# Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures*

Valentina Corradi[†]
Queen Mary, University of London

Walter Distaso[‡]
Queen Mary, University of London

Norman R. Swanson[§]
Rutgers University

July 2005

*Keywords.* Diffusions, integrated volatility, realized volatility measures, kernels, microstructure noise.

*JEL classification.* C22, C53, C14.

*We thank Yacine Aït-Sahalia, Tim Bollerslev, Christian Gourieroux, Nour Meddahi, Victoria Zinde Walsh and participants at the Cirano-Cireq Conference on "Forecasting in Macroeconomics and Finance", the SETA conference on "Recent Developments in Financial Econometrics", the conference "Changing Structures in International and Financial Markets and the Effects of Financial Decision Making" in Venice, for helpful comments and suggestions. 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.

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

**Abstract**

The main objective of this paper is to propose a feasible, model free estimator of the predictive density of integrated volatility. In this sense, we extend recent papers by Andersen, Bollerslev, Diebold and Labys (2003), and by Andersen, Bollerslev and Meddahi (2004, 2005), who address the issue of pointwise prediction of volatility via ARMA models, based on the use of realized volatility. Our approach is to use a realized volatility measure to construct a non parametric (kernel) estimator of the predictive density of daily volatility. We show that, by choosing an appropriate realized measure, one can achieve consistent estimation, even in the presence of jumps in prices and microstructure noise. More precisely, we establish that four well known realized measures, i.e. realized volatility; bipower variation, and two measures robust to microstructure noise, satisfy the conditions required for the uniform consistency of our estimator. Furthermore, we outline an alternative simulation based approach to predictive density construction. Finally, we carry out a simulation experiment in order to assess the accuracy of our estimators, and provide an empirical illustration that underscores the importance of using microstructure robust measures when using high frequency data.

# 1 Introduction

In a recent paper, Andersen, Bollerslev, Diebold and Labys (2003) suggest a novel, model free, approach for forecasting daily volatility. They advocate the use of simple, reduced form time series models for realized volatility, where the latter is constructed by summing up intradaily squared returns. The predictive ability of a given model is measured via the $R^2$ from the autoregressive or ARMA models constructed using (the log of) realized volatility. Their findings suggest that these ARMA based forecasts for realized volatility outperform most of the volatility models commonly used by practitioners, such as different varieties of GARCH models, for example. The rationale behind their approach is that, as the time interval between successive observations shrinks, realized volatility converges to the "true" daily volatility, whenever the underlying asset price is a continuous semimartingale. Although tick by tick and ultra high frequency data are now available, they are often contaminated by microstructure noise; therefore, in order to account for this potential problem, volatility has typically been constructed using 5 minutes interval returns, say, or even lower frequency observations. Hence, these reduced form time series forecasts for realized volatility imply a loss in efficiency relative to the infeasible optimal forecasts for the daily volatility process, based on the entire volatility path. For the class of eigenfunction stochastic volatility models of Meddahi (2001), an analytical expression for such loss in efficiency is provided by Andersen, Bollerslev and Meddahi (2004). In particular, they show that the error associated with realized volatility induces a downward bias in the estimated degree of predictability obtained via the $R^2$ approach mentioned above. To overcome this issue, Andersen, Bollerslev and Meddahi (2005) develop a general model free, feasible procedure to compute the adjusted $R^2$ used in model evaluation. All of the papers mentioned above are concerned with pointwise prediction of volatility via ARMA models based on realized measures. On the other hand, there are situations in which interest may focus on predictive conditional densities, as such densities yield information not only on the conditional mean of volatility, but also on all conditional aspects of the predictive distribution.

The main objective of this paper is to propose a feasible, model free estimator of the conditional predictive density of integrated volatility. From Meddahi (2003), we know that, within the context of eigenfunction stochastic volatility models, integrated volatility follows an $ARMA(p, p)$ structure, where $p$ denotes the number of eigenfunctions. However, we only have a complete characterization of the autoregressive part of the model. Furthermore, we do not know the marginal distribution of the innovation. For these reasons, we cannot exploit the ARMA representation in order to construct predictive densities for integrated volatility. Thus, we need to follow a different route. Our approach is to construct a kernel estimator of the conditional density of a given realized volatility measure, conditional on recent observed values of the realized measure itself. We provide general conditions on the measurement error that differentiates the realized measure from integrated volatility, in terms of its moment and covariance structure. Given these conditions, we define a sequence of bandwidth parameters under which the kernel estimator of the conditional density is uniformly consistent. We also provide a uniform rate of convergence, which depends on the bias and variance of the kernel estimator, as well as on the measurement error. Finally, we derive the relative rate, in terms of the number of days, $T$, at which the bandwidth parameter and the moments of the measurement error have to approach zero, in order to ensure that all three components (bias, variance and contribution of measurement error) approach zero at the same speed. Also, we show that four well known realized measures (realized volatility; bipower variation, Barndorff-Nielsen and Shephard, 2004b,c; and the robust subsampled realized volatility measures of (i) Zhang, Mykland and Aït-Sahalia, 2004 and (ii) Aït-Sahalia, Mykland and Zhang, 2005b, Zhang, 2004 and Barndorff-Nielsen, Hansen, Lunde and Shephard, 2004) satisfy the conditions on the measurement error required for the uniform consistency of the estimator. This means that we can provide a feasible model free estimator of the conditional predictive density of integrated volatility even in presence of jumps or microstructure noise.

Suppose that we knew the data generating process for the instantaneous volatility. While this information suffices to characterize the autoregressive structure of the integrated volatility process, often it does not suffice to recover the "entire" data generating process. Nevertheless, in this case we can construct a kernel density estimator using the integrated volatility values simulated under the null model (and "evaluated" at the estimated parameters) instead of using a realized measure. Under mild regularity conditions, and if the null model is correct, as the sample size and the number of simulations grow at an appropriate rate, the conditional density based on kernel estimators of simulated volatility converges to the "true" conditional density of integrated volatility. A natural question is whether there is some advantage, in terms of a faster rate of convergence, in using simulated volatility rather than realized measures. We show that the answer to this question depends on the relative rate at which the number of intradaily observations, $M$, grows, relative to the number of days $T$, and on the specific realized measure used.

In order to evaluate the accuracy of our proposed estimator constructed using realized measures, we carry out a simulation experiment in which the pseudo true predictive density is compared with the one estimated using our methodology. This is done for various daily sample sizes and for a variety of different intraday data frequencies and for different data generating processes, including jumps and microstructure noise. As expected, our subsampled realized volatility measures yield substantially more accurate predictions than the other measures, when data are subject to microstructure noise. Furthermore, the predictive estimator is seen to perform quite well, overall, based on the examination of mean square error and related predictive error loss measures. Finally, we provide an empirical illustration that 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. Section 2 describes the model. Section 3 provides a uniform rate of convergence for the conditional density estimator based on a given realized measure. Section 4 provides a uniform rate of convergence for the conditional density estimator based on simulated integrated volatility, for the case in which we know the data generating process of the instantaneous volatility process. Section 5 provides conditions under which realized volatility, bipower variation and the microstructure robust measures of realized volatility satisfy the conditions on the measurement error that are required for the uniform consistency of the kernel estimator based on realized measures. Section 6 reports the results from our simulation experiment, and our empirical illustration is discussed in Section 7. Finally, Section 8 contains some concluding remarks. All proofs are gathered in the Appendix.

## 2 The Model

The observable state variable, $Y_t = \log S_t$, where $S_t$ denotes the price of a financial asset or the exchange rate between two currencies, is modelled as a jump diffusion process with constant drift term and variance term modelled as a measurable function of a latent factor, $h_t$, which is also generated by a diffusion process. Thus,

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

$W_{1,t}$ and $W_{2,t}$ refer to two independent Brownian motions and volatility is modelled according to the eigenfunction stochastic volatility model of Meddahi (2001), so that

$$\begin{aligned} \sigma_t^2 &= \psi(h_t) = \sum_{i=1}^{p} a_i \mathrm{P}_i(h_t) \\ \mathrm{d}h_t &= \mu(h_t, \boldsymbol{\theta})\mathrm{d}t + \sigma(h_t, \boldsymbol{\theta})\mathrm{d}W_{2,t}, \end{aligned} \tag{2}$$

for some $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\mathrm{P}_i(h_t)$ denotes the $i$-th eigenfunction of the infinitesimal generator $\mathcal{A}$ associated with the unobservable state variable $h_t$.[1] The pure jump process $\mathrm{d}z_t$ specified in (1) is such that

$$Y_t = mt + \int_0^t \sqrt{\sigma_s^2}\left(\sqrt{1-\rho^2}\mathrm{d}W_{1,s} + \rho\mathrm{d}W_{2,s}\right) + \sum_{i=1}^{N_t} c_i,$$

where $N_t$ is a finite activity counting process, and $c_i$ is a nonzero i.i.d. random variable, independent of $N_t$. As $N_t$ is a finite activity counting process, we confine our attention to models characterized by only a finite number jumps over any fixed time span.

As is customary in the literature on stochastic volatility models, the volatility process is assumed to be driven by (a function of) the unobservable state variable $h_t$. Rather than assuming an ad hoc

[1] The infinitesimal generator $\mathcal{A}$ associated with $h_t$ is defined by

$$\mathcal{A}\phi(h_t) \equiv \mu(h_t)\phi'(h_t) + \frac{\sigma^2(h_t)}{2}\phi''(h_t)$$

for any square integrable and twice differentiable function $\phi(\cdot)$. The corresponding eigenfunctions $\mathrm{P}_i(h_t)$ and eigenvalues $-\lambda_i$ are given by $\mathcal{A}\mathrm{P}_i(h_t) = -\lambda_i \mathrm{P}_i(h_t)$.

function for $\psi(\cdot)$, the eigenfunction stochastic volatility model adopts a more flexible approach. In fact $\psi(\cdot)$ is modeled as a linear combination of the eigenfunctions of $\mathcal{A}$ associated with $h_t$. Notice that the $a_i$'s are real numbers and that $p$ may be infinite. For normalization purposes, it is further assumed that $\mathrm{P}_0(h_t) = 1$ and that $\mathrm{var}(\mathrm{P}_i(h_t)) = 1$, for any $i \neq 0$. Also, when $p$ is infinite, we also require $\sum_{i=1}^{\infty} a_i < \infty$. The generality and embedding nature of the approach just outlined stems from the fact that any square integrable function $\psi(h_t)$ can be written as a linear combination of the eigenfunctions associated with the state variable $h_t$. As a result, most of the widely used stochastic volatility models can be derived as special cases of the general eigenfunction stochastic volatility model. For more details on the properties of these models, see Meddahi (2001, 2003). Finally, notice that we have assumed a constant drift term.[2]

Following the widespread consensus that transaction data occurring in financial markets are contaminated by measurement errors, we assume that there are a total of $MT$ observations, consisting of $M$ intradaily observations for $T$ days. Namely, define

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

where

$$\epsilon_{t+j/M} \sim \text{i.i.d.}(0, \nu) \text{ and } \mathrm{E}(\epsilon_{t+j/M} Y_{s+i/M}) = 0 \text{ for all } t, s, j, i. \tag{3}$$

Thus, we allow for the possibility that the observed transaction price can be decomposed into the efficient one plus a "noise" due to measurement error, which captures generic microstructure effects.

