}_{RM_{T+1,M}|RM_{T,M}^{(d)}}(x|RM_{T,M}^{(d)})-\widehat{F}_{IV_{T+1}|IV_T^{(d)}}(x|RM_{T,M}^{(d)})\\&=\frac{1}{\widehat{f}_{IV_T^{(d)}}(RM_{T,M}^{(d)})}\\&\times\left(\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{RM_{t+1,M}\leq u\}}\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)-1_{\{IV_{t+1}\leq u\}}\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right)\right)\\&+\left(\frac{\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)-\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right)}{\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)}\right)\end{aligned}\quad(18)$$

$$\times \frac{1}{T\xi^d}\sum_{t=d}^{T-1} 1_{\{RM_{t+1,M}\leq u\}}\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right). \tag{19}$$

We begin by considering the term in (18). Since:

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

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:

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

$$= \frac{1}{T\xi^d}\sum_{t=d}^{T-1} 1_{\{IV_{t+1}\leq u\}}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right) - \mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right) \tag{20}$$

$$+ \frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{IV_{t+1}\leq u\}} - 1_{\{RM_{t+1,M}\leq u\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right) \tag{21}$$

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

$$\left.\times\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right) - \mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right)\right). \tag{22}$$

The term in (20) can be treated as in the proof of Theorem 1, part (i), and is thus $O_P(T^{-\frac{1}{2k-1}}b_M^{-1/2})$.

We now consider (21):

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

$$\leq \frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{u-\sup_{t\leq T}|N_{t+1,M}|\leq IV_{t+1}\leq u+\sup_{t\leq T}|N_{t+1,M}|\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right). \tag{23}$$

Now, given A5:

$$\begin{aligned}\Pr\left(\sup_{t\leq T} T^{-\frac{3}{2k-1}}b_M^{1/2}\left|N_{t,M}\right| > \varepsilon\right) &\leq \sum_{t=d}^{T-1}\Pr\left(T^{-\frac{3}{2k-1}}b_M^{1/2}\left|N_{t,M}\right| > \varepsilon\right)\\ &\leq \frac{1}{\varepsilon^k}TT^{-\frac{3k}{2k-1}}b_M^{k/2}\sup_{t\leq T}\mathrm{E}\left(\left|N_{t,M}\right|^k\right)\\ &\leq \frac{1}{\varepsilon^k}TT^{-\frac{3k}{2k-1}}b_M^{k/2}O(T^{1/2}b_M^{-k/2}) \to 0, \text{ as } T,M\to\infty.\end{aligned}$$

Thus, for all samples except a subset with probability measure approaching zero as $T, M \to \infty$, there is a constant c such that:

$$\begin{aligned}
&\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{u-\sup_{t\leq T}|N_{t+1,M}|\leq IV_{t+1}\leq u+\sup_{t\leq T}|N_{t+1,M}|\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\\
&\leq \frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{u-c\varepsilon b_M^{-1/2}T^{3/(2k-1)}\leq IV_{t+1}\leq u+c\varepsilon b_M^{-1/2}T^{3/(2k-1)}\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right). \qquad (24)
\end{aligned}$$

Hereafter, let $c\varepsilon b_M^{-1/2}T^{\frac{3}{2k-1}} = d_{T,M}$. Then, using (23) and (24):

$$\begin{aligned}
&\left|\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\{IV_{t+1}\leq u\}}-1_{\{RM_{t+1,M}\leq u\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right|\\
&\leq \frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\\
&\leq \left|\frac{1}{T}\sum_{t=d}^{T-1}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}-\mathrm{E}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\right)\right.\\
&\left.\times\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}-\mathrm{E}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}\right)\right)\right| \qquad (25)\\
&+\left|\frac{1}{T}\sum_{t=d}^{T-1}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}-\mathrm{E}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\right)\right.\\
&\left.\times\mathrm{E}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}\right)\right| \qquad (26)\\
&+\left|\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}-\mathrm{E}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}\right)\right)\right.\\
&\left.\times\mathrm{E}\left(\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\right)\right| \qquad (27)\\
&+\mathrm{E}\left(\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\right)\mathrm{E}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}\right). \qquad (28)
\end{aligned}$$

Now, $\mathrm{E}\left(\frac{1}{\xi^d}\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right) = O(1)$ and

$$\mathrm{E}\left(1_{\left\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\right\}}\right)\leq 2\sup_x f_{IV_{t+1}}(x)d_{T,M}.$$

Thus, the term in (28) is $O_P(d_{T,M}) = O_P\left(T^{\frac{3}{2k-1}}b_M^{-1/2}\right)$.

Also, by the law of large numbers:

$$\frac{1}{T}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}-\mathrm{E}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}\right)\right) = o_P(1),$$

so that the term in (27) is $o_P\left(T^{\frac{3}{2k-1}}b_M^{-1/2}\right)$.

Now, for all samples, except a subset with probability measure approaching zero as $T, M \to \infty$,

$$\begin{aligned}
&\operatorname{var}\left(\frac{1}{T}\sum_{t=d}^{T-1}\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}-\mathrm{E}\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\right)\right)\right)\\
&=\frac{1}{T^2}\sum_{t=d}^{T-1}\sum_{s=d}^{T-1}\mathrm{E}\left(\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}-\mathrm{E}\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\right)\right)\right.\\
&\left.\times\left(1_{\{u-d_{T,M}\leq IV_{s+1}\leq u+d_{T,M}\}}-\mathrm{E}\left(1_{\{u-d_{T,M}\leq IV_{s+1}\leq u+d_{T,M}\}}\right)\right)\right)\\
&\leq\frac{1}{T^2}\sum_{t=d}^{T-1}\sum_{s=d}^{T-1}\alpha_{|t-s|}^{1/2-1/r}\mathrm{E}\left(\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}-\mathrm{E}\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\right)\right)^2\right)^{1/2}\\
&\times\mathrm{E}\left(\left(1_{\{u-d_{T,M}\leq IV_{s+1}\leq u+d_{T,M}\}}-\mathrm{E}\left(1_{\{u-d_{T,M}\leq IV_{s+1}\leq u+d_{T,M}\}}\right)\right)^r\right)^{1/r}\\
&=O\left(T^{-1}d_{T,M}^{1/2+1/r}\right),
\end{aligned}$$

because of A1. Therefore, the term in (26) is $O_P\left(T^{-\frac{1}{2}+\frac{3}{4k-2}\left(\frac{1}{2}+\frac{1}{r}\right)}b_M^{-\frac{1}{4}\left(\frac{1}{2}+\frac{1}{r}\right)}\right)$.