The microstructure noise is assumed to be identically and independently distributed and independent of the underlying prices. This is consistent with the model considered by Aït-Sahalia, Mykland and Zhang (2005a), Zhang, Mykland and Aït-Sahalia (2004), Bandi and Russell (2004, 2005).[3] Needless to say, when $\nu = 0$, then $\epsilon_{t+j/M} = 0$ (almost surely), and therefore $X_{t+j/M} = Y_{t+j/M}$.

The daily integrated volatility process at day $t$ is defined as

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

Since $IV_t$ is not observable, different realized measures, based on the sample $X_{t+j/M}$, $t = 1, \ldots, T$ and $j = 1, \ldots, M$, are used as proxies for $IV_t$. The realized measure, say $RM_{t,M}$, is a noisy measure of the true integrated volatility process; in fact

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

[2] This is in line with Bollerslev and Zhou (2002), who assume a zero drift term and justify this with the fact that there is very little predictive variation in the mean of high frequency returns, as supported the empirical findings of Andersen and Bollerslev (1997). Indeed, the test statistics suggested below do not require the knowledge of the drift term. However, some of the proofs make use of the fact that the drift is constant.

[3] Recently, Aït-Sahalia, Mykland and Zhang (2005b), and Barndorff-Nielsen, Hansen, Lunde and Shephard (2004) allow for some dependence in the microstructure noise, while Awartani, Corradi and Distaso (2004) allow for correlation between the underlying price process and the microstructure noise.

where $N_{t,M}$ denotes the measurement error associated with the realized measure $RM_{t,M}$. Note that, in the case where $\nu > 0$, any realized measure of integrated volatility is contaminated by two measurement errors, given that the realized measure is constructed using contaminated data.

In the sequel, we first construct functionals of kernel estimator of conditional densities based on realized measure. Then, we provide primitive conditions on the measurement error $N_{t,M}$, in terms of its moments structure, ensuring that the kernel conditional density estimators based on realized measures are consistent for the conditional density of integrated volatility; also, we provide a uniform rate. Finally, we adapt the given primitive conditions on $N_{t,M}$ to the four considered realized measures of integrated volatility: namely,

(a) realized volatility, defined as:

$$RV_{t,M} = \sum_{j=1}^{M-1} \left( X_{t+(j+1)/M} - X_{t+j/M} \right)^2 ; \tag{5}$$

(b) normalized bipower variation, given by:

$$(\mu_1)^{-1} BV_{t,M} = (\mu_1)^{-1} \frac{M}{M-1} \sum_{j=2}^{M-1} \left| X_{t+(j+1)/M} - X_{t+j/M} \right| \left| X_{t+j/M} - X_{t+(j-1)/M} \right| , \tag{6}$$

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

(c) a microstructure robust subsampled based realized volatility measure, $\widehat{RV}_{t,l,M}$, suggested by Zhang, Mykland and Ait-Sahalia (2004), defined as

$$\widehat{RV}_{t,l,M} = RV_{t,l,M}^{avg} - 2l \widehat{\nu}_{t,M}, \tag{7}$$

where

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

and

$$RV_{t,l,M}^{avg} = \frac{1}{B} \sum_{b=0}^{B-1} RV_{t,l}^{b} = \frac{1}{B} \sum_{b=0}^{B-1} \sum_{j=b+1}^{M-(B-b-1)} \left( X_{t+\frac{jB}{M}} - X_{t+\frac{(j-1)B}{M}} \right)^2 . \tag{8}$$

Here $Bl \cong M$, $l = O(M^{1/3})$, $l$ denotes the subsample size and $B$ the number of subsamples. The logic underlying (7) is the following: first construct $B$ realized volatility measures using $l$ non overlapping subsamples, then take an average of this $B$ realized volatility measures and correct this average by an estimator of the bias term due to market microstructure, where the bias estimator is constructed using a finer grid;

(d) another microstructure robust subsampled based realized volatility, $\widetilde{RV}_{t,e,M}$, which has been proposed by Zhang (2004), Aït-Sahalia, Mykland and Zhang (2005b), and Barndorff-Nielsen, Hansen, Lunde and Shephard (2004). Define,

$$\begin{aligned}\widetilde{RV}_{t,e,M} &= \sum_{i=1}^{e} a_i \widetilde{RV}_{t,e_i,M} + \frac{RV_{t,M}}{M} \\ &= \sum_{i=1}^{e} a_i \frac{1}{e_i} \left( \sum_{j=1}^{M-e_i} \left( X_{t+\frac{j+e_i}{M}} - X_{t+\frac{j}{M}} \right)^2 \right) + \frac{RV_{t,M}}{M},\end{aligned} \tag{9}$$

so that $\widetilde{RV}_{t,e,M}$ is a linear weighted combination of $e$ realized volatilities compute over non-overlapping subsamples of $e_i$ observations each, plus a bias correction term. For $e_i = i$,

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

Zhang (2004) has shown that if $\sqrt{M}/e \to \pi$, $0 < \pi < \infty$, then $M^{1/4}\left(\widetilde{RV}_{t,e,M} - IV_t\right)$ is $O_P(1)$, and, in the case of finite time span, satisfies a central limit.

In particular, for each considered realized measure we will provide regularity conditions about the relative speed at which $T, M$ have to go to infinity, for the asymptotic validity of the associated specification test for integrated volatility.

# 3 A Predictive Density Estimator for Volatility Based on the Use of Realized Volatility Measures

Our objective is to construct a nonparametric estimator of the density of integrated volatility, conditional on a given realized volatility measure actually observed at time $T$. Define the conditional density kernel estimator based on realized measure as:[4]

$$\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) = \frac{\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{RM_{t+1,M}-x}{\xi_{2,T}}, \frac{RM_{t,M}-RM_{T,M}}{\xi_{2,T}}\right)}{\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right)}. \tag{10}$$

Hereafter, let's define $f_{IV_{T+1}|IV_T}(\cdot|\cdot)$ and $f_{IV_T}(\cdot)$ the conditional density of $IV_{T+1}$ given $IV_T$ and the marginal density of $IV_T$, respectively. Recall that $IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s = \int_{t-1}^{t} \psi(h_s)\mathrm{d}s$, and note that, because of assumption A2 below, $IV_t$ is a strictly stationary process, and so $f_{IV_{T+1}|IV_T}(\cdot|\cdot) = f_{IV_{t+1}|IV_t}(\cdot|\cdot)$ and $f_{IV_T}(\cdot) = f_{IV_t}(\cdot)$, for $t = 1, 2, \ldots, T$.

In the sequel, we will need the following assumptions.

**Assumption A1:** There is a sequence $b_M$, with $b_M \to \infty$ as $M \to \infty$, such that, uniformly in $t$,

[4] In the sequel, we consider one step ahead predictive densities; however, the case of $\tau-$step ahead predictive densities, for $\tau$ finite, can be treated in an analogous manner.

(i) $\mathrm{E}\left(N_{t,M}\right) = O\left(b_M^{-1}\right),$

(ii) $\mathrm{E}\left(\left|N_{t,M}\right|^k\right) = O\left(b_M^{-k/2}\right)$, for $k = 1, \ldots, 16$,

(iii) either

(a) $N_{t,M}$ is strong mixing with size $-r$, where $r > 2$; or

(b) $\mathrm{E}\left(N_{t,M}N_{s,M}\right) = O\left(b_M^{-2}\right) + \alpha_{t-s}O\left(b_M^{-1}\right)$, where $\alpha_{t-s} = O\left(|t-s|^{-2}\right).$

**Assumption A2**: $h_t$ is a time reversible process.

**Assumption A3**: the spectrum of the infinitesimal generator operator $\mathcal{A}$ of $h_t$ is discrete, and denoted by $0 < \lambda_1 < \ldots < \lambda_i < \ldots < \lambda_N$, where $\lambda_i$ is the eigenvalue associated with the $i-th$ eigenfunction $\mathrm{E}_i(h_t)$.

**Assumption A4**:

(i) The kernel $\mathbf{K} : \mathbb{R} \to \mathbb{R}^+$, is a symmetric, nonnegative, continuous function with bounded support, twice differentiable in the interior of the support, and

$$\int \mathbf{K}(x)\mathrm{d}x = 1, \int xK(x)\mathrm{d}x = 0,$$

(ii) $f_{IV_T}(\cdot)$ and $f_{IV_{T+1}|IV_T}(\cdot|\cdot)$ are absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}$ and $\mathbb{R}^2$, respectively, are $\omega-$times continuously differentiable on $\mathbb{R}$ and $\mathbb{R}^2$, with $\omega \geq 2$, are bounded and have bounded first derivatives.

Note that in A4(i) we have required that the kernel is a function with a bounded support, thus ruling out a gaussian kernel for example. The reason is the following. Realized measures are by definition positive, then if we used an unbounded kernel the estimated density would result downward biased, as some weight would be given also to negative observations. On the other hand, by employing a bounded kernel, it suffices to discard those values $x$ and $RM_{T,M}$, in the definition of (10), which are too close to the lower bound in the support. Broadly speaking, this means that we rule out prediction of the very left tail of the volatility distribution. An alternative route would be the use of asymmetric kernels. However, symmetry of the kernel is a key ingredient in the derivation of the uniform consistency results below.

Then, we can state the following.

**Theorem 1.** *Let assumptions A1(ii), A2-A4 hold and let $M/T^{1/2} \to \infty$. Then, for all $RM_{T,M}$ such that $\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right) > d_T$, with $d_T = O_P(1)$, uniformly in $x$,*

$$\sup_{\left\{RM_{T,M}:\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right)>d_T\right\}} \left|\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right|$$

$$
\begin{aligned}
&= O_P(b_M^{-1/2}\xi_{2,T}^{-3}d_T^{-1}) + O_P(b_M^{-1/2}\xi_{1,T}^{-2}d_T^{-2}) + O_P(T^{-1/2}\xi_{2,T}^{-2}d_T^{-1}) \\
&\quad + O_P(T^{-1/2}\xi_{1,T}^{-1}d_T^{-2}) + O\left(\xi_{2,T}^{2}d_T^{-1}\right) + O\left(\xi_{1,T}^{2}d_T^{-2}\right). \qquad (11)
\end{aligned}
$$

Note that the first two terms above are due to the measurement error due to the fact that we use a realized measure instead of the "true integrated volatility", the second two terms reflect the variance component and the last two terms the bias component. In the proposition below we shall provide conditions on the relative rate of growth of $\xi_{1,T}, \xi_{2,T}, d_T, b_M$ relative to $T$, under which (i) the order of magnitude of the second, fourth and sixth terms is smaller than the order of the other terms, so that the total error component due to the estimation of the marginal density is negligible; (ii) the first, third and fifth terms approach zero uniformly in $x$, so that we have a uniform rate; (iii) the measurement error component is of the same order or of a smaller order than the variance and the bias component, so that that it does not "slow down" the convergence to the true conditional density.

**Proposition 1.** *Let $\xi_{1,T} = cT^{-\phi_1}$, $\xi_{2,T} = cT^{-\phi_2}$, $b_M = cT^{\psi}$, $d_T = cT^{-\delta}$, $\psi, \delta \geq 0$, $\phi_1 > \phi_2$, $\psi > 1/2$, with c denoting a generic positive constant, and let assumptions of Theorem 1 hold.*
*(a) If (i) $\phi_2 < \phi_1 < 3/2\phi_2$, and (ii) $\delta < \min\{3\phi_2 - 2\phi_1, 2(\phi_1 - \phi_2)\}$, then*

$$
\begin{aligned}
&\sup_{\left\{RM_{T,M}: \frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right) > d_T\right\}} \left|\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| \\
&= O_P(b_M^{-1/2}\xi_{2,T}^{-3}d_T^{-1}) + O_P(T^{-1/2}\xi_{2,T}^{-2}d_T^{-1}) + O_P(\xi_{2,T}^{-2}d_T^{-1}); \qquad (12)
\end{aligned}
$$

*(b) if, in addition to (i), (ii) above, (iii) $\phi_2 < \frac{1}{6}$ and (iv) $\psi > 6\phi_2 + 2\delta$, then*

$$
\begin{aligned}
&\sup_{\left\{RM_{T,M}: \frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right) > d_T\right\}} \left|\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| \\
&= o_P(1);
\end{aligned}
$$

*(c) if, in addition to (i)-(iv) above, (v) $\phi_2 = \frac{1}{8}$ and (v) $\psi \geq 5/4$, then*

$$
\begin{aligned}
&\sup_{\left\{RM_{T,M}: \frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right) > d_T\right\}} \left|\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| \\
&= O_P(T^{-1/4}d_T^{-1}).
\end{aligned}
$$

First, from (a) above note that the error due to the variance component, i.e. the second term on the right hand side of (12), is of a larger order of probability than the typical one occurring in the pointwise case (see, e.g., Bosq, Ch. 2, 1998). In fact, in the pointwise case we would have $O_P\left(T^{-1/2}\xi_{2,T}^{-1}\right)$, instead of $O_P\left(T^{-1/2}\xi_{2,T}^{-2}\right)$. The slower rate is due to the need of deriving a result which holds uniformly on $\mathbb{R}^+$; it comes from a proof based on the Fourier transform of the kernel, firstly introduced by Bierens (1983) for regression functions with strong mixing processes and then extended to the case of generic derivatives of density and/of regression functions for general near epoch dependent, possibly heterogeneous, processes by Andrews (1990,1995).

Second, from (b) and (c) above we see that the faster the trimming parameter $d_T$ approaches zero, the slower is the rate of convergence of the estimator of the conditional density. In fact, if $\delta = 0$ we have the fastest rate; however, in this case we have to "give up" constructing a predictive density for all values of $RM_{t,M}$ such that $\frac{1}{T\xi_{1,T}} \sum_{t=1}^{T-1} K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right) < c$.

Third, we note that when $\phi_2 = \frac{1}{8}$ and $\psi = \frac{5}{4}$, all the error components, due to measurement error, variance and bias, are of the same order and get "pseudo" optimal rate $O_P\left(T^{-\frac{1}{4}} d_T^{-1}\right)$. This requires that $b_M$ grows at a rate faster than $T$ (later in the paper it will be shown that $b_M = M$ for realized volatility and for bipower variation, while $b_M = M^{1/3}$ or $M^{1/2}$ for the two microstructure robust versions of realized volatility).

In practice, we have $M$ intraday observations and and $T$ days, and once a realized measure has been chosen, we know how $b_M$ grows with $M$. Thus, in practice we have to fix $\psi$, as it is implied by our measure. First, it is immediate to see that, whenever $\psi < \frac{5}{4}$, then the measurement error term converges slower than the variance term; in fact the former converges at rate $T^{-1/2\psi} T^{3\phi_2} d_T^{-1}$, while the latter converges at rate $T^{-1/2} T^{2\phi_2} d_T^{-1}$. Therefore, for a given $\psi$, we define that value for $\phi_2$ which equalizes the order of magnitude of the measurement error term and of the bias term.

Thus, we want to find $\phi_2$, such that $-1/2\psi + 3\phi_2 = -2\phi_2$; this gives

$$\phi_2 = \frac{\psi}{10},$$

which implies a uniform convergence at rate $T^{-\frac{2\psi}{10}}$.

## 4 A Predictive Density Estimator for Volatility Based on the Use of Simulated Daily Volatility

In this section we consider the case in which we know the model generating the instantaneous volatility process, though we do not know the closed form of the conditional density of the integrated volatility process. We proceed in the following way: for any value in the parameter space, we generate $S$ (instantaneous) volatility paths of $k$ days (with $k \geq 1$), using as initial value a draw from the invariant distribution, and construct the associated daily integrated volatility. Parameters can be estimated by the method of Simulated Generalized Method of Moments (SGMM), as in Corradi and Distaso (2004, Theorem 2). Then, we simulate $S$ paths of 2 days using the estimated parameters, again drawing the initial value from the invariant distribution. More formally: for any simulation $i = 1, \ldots, S$, for $j = 1, \ldots, N$ and for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we simulate the volatility paths of length $k+1$ using a Milstein scheme, i.e.

$$h_{i,j\frac{k+1}{N}}(\boldsymbol{\theta})$$

$$
\begin{aligned}
&= h_{i,(j-1)\frac{k+1}{N}}\left(\boldsymbol{\theta}\right) + \mu(h_{i,(j-1)\frac{k+1}{N}}\left(\boldsymbol{\theta}\right), \boldsymbol{\theta}) - \frac{1}{2}\sigma'(h_{i,(j-1)\frac{k+1}{N}}\left(\boldsymbol{\theta}\right), \boldsymbol{\theta})\sigma(h_{i,(j-1)\frac{k+1}{N}}\left(\boldsymbol{\theta}\right), \boldsymbol{\theta})\frac{k+1}{N} \\
&\quad + \sigma(h_{i,(j-1)\frac{k+1}{N}}\left(\boldsymbol{\theta}\right), \boldsymbol{\theta})\left(W_{i,j\frac{k+1}{N}} - W_{i,(j-1)\frac{k+1}{N}}\right) \\
&\quad + \frac{1}{2}\sigma'(h_{i,(j-1)\frac{k+1}{N}}\left(\boldsymbol{\theta}\right), \boldsymbol{\theta})\sigma(h_{i,(j-1)\frac{k+1}{N}}\left(\boldsymbol{\theta}\right), \boldsymbol{\theta})\left(W_{i,j\frac{k+1}{N}} - W_{i,(j-1)\frac{k+1}{N}}\right)^2, \qquad (13)
\end{aligned}
$$

where $\sigma'(\cdot)$ denotes the derivative of $\sigma(\cdot)$ with respect to its first argument, $\left\{W_{i,j\frac{k+1}{N}} - W_{i,(j-1)\frac{k+1}{N}}\right\}$ is $i.i.d.$ $\mathrm{N}(0, \frac{k+1}{N})$ and $h_{i,0}\left(\boldsymbol{\theta}\right)$ is drawn from the invariant distribution of the volatility process under the given model. As discussed above, $\sigma^2_{i,j\frac{k+1}{N}} = \psi(h_{i,j\frac{k+1}{N}})$. Now, for each $i$ it is possible to compute the simulated integrated volatility as:

$$
IV_{i,\tau,N}\left(\boldsymbol{\theta}\right) = \frac{1}{N/(k+1)} \sum_{j=1}^{N/(k+1)} \sigma^2_{i,\tau-1+j\frac{k+1}{N}}(\boldsymbol{\theta}), \quad \tau = 1, \ldots, k+1, \qquad (14)
$$

and

$$
\sigma^2_{i,\tau-1+jh}(\boldsymbol{\theta}) = \psi(h_{i,\tau-1+jh}\left(\boldsymbol{\theta}\right)).
$$

Also, averaging the quantity calculated in (14) over the number of simulations $S$ and over the length of the path $k+1$ yields respectively

$$
\overline{IV}_{S,\tau,N}\left(\boldsymbol{\theta}\right) = \frac{1}{S}\sum_{i=1}^{S} IV_{i,\tau,N}\left(\boldsymbol{\theta}\right),
$$

and

$$
\overline{IV}_{S,N}\left(\boldsymbol{\theta}\right) = \frac{1}{k+1}\sum_{\tau=1}^{k+1} \overline{IV}_{S,\tau,N}\left(\boldsymbol{\theta}\right).
$$

We are now in a position to define the set of moment conditions as

$$
\overline{\mathbf{g}}^*_{T,M} - \overline{\mathbf{g}}_{S,N}\left(\boldsymbol{\theta}\right) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{g}^*_{t,M} - \frac{1}{S}\sum_{i=1}^{S} \mathbf{g}_{i,N}\left(\boldsymbol{\theta}\right), \qquad (15)
$$

where $\mathbf{g}^*_{t,M}$ is defined as

$$
\mathbf{g}^*_{t,M} = \begin{pmatrix} RM_{t,M} \\ \left(RM_{t,M} - \overline{RM}_{T,M}\right)^2 \\ \left(RM_{t,M} - \overline{RM}_{T,M}\right)\left(RM_{t-1,M} - \overline{RM}_{T,M}\right) \\ \vdots \\ \left(RM_{t,M} - \overline{RM}_{T,M}\right)\left(RM_{t-k,M} - \overline{RM}_{T,M}\right) \end{pmatrix}, \qquad (16)
$$

$RM_{t,M}$ denotes the particular realized measure used, and $\overline{RM}_{T,M} = \sum_{t=1}^{T} RM_{t,M}$. Also

$$
\frac{1}{S}\sum_{i=1}^{S} \mathbf{g}_{i,N}\left(\boldsymbol{\theta}\right) = \begin{pmatrix} \frac{1}{S}\sum_{i=1}^{S} IV_{i,1,N}\left(\boldsymbol{\theta}\right) \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1,N}\left(\boldsymbol{\theta}\right) - \overline{IV}_{S,N}\left(\boldsymbol{\theta}\right)\right)^2 \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1,N}\left(\boldsymbol{\theta}\right) - \overline{IV}_{S,N}\left(\boldsymbol{\theta}\right)\right)\left(IV_{i,2,N}\left(\boldsymbol{\theta}\right) - \overline{IV}_{S,N}\left(\boldsymbol{\theta}\right)\right) \\ \vdots \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1,N}\left(\boldsymbol{\theta}\right) - \overline{IV}_{S,N}\left(\boldsymbol{\theta}\right)\right)\left(IV_{i,k+1,N}\left(\boldsymbol{\theta}\right) - \overline{IV}_{S,N}\left(\boldsymbol{\theta}\right)\right) \end{pmatrix}. \qquad (17)
$$