We are left with the term in (25). Let:

$$Q_{t+1,M}=1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}-\mathrm{E}\left(1_{\{u-d_{T,M}\leq IV_{t+1}\leq u+d_{T,M}\}}\right)$$

and

$$U_t=\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}-\mathrm{E}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\frac{1}{\xi^d}\right).$$

Now:[14]

$$\operatorname{var}\left(\frac{1}{T}\sum_{t=1}^{T}Q_{t+1,M}U_t\right)=\frac{1}{T^2}\sum_{t=d}^{T-1}\sum_{s=d}^{T-1}\mathrm{E}\left(Q_{t+1,M}U_tQ_{s+1,M}U_s\right).$$

For $s<t-1$, define the sigma field $\mathcal{F}_{s+1}^{Q,U}=\sigma\left(Q_{\tau+1,M},U_\tau,\ \tau\leq s\right)$. Then, denoting by C a generic constant, we know that the following mixing inequality holds:

$$\begin{aligned}
\left|\mathrm{E}\left(Q_{t+1,M}U_tQ_{s+1,M}U_s\right)\right|&\leq\left|\mathrm{E}\left(\mathrm{E}\left(Q_{t+1,M}U_tQ_{s+1,M}U_s\right)|\mathcal{F}_{s+1}^{Q,U}\right)\right|=\left|\mathrm{E}\left(Q_{s+1,M}U_s\mathrm{E}\left(Q_{t+1,M}U_t|\mathcal{F}_{s+1}^{Q,U}\right)\right)\right|\\
&\leq\left(\mathrm{E}\left(Q_{s+1,M}^2U_s^2\right)\right)^{1/2}\left(\mathrm{E}\left(Q_{s+1,M}^rU_s^r\right)\right)^{1/r}\alpha_{t-(s+1)}^{1/2-1/r}\\
&\leq\mathrm{E}(Q_{t+1,M}^{4+2r})^{1/4+1/2r}\mathrm{E}(U_t^{4+2r})^{1/4+1/2r}\alpha_{t-(s+1)}^{1/2-1/r}\\
&\leq CT^{\frac{3}{2k-1}\left(\frac{1}{4}+\frac{1}{2r}\right)}b_M^{-\frac{1}{2}\left(\frac{1}{4}+\frac{1}{2r}\right)}\alpha_{t-(s+1)}^{1/2-1/r},
\end{aligned}$$

as $\mathrm{E}(Q_{t+1,M}^{4+2r})=O(T^{\frac{3}{2k-1}}b_M^{-\frac{1}{2}})$, and $\mathrm{E}\left(U_t^{4+2r}\right)=O(1)$.

Therefore, the term in (25) is $O_P\left(T^{\frac{3}{4k-2}\left(\frac{1}{4}+\frac{1}{2r}\right)-\frac{1}{2}}b_M^{-\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2r}\right)}\right)$. Finally, applying the Cauchy-Schwartz inequality, it follows that (22) is of a smaller order of probability than (20) and (21).

[14]Notice that since M enters only in the conditioning set, we write U_t instead of $U_{t,M}$.

Summarizing, the numerator of (18) is:

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

$$=O_P\left(T^{\frac{1}{2k-1}}b_M^{-1/2}\right)++O_P\left(T^{\frac{3}{2k-1}}b_M^{-1/2}\right)+O_P\left(T^{\frac{1}{4k-2}\left(\frac{1}{4}+\frac{1}{2r}\right)-\frac{1}{2}}b_M^{-\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2r}\right)}\right)$$

$$+O_P\left(T^{-\frac{1}{2}+\frac{3}{4k-2}\left(\frac{1}{2}+\frac{1}{r}\right)}b_M^{-\frac{1}{4}\left(\frac{1}{2}+\frac{1}{r}\right)}\right)$$

$$=O_P\left(T^{\frac{3}{2k-1}}b_M^{-1/2}\right)+O_P\left(T^{-\frac{1}{2}+\frac{3}{4k-2}\left(\frac{1}{4}+\frac{1}{2r}\right)}b_M^{-\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2r}\right)}\right).$$

It now remains to consider the term in (19). First, note that:

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

and that the term in the denominator is bounded away from zero. Also, by the same argument as the one used in the proof of Theorem 1, part (i):

$$\frac{1}{T\xi^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{IV_t^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)-\mathbf{K}\left(\frac{RM_{t,M}^{(d)}-RM_{T,M}^{(d)}}{\xi}\right)\right)=O_P\left(T^{\frac{1}{2k-1}}b_M^{-1/2}\right).$$

The statement in part (i) then follows.

(ii) From part (i), it is immediate to see that:

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

$$=O_P\left(T^{\frac{1}{2}+\frac{3}{2k-1}}b_M^{-1/2}\xi^{\frac{d}{2}}\right)+O_P\left(T^{\frac{3}{2k-1}\left(\frac{1}{4}+\frac{1}{2r}\right)}b_M^{-\frac{1}{4}\left(\frac{1}{4}+\frac{1}{2r}\right)}\xi^{\frac{d}{2}}\right). \tag{29}$$

The first term on the right hand side of (29) approaches zero, provided that $T^{\frac{2k+5}{2k-1}}\xi^d b_M^{-1}\to 0$. It is immediate to see that the establish convergence to zero implies convergence to zero of the second term as well. ■

Proof of Theorem 4:

Follows from Remark 6 in Hall, Wolff and Yao (1999). ■

Proof of Theorem 5:

(i) 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 $\mathbf{X}_{(M)}, \mathbf{W}_{(M)}, \mathbf{y}_{(M)}$, but replacing the realized measure series with integrated volatility. Note that:

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}}_{T,M}(x, RM_{T,M}^{(d)}) - \widehat{\boldsymbol{\beta}}_T(x, IV_T) \\
&= \left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} T^{-1}\left(\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{y}_{(M)} - \mathbf{X}'\mathbf{W}\mathbf{y}\right) \\
&+ \left(T^{-1}\left(\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)^{-1} - \left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\right) T^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \\
&+ \left(\left(T^{-1}\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{X}_{(M)}\right)^{-1} - \left(T^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\right)\left(T^{-1}\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{y}_{(M)} - T^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}\right). \qquad (30)
\end{aligned}$$

Now:

$$\begin{aligned}
&T^{-1}\left(\mathbf{X}'_{(M)}\mathbf{W}_{(M)}\mathbf{y}_{(M)} - \mathbf{X}'\mathbf{W}\mathbf{y}\right) \\
&= \begin{pmatrix}
\frac{1}{T\xi_1^d\xi_2}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right) - \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) K\left(\frac{IV_{t+1}-x}{\xi_2}\right)\right) \\
\frac{1}{T\xi_1^d\xi_2}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)(RM_{t,M} - RM_{T,M})\right. \\
\left. -\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) K\left(\frac{IV_{t+1}-x}{\xi_2}\right)(IV_t - RM_{T,M})\right) \\
\vdots \\
\frac{1}{T\xi_1^d\xi_2}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) K\left(\frac{RM_{t+1,M}-x}{\xi_2}\right)(RM_{t-(d-1),M} - RM_{T-(d-1),M})\right. \\
\left. -\mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) K\left(\frac{IV_{t+1}-x}{\xi_2}\right)(IV_{t-(d-1)} - RM_{T-(d-1),M})\right)
\end{pmatrix} \\
&= O_P(T^{\frac{1}{2k-1}} b_M^{-1/2}),
\end{aligned}$$

by the same argument used in the proof of Theorem 1(i).

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 $1 < i,j \leq d+1$, is given by:

$$\begin{aligned}
&\frac{1}{T\xi_1^d}\sum_{t=d}^{T-1}\left(\mathbf{K}\left(\frac{RM_{t,M}^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right)(RM_{t-(j-1),M} - RM_{T-(j-1),M})(RM_{t-(i-1),M} - RM_{T-(i-1),M})\right. \\
&\left. - \mathbf{K}\left(\frac{IV_t^{(d)} - RM_{T,M}^{(d)}}{\xi_1}\right) K\left(\frac{IV_{t+1}-x}{\xi_2}\right)(IV_{t-(j-1)} - RM_{T-(j-1),M})(IV_{t-(i-1)} - RM_{T-(i-1),M})\right) \\
&= O_P(T^{\frac{1}{2k-1}} b_M^{-1/2}),
\end{aligned}$$

and so, provided the two matrix are uniformly positive definite,

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

The statement of the Theorem then follows.

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

Proof of Theorem 6:

(i) Follows using the same argument as in the proof of Theorem 5, and by noting that the indicator function should be treated as in the proof of Theorem 3.

(ii) Follows from Remark 4, in Hall, Wolff and Yao (1999). ■

Proof of Lemma 1:

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

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

Also, for 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:

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

Define $\Omega^+=\{\omega:\limsup_T T^{-1/2}\sigma_T^{2k}>\varepsilon\}$, and note that:

$$\Pr\left(\limsup_T T^{-1/2}\sigma_T^{2k}>\varepsilon\right)\leq\sum_{t=1}^{T}\Pr\left(T^{-1/2}\sigma_t^{2k}>\varepsilon\right)\leq TT^{-2+\eta/2}\mathrm{E}\left(\left(\sigma_t^2\right)^{2k+\eta}\right)\to 0,$$

so that $\Pr(\Omega^{+}) = 1$. Therefore, for all $\omega \in \Omega^{+}$:

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

where C denotes a generic constant.

For $j_1 \neq j_2$, $\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s$ is independent of $\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s$. Thus,

$$\begin{aligned}
T^{-1/2}\mathrm{E}\left(\left(\sqrt{M}N_{t+1,M}^{(1)}\right)^{4}\right) &\leq CM^{2}\sum_{j_1}\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^{4}\right) \\
&+CM^{2}\sum_{j_1}\sum_{j_2\neq j_1}\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^{2}\right. \\
&\times\left.\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^{2}\right) \\
&=O(M^{-1})+CM^{2}\sum_{j_1}\sum_{j_2\neq j_1}\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^{2}\right) \\
&\times\left(\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\left(\int_{t+j_2/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^{2}\right) \\
&=O(M^{-1})+O(1)\text{ uniformly in }t,
\end{aligned}$$

given that $\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\left(\int_{t+j_1/M}^{s}\mathrm{d}W_u\right)\mathrm{d}W_s\right)^{4}\right) = O(M^{-4})$.

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

With regard to the drift term, its contribution to the measurement error on an interval of length $1/M$ is given by

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

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

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) By the same argument as that used in the proof of Lemma 5 in Corradi and Distaso (2005a), the orders of magnitude of the moments of the pure jump component and of the cross term component are respectively smaller and equal to the order of magnitude of the moments of the continuous component. Therefore we can ignore them. Let $\Delta X_{t+(j+1)/M}=X_{t+(j+1)/M}-X_{t+j/M}$. Thus:

$$\begin{aligned}
&BV_{t,M}-\mu_1^2 IV_t\\
&=\sum_{i=1}^{M-1}\sigma^2_{t+(i-1)/M}\left(\left|\int_{(i-1)/M}^{i/M}\mathrm{d}W_s\right|\left|\int_{i/M}^{(i+1)/M}\mathrm{d}W_s\right|-\frac{\mu_1^2}{M}\right)\\
&+\sum_{i=1}^{M-1}\left(\left|\Delta X_{t+(j+1)/M}\right|\left|\Delta X_{t+(j+1)/M}\right|-\sigma^2_{t+(i-1)/M}\left|\int_{(i-1)/M}^{i/M}\mathrm{d}W_s\right|\left|\int_{i/M}^{(i+1)/M}\mathrm{d}W_s\right|\right)\\
&-\mu_1^2\sum_{i=1}^{M-1}\int_{i/M}^{(i+1)/M}\left(\sigma^2_{(i-1)/M}-\sigma^2_s\right)\mathrm{d}s\\
&=\mu_1^2\left(N^{(1)}_{t,M}+N^{(2)}_{t,M}+N^{(3)}_{t,M}\right),
\end{aligned}\tag{31}$$

where μ_1 is defined as in the statement of the Lemma. As in the proof of Part (i), we proceed conditioning on $\omega\in\Omega^+$. Hereafter, let $\left|\int_{(i-1)/M}^{i/M}\mathrm{d}W_s\right|=\left|\Delta W_{i/M}\right|$, and note that