We can define the SGMM estimator as the minimizer of the quadratic form

$$\widehat{\boldsymbol{\theta}}_{T,S,M,N} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} (\overline{\mathbf{g}}^*_{T,M} - \overline{\mathbf{g}}_{S,N}(\boldsymbol{\theta}))'\mathbf{W}^{-1}_{T,M}(\overline{\mathbf{g}}^*_{T,M} - \overline{\mathbf{g}}_{S,N}(\boldsymbol{\theta})), \tag{18}$$

where $\mathbf{W}_{T,M}$ is defined as

$$\begin{aligned}\mathbf{W}_{T,M} =& \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{g}^*_{t,M} - \overline{\mathbf{g}}^*_{T,M}\right)\left(\mathbf{g}^*_{t,M} - \overline{\mathbf{g}}^*_{T,M}\right)' \\ &+ \frac{2}{T}\sum_{v=1}^{p_T} w_v \sum_{t=v+1}^{T}\left(\mathbf{g}^*_{t,M} - \overline{\mathbf{g}}^*_{T,M}\right)\left(\mathbf{g}^*_{t-v,M} - \overline{\mathbf{g}}^*_{T,M}\right)'.\end{aligned} \tag{19}$$

Also, define

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} (\overline{\mathbf{g}}^*_{\infty} - \overline{\mathbf{g}}_{\infty}(\boldsymbol{\theta}))'\mathbf{W}^{-1}_{\infty}(\overline{\mathbf{g}}^*_{\infty} - \overline{\mathbf{g}}_{\infty}(\boldsymbol{\theta})), \tag{20}$$

where $\overline{\mathbf{g}}^*_{\infty}$, $\overline{\mathbf{g}}_{\infty}(\boldsymbol{\theta})$ and $\mathbf{W}^{-1}_{\infty}$ are the probability limits, as $T$, $S$, $M$ and $N$ go to infinity, of $\overline{\mathbf{g}}^*_{T,M}$, $\overline{\mathbf{g}}_{S,N}(\boldsymbol{\theta})$ and $\mathbf{W}^{-1}_{T,M}$, respectively.

We can now construct kernel conditional density estimators based on the integrated volatility simulated under the estimated parameters. For $i = 1,\ldots,S$, define:

$$\begin{aligned}&\widehat{f}_{IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(u|RM_{T,M}) \\ =& \frac{\frac{1}{T\zeta_{2,T}^2}\sum_{i=1}^{S}\mathbf{K}\left(\frac{IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-x}{\zeta_{2,T}}, \frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\zeta_{2,T}}\right)}{\frac{1}{T\zeta_{1,T}}\sum_{i=1}^{S}K\left(\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\zeta_{1,T}}\right)}\end{aligned} \tag{21}$$

We also need the following further assumptions.

**Assumption A5**: The drift and variance functions $\mu(\cdot)$ and $\sigma(\cdot)$, as defined in (2), satisfy the following conditions:

(1a) $|\mu(f_r(\boldsymbol{\theta}_1),\boldsymbol{\theta}_1) - \mu(f_r(\boldsymbol{\theta}_2),\boldsymbol{\theta}_2)| \leq K_{1,r}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$,

$|\sigma(f_r(\boldsymbol{\theta}_1),\boldsymbol{\theta}_1) - \sigma(f_r(\boldsymbol{\theta}_2),\boldsymbol{\theta}_2)| \leq K_{2,r}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$,

for $0 \leq r \leq k+1$, where $\|\cdot\|$ denotes the Euclidean norm, any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, with $K_{1,r}$, $K_{2,r}$ independent of $\boldsymbol{\theta}$, and $\sup_{r\leq k+1} K_{1,r} = O_P(1)$, $\sup_{r\leq k+1} K_{2,r} = O_P(1)$.

(1b) $|\mu(f_{r,N}(\boldsymbol{\theta}_1),\boldsymbol{\theta}_1) - \mu(f_{r,N}(\boldsymbol{\theta}_2),\boldsymbol{\theta}_2)| \leq K_{1,r,N}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$ ,

$|\sigma(f_{r,N}(\boldsymbol{\theta}_1),\boldsymbol{\theta}_1) - \sigma(f_{r,N}(\boldsymbol{\theta}_2),\boldsymbol{\theta}_2)| \leq K_{2,r,N}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$, where $f_{r,N}(\boldsymbol{\theta}) = f_{\lfloor\frac{Nrh}{k+1}\rfloor}(\boldsymbol{\theta})$ and for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, with $K_{1,r,N}$, $K_{2,r,N}$ independent of $\boldsymbol{\theta}$, and $\sup_{r\leq k+1} K_{1,r,N} = O_P(1)$,

$\sup_{r\leq k+1} K_{2,r,N} = O_P(1)$, uniformly in $N$.

(2) $|\mu(x,\boldsymbol{\theta}) - \mu(y,\boldsymbol{\theta})| \leq C_1\|x-y\|$, $\quad |\sigma(x,\boldsymbol{\theta}) - \sigma(y,\boldsymbol{\theta})| \leq C_2\|x-y\|$,

where $C_1, C_2$ are independent of $\boldsymbol{\theta}$.

(3) $\sigma(\cdot)$ is three times continuously differentiable and $\psi(\cdot)$ is a Lipschitz-continuous function.

(4) $\mathrm{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} IV_{i,j,N}(\boldsymbol{\theta})\right)^{14} < \infty$, $\forall\, i$ and for $j = 1, 2$.

**Assumption A6**: $(\overline{\mathbf{g}}^*_\infty - \overline{\mathbf{g}}_\infty(\boldsymbol{\theta}^*))'\mathbf{W}_\infty^{-1}(\overline{\mathbf{g}}^*_\infty - \overline{\mathbf{g}}_\infty(\boldsymbol{\theta}^*)) < (\overline{\mathbf{g}}^*_\infty - \overline{\mathbf{g}}_\infty(\boldsymbol{\theta}))'\mathbf{W}_\infty^{-1}(\overline{\mathbf{g}}^*_\infty - \overline{\mathbf{g}}_\infty(\boldsymbol{\theta}))$, for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$.

**Assumption A7**:

(1) $\widehat{\boldsymbol{\theta}}_{T,S,M,N}$ and $\boldsymbol{\theta}^*$ are in the interior of $\boldsymbol{\Theta}$.

(2) $\overline{\mathbf{g}}_S(\boldsymbol{\theta})$ is twice continuously differentiable in the interior of $\boldsymbol{\Theta}$, where

$$\mathbf{g}_S(\boldsymbol{\theta}) = \frac{1}{S}\sum_{i=1}^{S} \mathbf{g}_i(\boldsymbol{\theta}), \tag{22}$$

where

$$\overline{\mathbf{g}}_S(\boldsymbol{\theta}) = \frac{1}{S}\sum_{i=1}^{S}\mathbf{g}_i(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{S}\sum_{i=1}^{S} IV_{i,1}(\boldsymbol{\theta}) \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right)^2 \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right)\left(IV_{i,2}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right) \\ \vdots \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right)\left(IV_{i,k+1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right) \end{pmatrix}, \tag{23}$$

and, for $\tau = 1, \ldots, k+1$,

$$IV_{i,\tau}(\boldsymbol{\theta}) = \int_{\tau-1}^{\tau} \sigma^2_{i,s}(\boldsymbol{\theta})\,\mathrm{d}s, \qquad \overline{IV}_S(\boldsymbol{\theta}) = \frac{1}{k+1}\sum_{\tau=1}^{k+1}\frac{1}{S}\sum_{i=1}^{S}\int_{\tau-1}^{\tau}\sigma^2_{i,s}(\boldsymbol{\theta})\,\mathrm{d}s.$$

(3) $\mathrm{E}(\partial\overline{\mathbf{g}}_1(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*})$ exists and is of full rank.

Note that A5-A7, together with some conditions on the relative rate of growth of $T, M, S, N$ ensure that $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N} - \boldsymbol{\theta}^\dagger\right) = O_P(1)$.

**Theorem 2.** *Let A1-A7 be satisfied. If, as $M, T, S, N \to \infty$, $T/b_M^2 \to 0$, $T/N^{(1-\delta)} \to 0$, $T/S \to 0$, $T^2/S \to \infty$, $\varsigma_{1,T}/T^{1/6} \to 0$, $\varsigma_{2,T}/T^{1/12} \to 0$, $p_T \to \infty$ and $p_T/T^{1/4} \to 0$, $d_T = O_P(1)$, then, uniformly in $x$,*

$$\begin{aligned} &\sup_{\left\{RM_{T,M}:\frac{1}{S\zeta_{1,T}}\sum_{i=1}^{S}K\left(\frac{IV_{i,1,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-RM_{T,M}}{\zeta_{1,T}}\right)>d_T\right\}} \left|\widehat{f}_{IV_{i,2,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)|IV_{i,1,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| \\ = \;& O_P(T^{-1/2}\zeta_{2,T}^{-3}d_T^{-1}) + O_P(T^{-1/2}\zeta_{1,T}^{-2}d_T^{-2}) + O\left(\zeta_{2,T}^2 d_T^{-1}\right) + O\left(\zeta_{1,T}^2 d_T^{-2}\right). \end{aligned} \tag{24}$$

Theorem 2 reports the uniform rate of convergence for the case where we construct kernel density estimators based on integrated volatility, simulated using a $\sqrt{T}$-consistent estimator for the parameters.

The variance term in this case is different from the one in which we observe the true volatility process, i.e. the second term of the right hand side of (11); in fact in this case the exponent of the bandwidth is lower, and the difference is due to the fact that estimated parameters are used.

Broadly speaking, the simulation error is negligible, as $N$, the reciprocal of the discrete interval in the path simulation, and $S$, the number of simulations, grow at a rate faster than $T$; in fact $N$ and $S$ can be set arbitrarily large. On the other hand, the fact that we simulate the volatility paths using $\sqrt{T}$ estimated parameters increases the probability order of the variance component, which is $T^{-1/2}\zeta_{2,T}^{-3}$ instead of $T^{-1/2}\zeta_{2,T}^{-2}$. This result may seem a little bit surprising; in fact, generally kernel estimators constructed using estimated residuals are asymptotically equivalent to those constructed using true errors. The key condition for this result is that the derivative of the kernel with respect to the parameter has mean zero. In this case, this is in general not true, hence the extra $\zeta_{2,T}^{-1}$ in the right hand side of (24). This fact has been already pointed out by Andrews (1995). Needless to say, if the model used to simulate the instantaneous volatility process is not correctly specified, then the conditional density estimator based on simulated integrated volatility is no longer consistent for the true conditional density.