$$\mathrm{E}\left(\left(\left|\Delta W_{i/M}\right|-\frac{\mu_1^2}{M}\right)\left(\left|\Delta W_{s/M}\right|-\frac{\mu_1^2}{M}\right)\right)=0$$

for all s such that $|i-s|>1$. As in Part (i), for the sake of notational simplicity, we consider the case of $k=4$; again, the case of $k>4$ can be treated in an analogous manner. Hereafter, with the notation $\approx$ we mean "of the same order of magnitude". Thus:

$$\begin{aligned}
&T^{-1/2}\mathrm{E}\left(\left(\mu_1^2N^{(1)}_{t,M}\right)^4\right)\\
&\approx\sum_{j_1}\mathrm{E}\left(\left(\left(\left|\Delta W_{j_1/M}\right|-\frac{\mu_1^2}{M}\right)\left(\left|\Delta W_{(j_1+1)/M}\right|-\frac{\mu_1^2}{M}\right)\right)^4\right)\\
&+\sum_{j_1}\sum_{j_2\neq j_1}\mathrm{E}\left(\left(\left(\left|\Delta W_{j_1/M}\right|-\frac{\mu_1^2}{M}\right)\left(\left|\Delta W_{(j_1+1)/M}\right|-\frac{\mu_1^2}{M}\right)\right)^2\right.\\
&\left.\left(\left(\left|\Delta W_{j_2/M}\right|-\frac{\mu_1^2}{M}\right)\left(\left|\Delta W_{(j_2+1)/M}\right|-\frac{\mu_1^2}{M}\right)\right)^2\right)\\
&=O\left(M^{-2}\right),
\end{aligned}$$

by the same argument as that used in Part (i); and, as in Part (i), for any generic k, $\mathrm{E}\left(\left(\mu_1^2 N_{t,M}^{(1)}\right)^k\right) = O\left(T^{1/2}M^{-k/2}\right)$, uniformly in t.

With regard to $N_{t,M}^{(2)}$,

$$\mathrm{E}\left(\left(\mu_1^2 N_{t,M}^{(2)}\right)^k\right) = O\left(T^{1/2}M^{-k/2}\right),$$

uniformly in t, by the same argument as that used in the proof of Theorem 5.1 in Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006). Finally, $\mathrm{E}\left(\left(\mu_1^2 N_{t,M}^{(3)}\right)^k\right) = O\left(T^{1/2}M^{-k/2}\right)$, uniformly in t, by the same argument as that used in the proofs in Section 7 of Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006). It should be pointed out that Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard assume bounded drift and instantaneous variance. In our case σ_t^2 and μ_t are not bounded in general, a feature which is captured by the extra $T^{1/2}$ term in the orders of magnitude.

(iii) We can expand as follows:

$$\begin{aligned}
&\mathrm{E}\left(\left(\widehat{RV}_{t,l,M} - IV_t\right)^k\right) \\
&= \mathrm{E}\left(\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(X_{t+(jB+b)/M} - X_{t+((j-1)B+b)/M}\right)^2\right.\right.\right. \\
&\left.\left.\left.-\frac{l}{M}\sum_{j=1}^{M}\left(X_{t+j/M} - X_{t+((j-1)/M)}\right)^2\right) - IV_t\right)^k\right) \\
&\approx \mathrm{E}\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(Y_{t+(jB+b)/M} - Y_{t+((j-1)B+b)/M}\right)^2 - IV_t\right)^k\right) &(32)\\
&+ \mathrm{E}\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(Y_{t+(jB+b)/M} - Y_{t+((j-1)B+b)/M}\right)\left(\epsilon_{t+(jB+b)/M} - \epsilon_{t+((j-1)B+b)/M}\right)\right)^k\right) &(33)\\
&+ \mathrm{E}\left(\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(\left(\epsilon_{t+(jB+b)/M} - \epsilon_{t+((j-1)B+b)/M}\right)^2 - 2\nu\right)\right)^k\right) &(34)\\
&+ \mathrm{E}\left(\left(\frac{l}{M}\sum_{j=1}^{M}\left(\left(\epsilon_{t+j/M} - \epsilon_{t+(j-1)/M}\right)^2 - 2\nu\right)\right)^k\right). &(35)
\end{aligned}$$

Provided that $\mathrm{E}(\epsilon_{t+j/M}^{2k}) < \infty$, the term in (35) is $O(l^k/M^{k/2})$, when the microstructure noise is i.i.d.. For the geometrically mixing microstructure error case, (35) 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 (34) is $O(l^{k/2}/B^{k/2})$. Thus, (35) and (34) are $O(b_M^{-k/2})$, with $b_M = M^{1/3}$, provided that $B = M^{2/3}$ and $l = M^{1/3}$.

Given that the noise is independent of the price process, the term in (33) is $o(b_M^{-k/2})$. We are left with (32).

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

From the proof of part (i), it follows that term in (37) is $O(T^{1/2}M^{-k/2})$. With regard to (36), from the proof of Theorem 2 in Zhang, Mykland and Aït-Sahalia (2005), note that:

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

where the last term captures the end effects.

By the same argument used in the proof of part (i), given that $\mathrm{E}\left(\left(\sigma_t^2\right)^{2k+\eta}\right) < \infty$ and $\mathrm{E}\left((\mu_t)^{2k+\eta}\right) < \infty$, for all samples except a subset of measure zero, and $k = 4$:

$$\begin{aligned}
&\mathrm{E}\left(\left(\sum_{j=1}^{M}\left(Y_{t+(j+1)/M} - Y_{t+j/M}\right)\sum_{i=1}^{B\wedge j}\sum_{i=1}^{B\wedge j}\left(1 - \frac{j}{B}\right)\left(Y_{t+(j-i+1)/M} - Y_{t+(j-i)/M}\right)\right)^4\right) \\
&\approx \mathrm{E}\left(\left(\sum_{j=1}^{M-1}\sigma_{t+j/M}^{2k}\int_{t+j/M}^{t+(j+1)/M}\mathrm{d}W_s\sum_{i=1}^{B\wedge j}\int_{t+(j-i)/M}^{t+(j-i+1)/M}\mathrm{d}W_s\right)^4\right) \\
&\leq CT^{1/2}\sum_{j_1=1}^{M}\sum_{i_1=1}^{B\wedge j_1}\sum_{j_2\neq j_1}^{M}\sum_{i_2=1}^{B\wedge j_2}\mathrm{E}\left(\left(\int_{t+j_1/M}^{t+(j_1+1)/M}\mathrm{d}W_s\int_{t+(j_1-i_1)/M}^{t+(j_1-i_1+1)/M}\mathrm{d}W_s\right)^2\right. \\
&\left.\left(\int_{t+j_2/M}^{t+(j_2+1)/M}\mathrm{d}W_s\int_{t+(j_2-i_2)/M}^{t+(j_2-i_2+1)/M}\mathrm{d}W_s\right)^2\right)
\end{aligned}$$

$$= O\left(T^{1/2}B^2M^{-2}\right) = O(T^{1/2}b_M^{-2}), \text{ for } b_M = M^{1/3}. \tag{38}$$