**Proposition 2.** *Let the Assumptions of Theorem 2 hold. Also, let* $\zeta_{1,T} = T^{-\varphi_1}$, $\zeta_{2,T} = T^{-\varphi_2}$, $\varphi_1, \varphi_2 > 0$, *and* $d_T$ *be defined as in Proposition 1.*
*(a) If (i)* $\phi_2 < \phi_1 < 3/2\phi_2$, *and (ii)* $\delta < \min\{3\phi_2 - 2\phi_1, 2(\phi_1 - \phi_2)\}$, *then uniformly in* $x$,

$$\sup_{\left\{RM_{T,M}:\frac{1}{S\zeta_{1,T}}\sum_{i=1}^{S} K\left(\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\zeta_{1,T}}\right)>d_T\right\}} \left|\widehat{f}_{IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| = o_P(1).$$

*(b) If in addition to (i)-(ii) above,*

$$\varphi_2 = \frac{1}{10}$$

*then as* $T \to \infty$, *uniformly in* $x$,

$$\sup_{\left\{RM_{T,M}:\frac{1}{S\zeta_{1,T}}\sum_{i=1}^{S} K\left(\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\zeta_{1,T}}\right)>d_T\right\}} \left|\widehat{f}_{IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| = O_P\left(T^{-\frac{1}{5}}d_T^{-1}\right).$$

It is worthwhile to compare the uniform rates for kernel estimators based on realized measures with those based on simulated integrated volatility. This mainly depends on the relative rate at which $M$ and $T$ grow. Letting $b_M = T^{\psi}$, the speed at which the bandwidth parameter should approach zero is $\phi_2 = \frac{\psi}{10}$, which gives a rate equal to $T^{-\frac{2\psi}{10}} d_T^{-1}$. Therefore for $\psi = 1$, i.e. if $b_M$ grows as the same speed as $T$, then the uniform rate is the same, while for $\psi > 1$ (resp. $\psi < 1$), the uniform rate of convergence is faster for kernel estimators based on a realized measure (resp. simulated integrated volatility). This is quite intuitive, as the measurement error associated with the realized measure is of order $b_M^{-1}$, while parameter estimation error is of order $T^{-1/2}$.

# 5 Applications to Specific Volatility Realized Measures

Assumption A1 states some primitive conditions on the measurement error between integrated volatility and realized measure. Basically, A1(i) requires that the $k-$th moment of $|N_{t,M}|$, $k = 2, \ldots, 16$ is of order $b_M^{-k/2}$, condition which are used to prove the uniform convergence for kernel estimators based on realized measures. On the other hand, A1, stating conditions on the mean, higher moments and covariance structure of the measurement error, are required for the $\sqrt{T}-$consistency of the SGMM estimator, which is used in the proof of Theorem 2.
Realized volatility has been suggested as an estimator of integrated volatility by Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold and Labys (2001, 2003). When the (log) price process is a continuous semimartingale, then realized volatility is a consistent estimator of the increments of the quadratic variation (see e.g. Karatzas and Shreve, 1991, Ch.1). The relevant limit theory, under general conditions, also allowing for generic leverage effects, has been provided by Barndorff-Nielsen and Shephard (2004a), who have shown that, as $M \longrightarrow \infty$,

$$\sqrt{M}\left(RV_{\overline{T},M} - \int_0^{\overline{T}} \sigma_s^2 \mathrm{d}s\right) \xrightarrow{d} \mathrm{MN}\left(0, 2\int_0^{\overline{T}} \sigma_s^4 \mathrm{d}s\right),$$

for given $\overline{T}$.

**Proposition 3.** *Let* $\mathrm{d}z_t = 0$, *a.s. and* $\nu = 0$, *where* $\mathrm{d}z_t$ *and* $\nu$ *are defined in (1) and in (3), respectively. Then Assumption A1 holds with* $RM_{t,M} = RV_{t,M}$ *for* $b_M = O(M)$.

Bipower variation has been introduced by Barndorff-Nielsen and Shephard (2004b,c). Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2004) show that, as $M \longrightarrow \infty$,

$$\sqrt{M}\left(\mu_1^{-2} BV_{\overline{T},M} - \int_0^{\overline{T}} \sigma_s^2 \mathrm{d}s\right) \xrightarrow{d} \mathrm{MN}\left(0, 2.6090\int_0^{\overline{T}} \sigma_s^4 \mathrm{d}s\right).$$

**Proposition 4.** *Let* $\nu = 0$, *where* $\nu$ *is defined in (3). Then Assumption A1 holds with* $RM_{t,M} = BV_{t,M}$ *for* $b_M = O(M)$.

Thus, when using bipower variation (robust to the presence of jumps) instead of realized volatility, there is no cost in terms of slower convergence rate. Nevertheless, as Barndorff-Nielsen and Shephard (2004b,c) point out, bipower variation is a less efficient estimator than realized volatility.

As for the microstructure noise estimator suggested by Zhang, Mykland, and Aït-Sahalia (2004), we have the following Proposition.

**Proposition 5.** *Let* $\mathrm{d}z_t = 0$ *a.s., where* $\mathrm{d}z_t$ *is defined in (1). Let* $M/l^{1/3} \to \pi$, $0 < \pi < \infty$.. *Then Assumption A1 holds with* $RM_{t,M} = \widehat{RV}_{t,l,M}$ *for* $b_M = M^{1/3}$.

Finally, if our realized measure is the microstructure robust subsampled based realized volatility measure of Zhang (2004), Aït-Sahalia, Mykland and Zhang (2005b), and Barndorff-Nielsen, Hansen, Lunde and Shephard (2004), we have the following result.

**Proposition 6.** *Let* $\mathrm{d}z_t = 0$ *a.s., where* $\mathrm{d}z_t$ *is defined in (1). Let* $M/\sqrt{l} \to \pi$, $0 < \pi < \infty$. *Then Assumption A1 holds with* $RM_{t,M} = \widetilde{RV}_{t,e,M}$ *for* $b_M = M^{1/2}$.

As noted below in Theorem 2, we have no gain in using (model based) simulated integrated volatility whenever $b_M$ grows at a faster rate than $T$. This implies that the number of intradaily observations should grow at a rate faster than $T$ (resp. $T^3, T^2$) for the cases of realized volatility and bipower variation (resp. microstructure robust realized volatilities). Also, unless we let the bandwidth go to zero very slowly, even the necessary condition ensuring that contribution of measurement error approaches zero is quite stringent, and gets more and more stringent moving from realized volatility and bipower variation to modified subsampled realized volatility. Clearly, we also want to choose a rather large value for $T$, as the estimator for the conditional density is consistent (at a nonparametric rate) only for $T$ going to infinity. Thus, we need a very large number of intradaily observations, especially for the case of $\widehat{RV}_{t,l,M}$. However, the latter measure is robust to microstructure noise, at least to the type of noise outlined in Section 2, and therefore we can use all the available information and sample data at a very high frequency (such as 1-5 seconds).

# 6 Experimental Results: Predictive Density Estimator Accuracy

Our objective in this section is to assess the accuracy of the predictive density estimator outlined in Section 3 of this paper. This is the estimator constructed using realized volatility measures. Namely, recall that we define

$$\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) = \frac{\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{RM_{t+1,M}-x}{\xi_{2,T}}, \frac{RM_{t,M}-RM_{T,M}}{\xi_{2,T}}\right)}{\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right)}.$$

Analysis of the simulation based estimator outlined in Section 4 is left to future research, although it should be noted that the kernel density estimators used in the construction of the simulation based estimator have been found to be very sensitive to parameter estimation error. On the other hand, no parameters need to be estimated when realized measures are used, as in $\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})$.

Our experiment is carried out as follows. Using the notation defined in Section 4, we begin by simulating $S$ paths of $h_{i,j\frac{k+1}{N}}$, as defined in (13), where each path is of length $k+1$, and where data are simulated using the discrete interval $N^{-1}$. In order to carry out the simulations, we define the drift term, variance term, and derivative of the variance term of the instantaneous volatility process as follows:

$$\mu(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \kappa\left(\alpha + 1 - h_{i,(j-1)\frac{k+1}{N}}\right),$$
$$\sigma(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \sqrt{2\kappa} h_{i,(j-1)\frac{k+1}{N}},$$
$$\sigma'(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \frac{1}{2}\sqrt{2\kappa} h^{-1/2}_{i,(j-1)\frac{k+1}{N}}.$$

Thus, we are assuming that volatility follows a square root process. The reader is referred to Meddahi (2001) for a complete discussion of the one to one mapping between square root stochastic volatility models and eigenfunction stochastic volatility models of the variety posited here. Now, note that we can define the volatility as $\sigma^2_{i,j\frac{k+1}{N}} = \psi(h_{i,j\frac{k+1}{N}})$, as discussed above. Also, we set $k=1$ and $\boldsymbol{\theta} = \boldsymbol{\theta}^\dagger$. Now, we fix $\left\{W_{i,j\frac{k+1}{N}} - W_{i,(j-1)\frac{k+1}{N}}\right\}$ across simulations, $j = 1, \ldots, N/(k+1)$. However, $\left(W_{i,j\frac{k+1}{N}} - W_{i,(j-1)\frac{k+1}{N}}\right)$ is independent across $i$, for $j = N/(k+1)+1, \ldots, N$. Thus, $h_{i,j\frac{k+1}{N}}$ is fixed across simulations; and consequently does not depend on $i$, for $j = 1, \ldots, N/(k+1)$.

Given this framework, and noting that integrated volatility is thus fixed during the first day, across all simulations, we can define the first day $IV$ as:

$$IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right) = \frac{1}{N/(k+1)} \sum_{j=1}^{N/(k+1)} \sigma^2_{j\frac{k+1}{N}}(\boldsymbol{\theta}^\dagger), \tag{25}$$

where the index $i$ is dropped as this value is fixed across $i$. Furthermore, for all days beyond the first:

$$IV_{i,\tau+1,N}\left(\boldsymbol{\theta}^\dagger\right) = \frac{1}{N/(k+1)} \sum_{j=\tau N/(k+1)+1}^{(\tau+1)N/(k+1)} \sigma^2_{i,j\frac{k+1}{N}}(\boldsymbol{\theta}^\dagger),\ \tau = 1, ..., k, \quad i = 1, ..., S.$$

It thus follows that we can construct a pseudo-true conditional predictive volatility density as:

$$\widetilde{f}_{IV_{3,N}(\boldsymbol{\theta}^\dagger)|IV_{2,N}(\boldsymbol{\theta}^\dagger)}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right)\right) = \frac{\frac{1}{S\xi_S^2}\sum_{i=1}^S \mathbf{K}\left(\frac{IV_{i,3,N}(\boldsymbol{\theta}^\dagger)-x}{\xi_S}, \frac{IV_{i,2,N}(\boldsymbol{\theta}^\dagger)-IV_{1,N}(\boldsymbol{\theta}^\dagger)}{\xi_S}\right)}{\frac{1}{S\xi_S}\sum_{i=1}^S \mathbf{K}\left(\frac{IV_{i,2,N}(\boldsymbol{\theta}^\dagger)-IV_{1,N}(\boldsymbol{\theta}^\dagger)}{\xi_S}\right)}.$$

Note that, for $\xi_S = O(S^{-1/8})$,

$$\begin{aligned} & \widetilde{f}_{IV_{3,N}(\boldsymbol{\theta}^\dagger)|IV_{2,N}(\boldsymbol{\theta}^\dagger)}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right)\right) - f_{IV_{3,N}(\boldsymbol{\theta}^\dagger)|IV_{2,N}(\boldsymbol{\theta}^\dagger)}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right)\right) \\ = \ & O_P\left(\frac{1}{S^{1/4}}\right) + O_P\left(\frac{1}{N^{1/2-\delta/2}}\right), \text{ for any } \delta > 0, \end{aligned} \tag{26}$$

where the error on the right hand side above can be made arbitrarily small by choosing $S$ and $N$ sufficiently large. Note that the first term on the right hand side of (26) comes straightforwardly from Theorem 1 in Andrews (1995), and the second term is due to the discretization error (see e.g. Pardoux and Talay, 1985, Corollary 1.8).