By the same argument, for any generic k, the term in (36) is $O\left(T^{1/2}B^{k/2}M^{-k/2}\right) = O(T^{1/2}b_M^{-k/2})$.
(iv) Recalling that $\sum_{i=1}^{e}\frac{a_i}{i} = 0$, along the lines of Zhang (2004), it follows that:

$$\begin{aligned}
&\widetilde{RV}_{t,e,M} - IV_t \\
&= \left(\sum_{i=1}^{e}\frac{a_i}{i}\sum_{j=1}^{M-i}\left(Y_{t+(j+i)/M} - Y_{t+j/M}\right)^2 - IV_t\right) && (39)\\
&- 2\sum_{i=1}^{e}\frac{a_i}{i}\sum_{j=1}^{M-i}\epsilon_{t+(j+i)/M}\epsilon_{t+j/M} && (40)\\
&+ 2\sum_{i=1}^{e}\frac{a_i}{i}\sum_{j=1}^{M-i}\left(Y_{t+(j+i)/M} - Y_{t+j/M}\right)\left(\epsilon_{t+(j+i)/M} - \epsilon_{t+j/M}\right) && (41)\\
&- \left(\sum_{i=1}^{e}\frac{a_i}{i}\sum_{j=1}^{i}\left(\epsilon_{t+j/M} - \nu\right)\right) && (42)\\
&- \left(\sum_{i=1}^{e}\frac{a_i}{i}\sum_{j=M-i}^{M}\left(\epsilon_{t+j/M} - \nu\right)\right) && (43)\\
&+ 2\left(\widehat{\nu}_{t,M} - \nu\right),
\end{aligned}$$

where $\widehat{\nu}_{t,M}$ is defined in (6) and $\mathrm{E}\left(\epsilon_{t+j/M}^2\right) = \nu$. By the same argument as that used in the proof of part (iii), it follows that $\mathrm{E}\left((\widehat{\nu}_{t,M} - \nu)^k\right) = O(M^{-k/2})$. Note that $a_i \approx i^2/e^3$. Therefore:

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

so that the expectations of the k-th moments of (42) and (43) are $O(e^{-k/2})$.

Now, with regard to the term in (40),

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

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

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

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

and 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 $\mathrm{E}(\epsilon_{t+j/M}^{2k+\eta}) < \infty$, with $\eta > 0$, by the same argument as that used in Yoshihara (1975, Lemma 1).

With regard to the term in (39), 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):

$$\mathrm{E}\left(\left(\sum_{i=1}^{e}\frac{a_i}{i}\sum_{j=1}^{M-i}\left(Y_{t+(j+i)/M} - Y_{t+j/M}\right)^2 - IV_t\right)^k\right) \tag{44}$$
$$= O\left(T^{1/2}e^{-k/2}\right) = O\left(T^{1/2}M^{-k/4}\right) = O\left(T^{1/2}b_M^{-k/2}\right),$$

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

Finally, the expecation of the k-th power of (41) is of smaller order than (44). ■

References

Aït-Sahalia, Y., J. Fan and H. Peng (2005). 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.

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.

Andersen, T.G. and T. Bollerslev (1997). Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance*, 4, 115-158.

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., V. Corradi and W. Distaso (2005). Testing and Modelling Market Microstructure Effects with an Application to the Dow Jones Industrial Average. FERC Working Paper, WP04-06.

Bandi, F.M. and J.R. Russell (2004). Microstructure Noise, Realized Volatility, and Optimal Sampling. Working Paper, University of Chicago, Graduate School of Business.

Bandi, F.M. and J.R. Russell (2005). Separating microstructure noise from volatility. *Journal of Financial Economics*, forthcoming.

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

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2006). Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. 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 (2004a). Econometric Analysis of Realised Covariation: High Frequency Covariance, Regression and Correlation in Financial Economics. *Econometrica*, 72, 885-925.

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

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.

Corradi, V. and W. Distaso (2005a). Semiparametric Comparison of Stochastic Volatility Models via Realized Measures. *Review of Economic Studies*, forthcoming.

Corradi, V. and W. Distaso (2005b). Testing for One-Factor Models versus Stochastic Volatility Models. Manuscript, Queen Mary, University of London.

Corradi, V., W. Distaso and N.R. Swanson (2005). Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures. Manuscript, Queen Mary, University of London.

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., 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.

Hansen, P.R. and A. Lunde (2006). Realized Variance and Market Microstructure Noise (with discussion). *Journal of Business & Economic Statistics*, forthcoming.

Hong, Y. and Li, H. (2005). Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates. *Review of Financial Studies*, 18, 37-84.

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

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.

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

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

Robinson, P.M. (1983). Nonparametric Estimators for Time Series. *Journal of Time Series Analysis,* 4, 185-207.

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.

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

Yoshihara, K. (1975). Weak Convergence of Multidimensional Empirical Processes for Strong Mixing Sequences of Stochastic Vectors. Z. Wahrscheinlichkeitstheprie verw. 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. (2004). Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. Manuscript, Carnegie Mellon University.

Table 1: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case I: No Microstructure Noise or Jumps in DGP

M	*Realized Volatility*	*Bipower Variation*	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,e,M}$
		Panel A: Interval $= \widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$		
		Sample Size = 100 Daily Realized Measure Observations		
72	0.233	0.253	0.533	0.396
144	0.260	0.270	0.313	0.366
288	0.253	0.223	0.306	0.340
576	0.236	0.240	0.246	0.233
		Sample Size = 300 Daily Realized Measure Observations		
72	0.146	0.173	0.576	0.436
144	0.173	0.153	0.383	0.336
288	0.156	0.136	0.243	0.233
576	0.143	0.110	0.223	0.193
		Sample Size = 500 Daily Realized Measure Observations		
72	0.143	0.133	0.633	0.496
144	0.123	0.096	0.420	0.386
288	0.086	0.093	0.313	0.270
576	0.086	0.120	0.160	0.236
		Panel B: Interval $= \widehat{\mu}_{IV_t} + 0.25\widehat{\sigma}_{IV_t}$		
		Sample Size = 100 Daily Realized Measure Observations		
72	0.156	0.220	0.543	0.463
144	0.163	0.170	0.316	0.410
288	0.196	0.163	0.263	0.296
576	0.166	0.183	0.210	0.193
		Sample Size = 300 Daily Realized Measure Observations		
72	0.160	0.153	0.726	0.556
144	0.143	0.110	0.496	0.456
288	0.143	0.113	0.320	0.306
576	0.070	0.086	0.220	0.216
		Sample Size = 500 Daily Realized Measure Observations		
72	0.126	0.153	0.840	0.716
144	0.083	0.096	0.603	0.560
288	0.103	0.086	0.376	0.360
576	0.076	0.100	0.206	0.310
		Panel C: Interval $= \widehat{\mu}_{IV_t} + 0.50\widehat{\sigma}_{IV_t}$		
		Sample Size = 100 Daily Realized Measure Observations		
72	0.153	0.190	0.590	0.500
144	0.106	0.140	0.416	0.423
288	0.166	0.146	0.216	0.296
576	0.130	0.146	0.183	0.203
		Sample Size = 300 Daily Realized Measure Observations		
72	0.116	0.103	0.876	0.750
144	0.123	0.113	0.620	0.553
288	0.113	0.100	0.340	0.286
576	0.120	0.113	0.190	0.230
		Sample Size = 500 Daily Realized Measure Observations		
72	0.126	0.143	0.970	0.883
144	0.130	0.136	0.716	0.690
288	0.140	0.136	0.446	0.420
576	0.160	0.136	0.203	0.286

* Notes: Entries denote rejection frequencies based on the construction of $G_{T,M}(u_1, u_2)$ values are compared with 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 *Realized Volatility*, *Bipower Variation*, $\widehat{RV}_{t,l,M}$ and $\widetilde{RV}_{t,e,M}$), for various different values of M, and for various daily sample sizes. 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 $\widehat{\mu}_{IV_t}$ and $\widehat{\sigma}_{IV_t}$ are the mean and standard error of the pseudo true data, and $\beta = \{0.125, 0.25, 0.50\}$. See Section 5 for further details.

Table 2: Conditional Confidence Interval Accuracy Assessment: Level Experiments

Case II: Microstructure Noise in DGP

M	*Realized Volatility*	*Bipower Variation*	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,e,M}$
		Panel A: Interval $= \widehat{\mu}_{IV_t} + 0.125\widehat{\sigma}_{IV_t}$		
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(3*1152)^{-1})$		
72	0.280	0.270	0.553	0.373
144	0.256	0.296	0.416	0.340
288	0.420	0.416	0.326	0.293
576	0.896	0.923	0.233	0.256
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(4*1152)^{-1})$		
72	0.246	0.280	0.556	0.376
144	0.236	0.263	0.393	0.316
288	0.290	0.303	0.300	0.306
576	0.653	0.683	0.263	0.296
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(5*1152)^{-1})$		
72	0.233	0.276	0.556	0.386
144	0.270	0.233	0.380	0.320
288	0.273	0.293	0.293	0.300
576	0.513	0.563	0.280	0.300
		Panel B: Interval $= \widehat{\mu}_{IV_t} + 0.25\widehat{\sigma}_{IV_t}$		
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(3*1152)^{-1})$		
72	0.163	0.180	0.516	0.410
144	0.160	0.170	0.403	0.363
288	0.213	0.240	0.316	0.300
576	0.830	0.840	0.200	0.226
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(4*1152)^{-1})$		
72	0.183	0.183	0.530	0.426
144	0.133	0.136	0.423	0.323
288	0.193	0.170	0.340	0.313
576	0.500	0.540	0.216	0.240
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(5*1152)^{-1})$		
72	0.166	0.160	0.513	0.433
144	0.120	0.143	0.416	0.343
288	0.163	0.173	0.323	0.313
576	0.343	0.366	0.226	0.240
		Panel C: Interval $= \widehat{\mu}_{IV_t} + 0.50\widehat{\sigma}_{IV_t}$		
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(3*1152)^{-1})$		
72	0.113	0.170	0.616	0.510
144	0.126	0.126	0.400	0.400
288	0.186	0.166	0.246	0.273
576	0.713	0.763	0.183	0.196
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(4*1152)^{-1})$		
72	0.120	0.153	0.613	0.490
144	0.203	0.163	0.403	0.380
288	0.190	0.160	0.260	0.293
576	0.386	0.403	0.183	0.190
		Noise $=$ *i.i.d.* $\mathrm{N}(0,(5*1152)^{-1})$		
72	0.133	0.170	0.616	0.496
144	0.163	0.143	0.386	0.396
288	0.183	0.160	0.263	0.310
576	0.266	0.310	0.190	0.196