Given the above pseudo true estimator, it remains to construct a second estimator in the usual way, as discussed in Section 3. This can be done by simulating a path of length $T$ for $X_t$, say, using constant drift and the same specification for instantaneous volatility as that given above. Namely, and in addition to the above volatility model, we specify

$$\mathrm{d}X_t = m\mathrm{d}t + \sqrt{\frac{\eta^2}{2\kappa}h_t}\mathrm{d}W_{1,t},$$

where $h_t$ follows the same square root process defined above. Given this framework, we generate paths for $X_t$ via use of the Milstein scheme, using a discrete interval of order $N^{-1}$. Then, we sample the simulated process for the $X_t$ at frequency $1/M$ of the actual data, and form the 4 realized volatility measures discussed in Section 2 based on $M$ intradaily observations. Finally, along the same lines as in Section 3, we construct

$$\widehat{f}_{RM_{T+1,M}|RM_{T,M}}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right) = \frac{\frac{1}{T\xi_T^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{RM_{t+1,M}-x}{\xi_T}, \frac{RM_{t,M}-IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)}{\xi_T}\right)}{\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1}K\left(\frac{RM_{t,M}-IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)}{\xi_T}\right)}, \tag{27}$$

where $IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)$ is the quantity computed in (25); and where the kernel that we use in our experiment is the Epanechnikov kernel defined as

$$\begin{aligned}\mathbf{K}(\cdot,\cdot) &= \frac{3}{4}\left(1-\left(\frac{RM_{t+1,M}-x}{\xi_T}\right)^2\right)1_{\left\{\left|\frac{RM_{t+1,M}-x}{\xi_T}\right|\leq 1\right\}} \\ &\quad\times\frac{3}{4}\left(1-\left(\frac{RM_{t,M}-IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)}{\xi_T}\right)^2\right)1_{\left\{\left|\frac{RM_{t,M}-IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)}{\xi_T}\right|\leq 1\right\}},\end{aligned}$$

for the bivariate case, and analogously for the univariate case.

We repeat the step above $\Delta$ times, i.e. we generate $\Delta$ paths of length $T$ for $X_t$. Hereafter, let $\widehat{f}_{RM_{T+1,M}|RM_{T,M},p}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right)$ denotes the conditional density estimator at replication $p$, with $p = 1, ..., \Delta$.

In order to measure the degree of accuracy of the conditional density estimator based on realized measures, for different values of $M$ (i.e. $M = \{48, 144, 360, 720\}$) and $T$ (i.e. $T = \{100, 300, 500\}$), at each replication, we construct the integrated mean square error, and then we average over the number of replications. That is, we construct,

$$\frac{1}{\Delta}\sum_{p=1}^{\Delta}\left(\frac{1}{\Lambda}\sum_{j=1}^{\Lambda}\left(\widehat{f}_{RM_{T+1,M}|RM_{T,M},p}\left(x_j|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right) - \widetilde{f}_{IV_{3,N}\left(\boldsymbol{\theta}^{\dagger}\right)|IV_{2,N}\left(\boldsymbol{\theta}^{\dagger}\right)}\left(x_j|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right)\right)^2\right), \tag{28}$$

where $\Lambda$ is the number of points in the range of the data across which to evaluate the density. Here $\Lambda$ has been set equal to 100 equally spaced values across a large simulated sample of $IV$ values, for the interval $[\overline{IV} - \widehat{\mathrm{SE}}(IV), \overline{IV} + \widehat{\mathrm{SE}}(IV)]$. Also, in order to evaluate the variability across replications we also consider the standard error of the integrated mean square error across Monte Carlo replications, that is:

$$\left\{ \frac{1}{\Delta} \sum_{p=1}^{\Delta} \right.$$

$$\left( \frac{1}{\Lambda} \sum_{j=1}^{\Lambda} \left( \widehat{f}_{RM_{T+1,M}|RM_{T,M},p} \left( x_j | IV_{1,N} \left( \boldsymbol{\theta}^{\dagger} \right) \right) - \widetilde{f}_{IV_{3,N}(\boldsymbol{\theta}^{\dagger})|IV_{2,N}(\boldsymbol{\theta}^{\dagger})} \left( x_j | IV_{1,N} \left( \boldsymbol{\theta}^{\dagger} \right) \right) \right)^2 \right)^2$$

$$- \left[ \frac{1}{\Delta} \sum_{p=1}^{\Delta} \right.$$

$$\left. \left. \left( \frac{1}{\Lambda} \sum_{j=1}^{\Lambda} \left( \widehat{f}_{RM_{T+1,M}|RM_{T,M},p} \left( x_j | IV_{1,N} \left( \boldsymbol{\theta}^{\dagger} \right) \right) - \widetilde{f}_{IV_{3,N}(\boldsymbol{\theta}^{\dagger})|IV_{2,N}(\boldsymbol{\theta}^{\dagger})} \left( x_j | IV_{1,N} \left( \boldsymbol{\theta}^{\dagger} \right) \right) \right)^2 \right) \right]^2 \right\}^{1/2}.$$

Additionally, we report analogous integrated mean absolute deviation, mean absolute percentage error, and associated standard errors. All results are based upon 1000 Monte Carlo iterations. In all tables, the first column reports results for $RV_{t,M}$, the second for $BV_{t,M}$, the third for $\widehat{RV}_{t,l,M}$ and the fourth for $\widetilde{RV}_{t,e,M}$. In parenthesis we report the associated standard errors.

The findings of this experiment are gathered in Table 1 (Case I: no microstructure noise or jumps); Table 2 (Case II: microstructure noise); and Table 3 (Case III: jumps). The magnitude and frequency of the noise and jumps are as follows: for microstructure noise, we add i.i.d. $N(0, 1/\alpha 1440)$ increments to the generated data, where $\alpha = \{0.5, 1, 2\}$; and for jumps, we add i.i.d. $N(0, \alpha \times 0.64 \times \overline{IV})$ jumps occurring once (on average) every five days, where $\alpha = \{1, 3, 5\}$. A number of clear conclusions emerge upon examination of the tables. First, for Case I, $RV_{t,M}$ and $BV_{t,M}$ perform approximately equally well. Additionally, both of these measures yield more accurate density estimates than the two subsampled measures, as expected. Furthermore, the mean square error reduces approximately 20-40% as the number of intraday observations increases from 48 to 720; and there is a reduction of approximately 30-50% as the number of daily observations is increased from 100 to 300. Finally, the above results hold for all three accuracy measures (i.e mean square error - MSE, mean absolute deviation, and mean absolute percentage error). Second, for Case II, $\widehat{RV}_{t,l,M}$ 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.[5] For example, consider Panel A in Table 2. The

[5]Note that no values are reported for $BV_{t,M}$ for the case where $M = 720$. This omission is due to numerical instability of the particular kernel density estimator that we use in this experiment, and is indicative of the fact that the non-robust measures exhibit increasingly dramatic performance deterioration as $M$ increases.

MSE for $RV_{t,M}$ is 0.023634 for $M = 48$; and increases to 0.481822 when $M = 720$, again as expected. Furthermore, $\widetilde{RV}_{t,e,M}$ is 0.029185 for $M = 48$; and slowly shrinks as $M$ increases, to a level of 0.021236 when $M = 720$. The $\widehat{RV}_{t,l,M}$ measure performs similarly well for most values of $M$, other than when $M = 48$, in which case it performs substantially worse than all other measures. Overall, $\widetilde{RV}_{t,e,M}$ performs marginally better than $\widehat{RV}_{t,l,M}$, although the performance of the former measure depends somewhat upon the number of realized volatilities ($e$) used in the construction. Thus, the results here that are based on their comparison are only preliminary, as we did not experiment with different weights and truncation values. Third, examination of Table 3 suggests that when there are jumps of sufficient magnitude (see Panel C), $BV_{t,M}$ outperforms all other measures, with MSE values less than one half that or any other realized measure. Furthermore, in this case, the standard error associated with the MSE, across simulations, is also less than one half that of any other measure. The same result holds for the case when $\alpha = 3$ (see Panel B), although the magnitude of the jumps is sufficiently small in Panel A of the table so as to leave little to choose between the 4 realized measures. In summary, the predictive integrated volatility density estimator appears to be performing adequately, and as expected, given the theoretical results outlined in the prior sections of this paper.

# 7 Empirical Illustration: forecasting the conditional distribution of daily volatility of Intel

## 7.1 Data description

Data are retrieved from the Trade and Quotation (TAQ) database at the New York Stock Exchange (NYSE). The TAQ database contains intraday trades and quotes for all securities listed on the NYSE, the American Stock Exchange (AMEX) and the Nasdaq National Market System (NMS). The data is published monthly on CD-ROM since 1993 and on DVD since June 2002. Our sample contains data for the stock Intel and extends from January 1, 1997 until May the 20th, for a total of 100 trading days.

Therefore, we predict the conditional distribution of daily integrated volatility for Wednesday May 21st, 1997.

From the original data set, which includes prices recorded for every trade, we extracted 1 second and 5 minutes interval data. Provided that there is sufficient liquidity in the market, the 5 minutes 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, the choice of the two frequencies, in order to evaluate the effect of microstructure noise on the estimated densities.

The price figures for each 1 second and 5 minutes intervals are determined using the last tick method, which was first proposed by Wasserfallen and Zimmermann (1985). Specifically, when no trade occurs at the required point in time, the price is calculated as last observed price. Another way of obtaining equidistant artificial prices is the linear interpolation method, where the price figure is computed as the interpolated average between the preceding and the immediately following trades, weighted linearly by their inverse relative distance to the required point in time (see Andersen and Bollerslev, 1997). Unfortunately, constructing realized measures of integrated volatility using this kind of artificial price process has some disadvantages. See, e.g., Lemma 1 in Hansen and Lunde (2004); realized volatility constructed from linearly interpolated returns converges in probability to zero as the time interval shrinks to zero.

From the calculated series we we have obtained 1 second and 5 minutes intradaily returns as the difference between successive log prices expressed in percentages. Formally

$$R_{t+\frac{i}{M}} = 100 \times \left( \log(X_{t+\frac{i}{M}}) - \log(X_{t+\frac{i-1}{M}}) \right),$$

where $R_{t+i/M}$ denotes the return for intraday period $i/M$ on trading day $t$ , with $t = 0, \ldots, T-1$. The New York Stock Exchange opens at 9:30 a.m. and closes at 4.00 p.m.. Therefore a full trading day consists of 2341 (resp. 79) intraday returns calculated over an interval of one second (resp. five minutes).[6]

### 7.2 Main Results

Using the two series of returns at different frequencies, the predictive densities have been calculated for each of the four considered realized measures. Results are shown in Figures 1 to 8 using 1000 evaluation points for the conditional density.

The graphs reveal some interesting facts. First of all, the graphs for the realized volatility and bipower variation are quite similar. This means that jumps occur occasionally in the price process, and therefore do not affect a procedure which is based on sample sizes containing a large number of daily observations.

On the contrary, microstructure noise seems to have a more tangible effect on our estimates. In fact, by looking at the range of the densities of realized volatility and bipower variation for the two different frequencies, one can immediately appreciate the effect of microstructure noise. As predicted by the theory (see Aït-Sahalia, Mykland and Zhang, 2005a), and confirmed by simulation results, when the time interval between successive observations becomes small, than the signal to noise ratio contained in the data also becomes very small, and realized volatility and bipower variation tend to explode, instead of converging to the quadratic variation. This can be seen in

[6]Highly liquid stocks may have more than one price at certain points in time (for example, 5 or 10 quotations at the same time stamp is normal for Intel and Microsoft); hence, when there exists more than one price at the required time, we select the last provided one.

the pictures; in fact, the range of the density of the two estimators, estimated with data at a high frequency, is considerably wider than the corresponding one obtained with a lower frequency.

On the contrary, the microstructure robust realized measures display a more stable picture. Increasing the frequency at which data are sampled does not seem to induce any distortion; the results shown in this example provide another warning remark against trying to estimate integrated volatility with data sampled at a very high frequency, when using estimators which are not robust to the presence of market microstructure noise.

# 8 Concluding Remarks

In this paper we have proposed a feasible, model free estimator of the conditional predictive density of integrated volatility. The estimator, which is constructed using either realized volatility, or using simulated integrated volatility, is shown to be both empirically tractable and uniformly consistent under both microstructure noise, and in the presence of jumps. In this sense, the estimator discussed in this paper can be viewed as a natural model free extension of the point predictive estimator developed in Andersen, Bollerslev, Diebold and Labys (2003) and Andersen, Bollerslev and Meddahi (2004, 2005). A simulation experiment is carried out, illustrating that the estimator is accurate, and yields sensible answers in a variety of scenarios, including cases where there is microstructure noise and when there are jumps. Finally, we illustrate the ease with which the estimator can be applied via an empirical illustration.

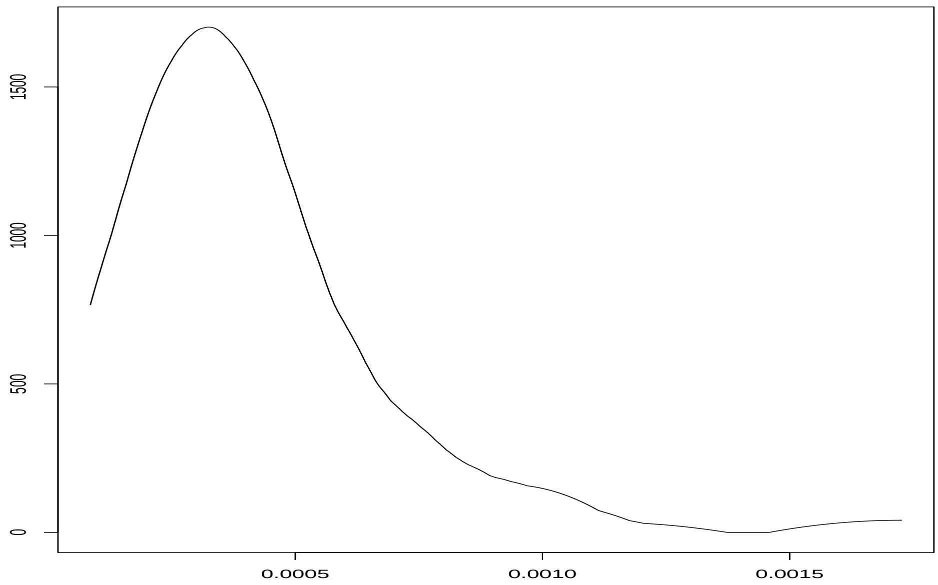


Figure 1: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on realized volatility. $M = 79$, $T = 100$.

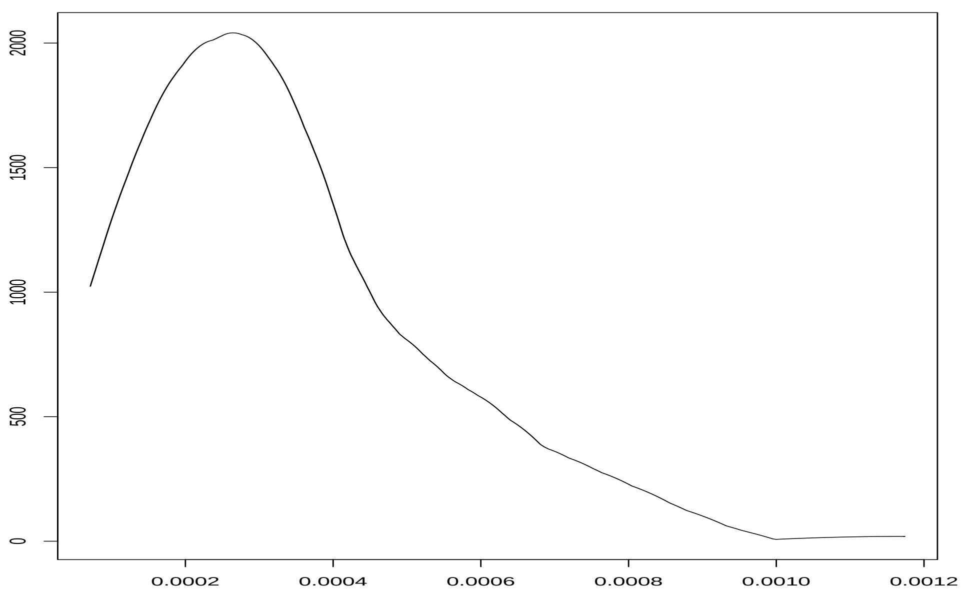


Figure 2: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on bipower variation. $M = 79$, $T = 100$.

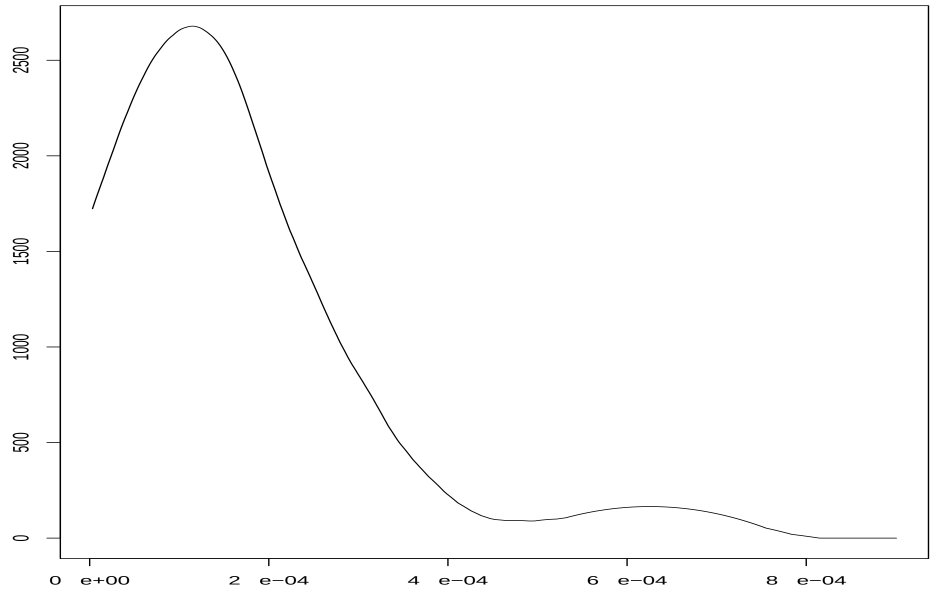


Figure 3: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on $\widehat{RV}_{t,l,M}$. $M = 79$, $T = 100$.

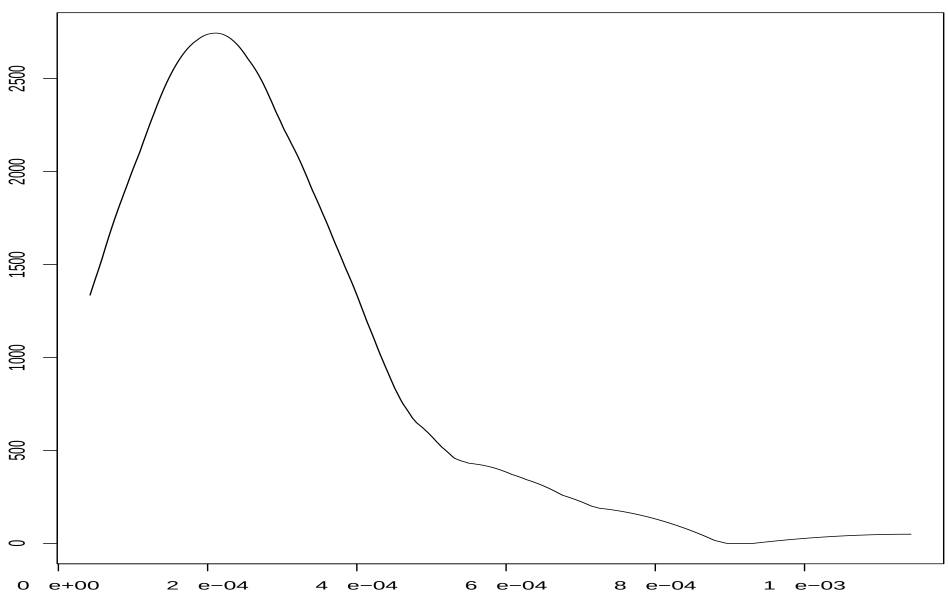


Figure 4: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on $\widetilde{RV}_{t,e,M}$. $M = 79$, $T = 100$.

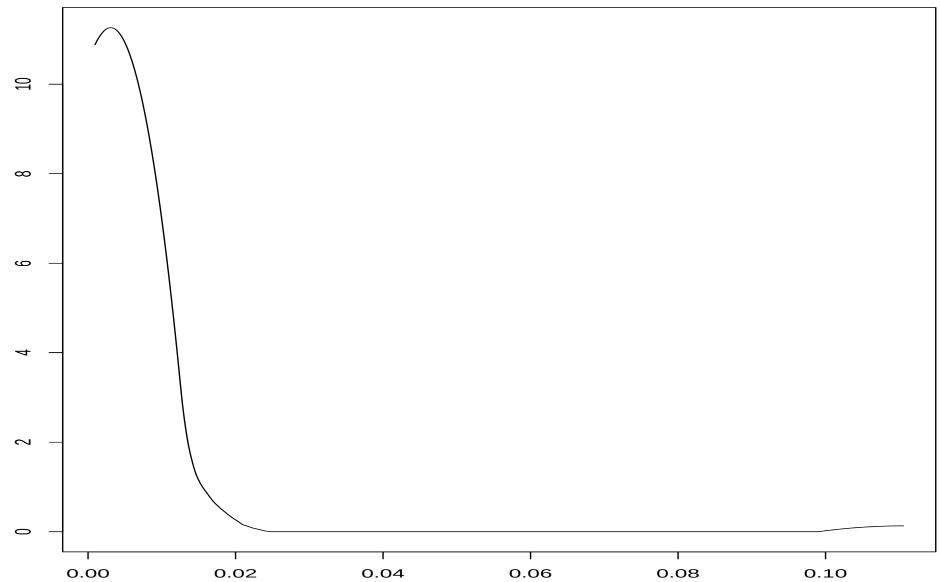


Figure 5: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on realized volatility. $M = 2341$, $T = 100$.