* 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	*Realized Volatility*	*Bipower Variation*	$\widehat{RV}_{t,l,M}$	$\widetilde{RV}_{t,e,M}$
		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*		
72	0.373	0.336	0.620	0.583
144	0.333	0.276	0.596	0.510
288	0.290	0.286	0.390	0.456
576	0.280	0.270	0.453	0.460
		One i.i.d. $\mathrm{N}(0, 5*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.370	0.366	0.630	0.586
144	0.346	0.283	0.613	0.530
288	0.290	0.300	0.386	0.463
576	0.306	0.273	0.426	0.463
		One i.i.d. $\mathrm{N}(0, 10*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.383	0.353	0.620	0.610
144	0.363	0.303	0.616	0.523
288	0.320	0.276	0.403	0.466
576	0.330	0.266	0.480	0.476
		Panel B: Interval $= \widehat{\mu}_{IV_t} + 0.25\widehat{\sigma}_{IV_t}$		
		One i.i.d. $\mathrm{N}(0, 3*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.383	0.373	0.703	0.686
144	0.356	0.306	0.610	0.593
288	0.293	0.273	0.470	0.526
576	0.246	0.226	0.460	0.486
		One i.i.d. $\mathrm{N}(0, 5*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.393	0.396	0.720	0.656
144	0.346	0.310	0.630	0.630
288	0.310	0.263	0.490	0.563
576	0.273	0.216	0.463	0.520
		One i.i.d. $\mathrm{N}(0, 10*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.410	0.410	0.700	0.686
144	0.363	0.320	0.690	0.620
288	0.313	0.260	0.500	0.580
576	0.306	0.223	0.496	0.550
		Panel C: Interval $= \widehat{\mu}_{IV_t} + 0.50\widehat{\sigma}_{IV_t}$		
		One i.i.d. $\mathrm{N}(0, 3*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.526	0.416	0.846	0.840
144	0.420	0.296	0.770	0.783
288	0.360	0.240	0.636	0.750
576	0.346	0.240	0.566	0.680
		One i.i.d. $\mathrm{N}(0, 5*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.546	0.443	0.890	0.863
144	0.440	0.300	0.806	0.806
288	0.410	0.270	0.656	0.766
576	0.360	0.233	0.573	0.706
		One i.i.d. $\mathrm{N}(0, 10*0.64*\widehat{\mu}_{IV_t})$ *Jump Every 5 Days*		
72	0.576	0.483	0.883	0.883
144	0.480	0.353	0.826	0.853
288	0.423	0.270	0.713	0.796
576	0.390	0.223	0.626	0.740

* Notes: See notes to Table 2.

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

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

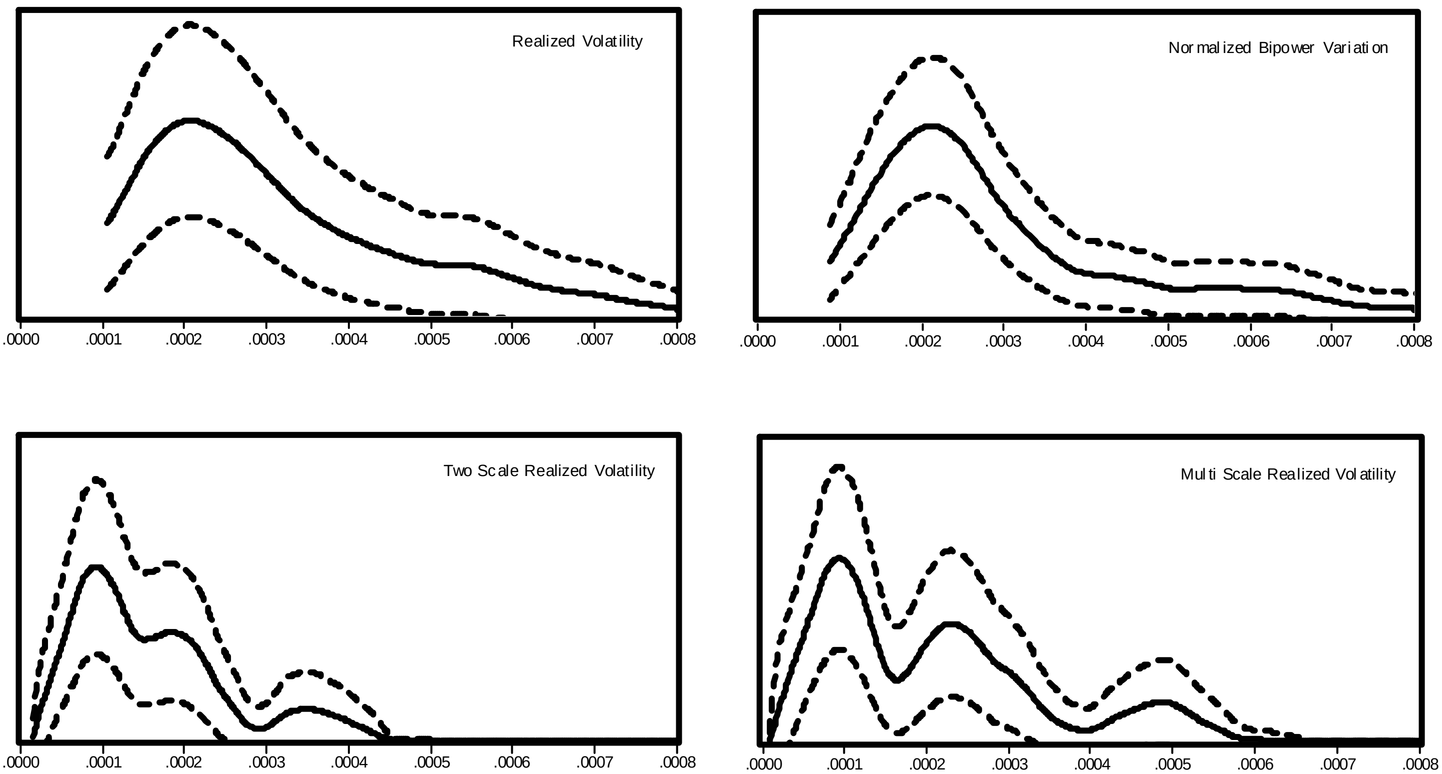

Figure 2: Predictive Conditional Densities for Intel Integrated Volatility Based on Various Realized Measures

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

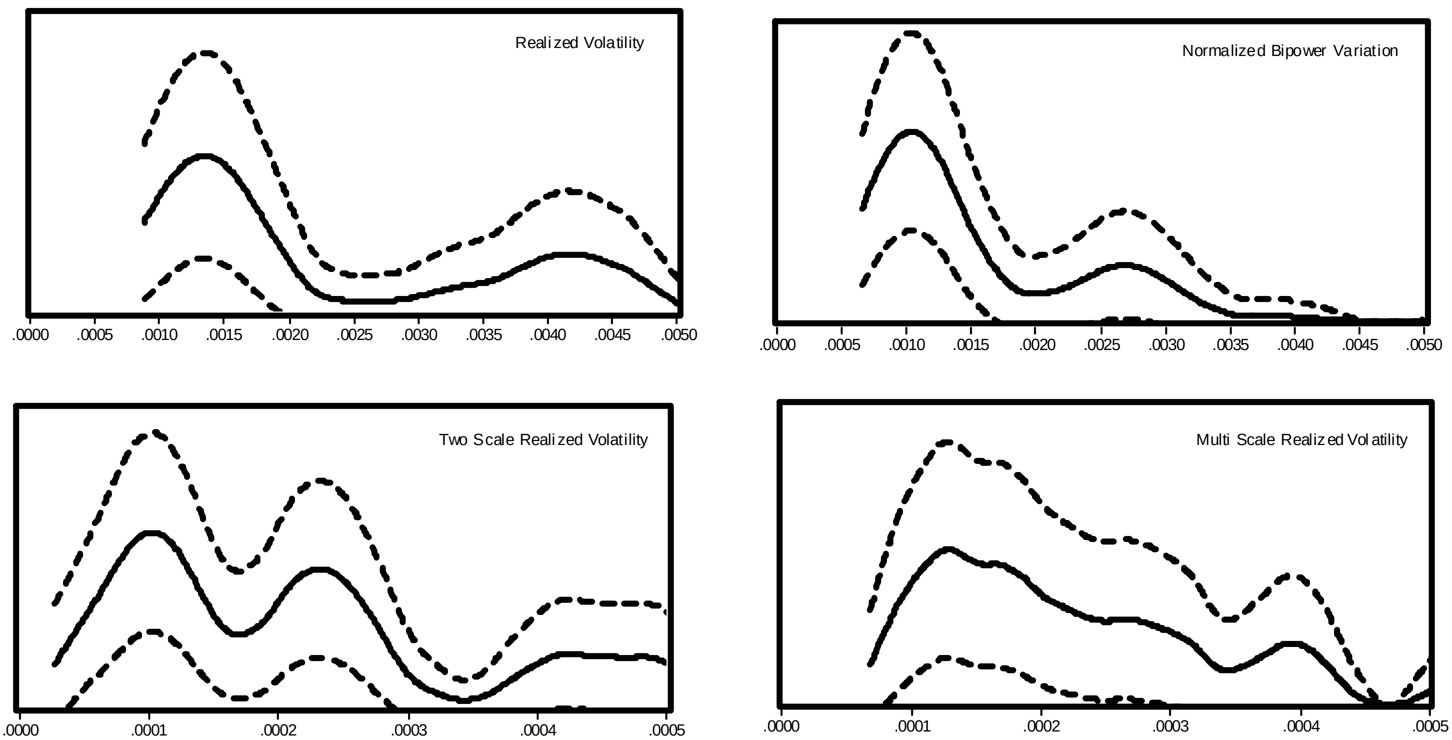

Figure 3: Predictive Conditional Densities for Intel Integrated Volatility Based on Various Realized Measures

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

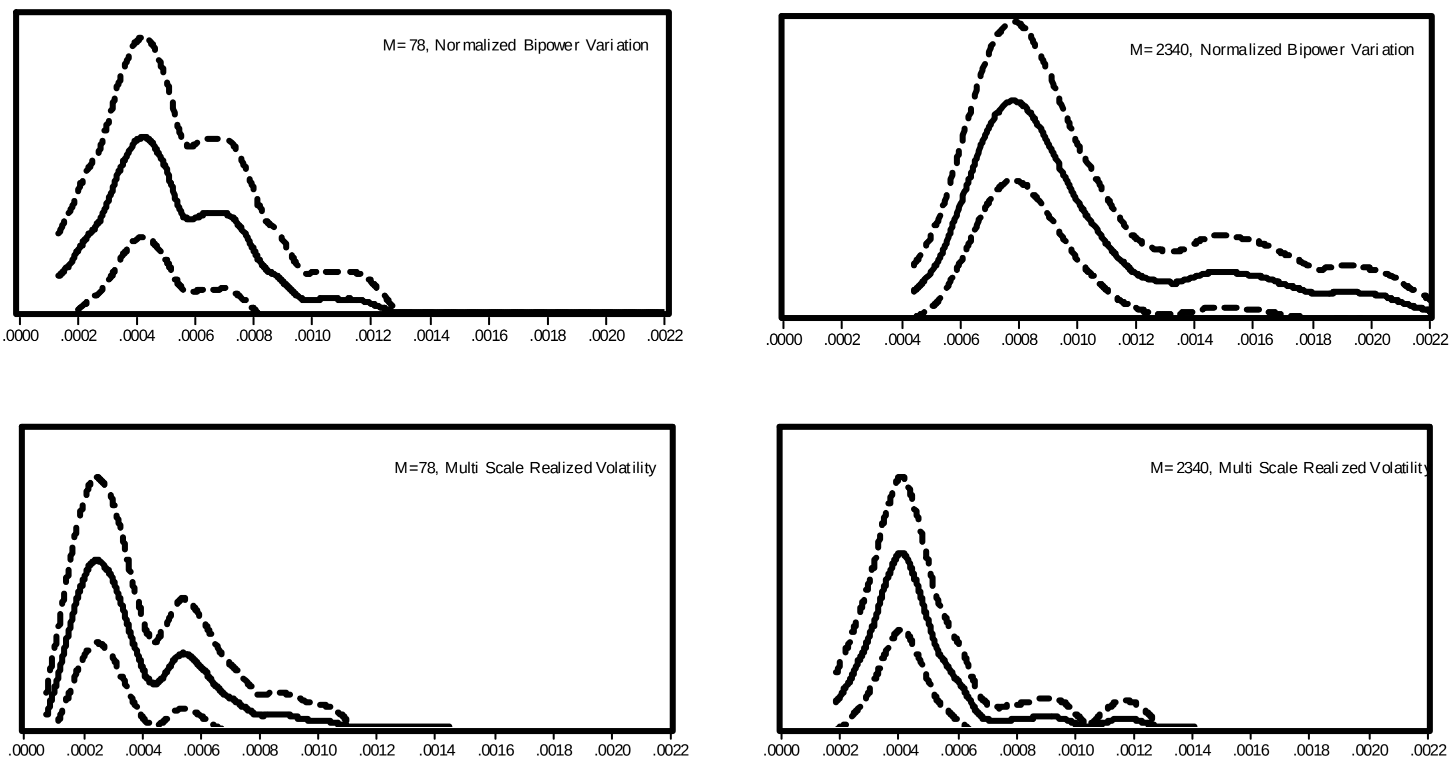

Figure 4: Predictive Conditional Densities for Intel Logged Integrated Volatility Based on Various Logged Realized Measures

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