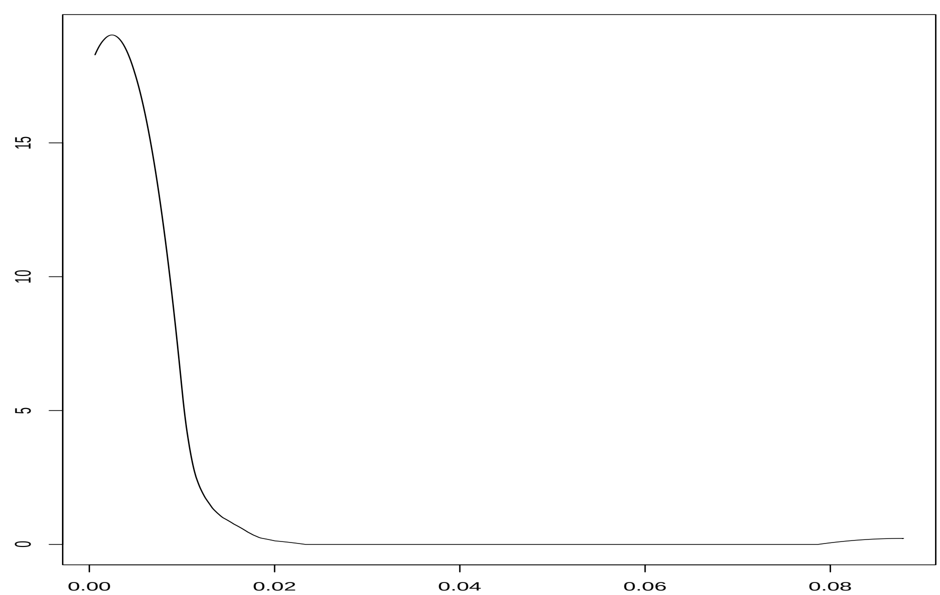


Figure 6: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on bipower variation. $M = 2341$, $T = 100$.

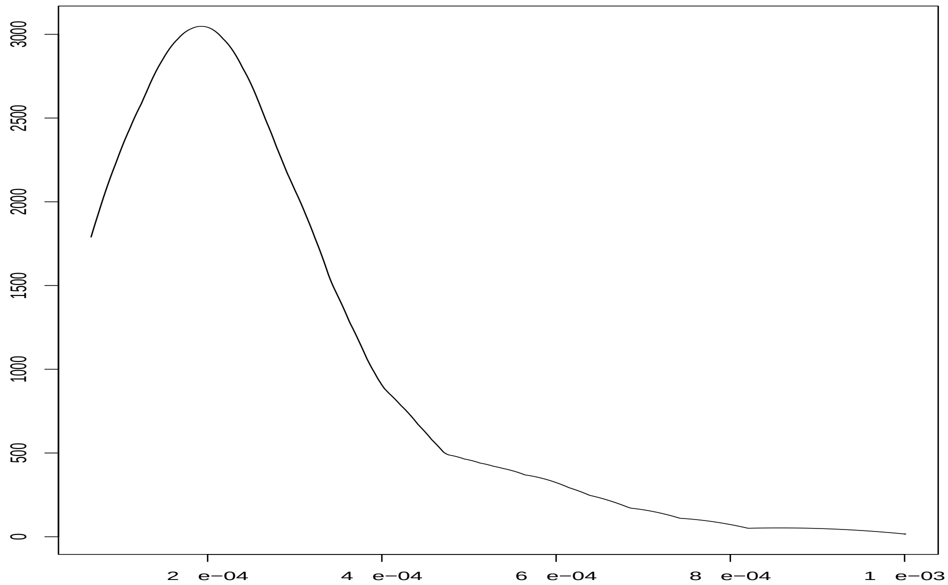


Figure 7: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on $\widehat{RV}_{t,l,M}$. $M = 2341$, $T = 100$.

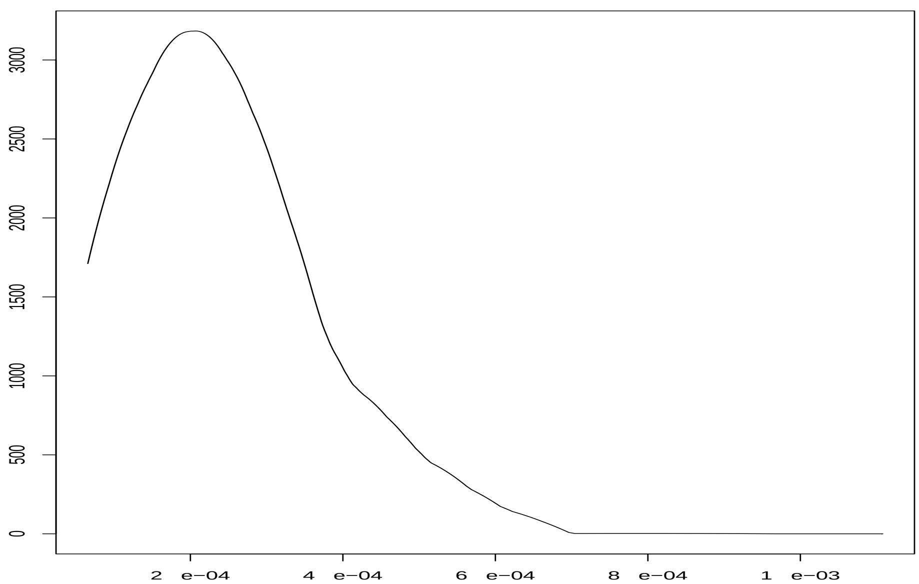


Figure 8: Predictive conditional density of $IV_{T+1}$ for the Intel stock, based on $\widetilde{RV}_{t,e,M}$. $M = 2341$, $T = 100$.

## Appendix

In the sequel, let $u_M = (x, RM_{T,M})'$, $\widehat{U}_{t+1,M} = (RM_{t+1,M}, RM_{t,M})'$, $U_{t+1} = (IV_{t+1}, IV_t)$, and $\widehat{U}_{t+1,M} - U_{t+1} = (N_{t+1,M}, N_{t,M})'$. Thus

$$\begin{aligned}
\widehat{f}_{RM_{T+1}|RM_{T,M}}(x|RM_{T,M}) &= \frac{\widehat{f}_{RM_{T+1,M},RM_{T,M}}(x, RM_{T,M})}{\widehat{f}_{RM_{T,M}}(RM_{T,M})} \\
&= \frac{\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{RM_{t+1,M}-x}{\xi_{2,T}}, \frac{RM_{t,M}-RM_{T,M}}{\xi_{2,T}}\right)}{\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right)} \\
&= \frac{\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{\widehat{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)}{\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1}K\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right)}.
\end{aligned}$$

Also, in order to stress the fact the kernel has bounded support, we shall write $\mathbf{K}(u) = \widetilde{\mathbf{K}}(u)\mathbf{1}_{\{|u|\leq 1\}}$, so that

$$\widehat{f}_{RM_{T+1}|RM_{T,M}}(x|RM_{T,M}) = \frac{\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\widetilde{\mathbf{K}}\left(\frac{\widehat{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\{|\widehat{U}_{t+1,M}-u_M|\leq\xi_{2,T}\}}}{\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1}\widetilde{K}\left(\frac{RM_{t,M}-RM_{T,M}}{\xi_{1,T}}\right)1_{\{|RM_{t,M}-RM_{T,M}|\leq\xi_{1,T}\}}},$$

where $\mathbf{1}_{\{|\widehat{U}_{t+1,M}-u_M|\leq\xi_{2,T}\}} = 1_{\{|RM_{t+1,M}-x|\leq\xi_{2,T}\}}1_{\{|RM_{t,M}-RM_{T,M}|\leq\xi_{2,T}\}}$. The proof of Theorem 1 requires the following Lemma.

**Lemma 1.** *Let assumptions A1 and A4 hold. Then*

$$\widehat{f}_{RM_{T+1,M},RM_{T,M}}(x, RM_{T,M}) - f_{IV_{t+1|IV_t}}(x, RM_{T,M}) = O_P\left(b_M^{-1/2}\xi_{2,T}^{-3}\right) + O_P\left(T^{-1/2}\xi_{2,T}^{-2}\right) + O\left(\xi_{2,T}^2\right), \tag{29}$$

*uniformly in $x$, and*

$$\widehat{f}_{RM_{t,M}}(RM_{T,M}) - f_{IV_t}(RM_{T,M}) = O_P\left(b_M^{-1/2}\xi_{1,T}^{-2}\right) + O_P\left(T^{-1/2}\xi_{1,T}^{-1}\right) + O\left(\xi_{1,T}^2\right) \tag{30}$$

### Proof of Lemma 1

We show the result in (29). The result in (30) follows by the same argument. By simple algebraic manipulation

$$\begin{aligned}
&\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\widetilde{\mathbf{K}}\left(\frac{\widehat{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\{|\widehat{U}_{t+1,M}-u_M|\leq\xi_{2,T}\}} - f_{U_{t+1}}(u_M) \\
= &\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\{|\widehat{U}_{t+1,M}-u_M|\leq\xi_{2,T}\}}\right. \\
&\left.-\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\{|\widehat{U}_{t+1,M}-u_M|\leq\xi_{2,T}\}}\right)\right)
\end{aligned} \tag{31}$$

$$+\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\left(\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\left\{\left|\widehat{U}_{t+1,M}-u_M\right|\leq\xi_{2,T}\right\}}\right)\right.$$

$$\left.-\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{t+1}-u_M\right|\leq\xi_{2,T}\right\}}\right)\right) \tag{32}$$

$$+\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{t+1}-u_M\right|\leq\xi_{2,T}\right\}}\right)-f_{U_{t+1}}(u_M) \tag{33}$$

$$+\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\left(\widetilde{\mathbf{K}}\left(\frac{\widehat{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)-\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\right)\mathbf{1}_{\left\{\left|\widehat{U}_{t+1,M}-u_M\right|\leq\xi_{2,T}\right\}}. \tag{34}$$

We begin by considering the term in (34), which via a mean value expansion can be rewritten as

$$\begin{aligned}
&\left|\frac{1}{T\xi_{2,T}^3}\sum_{t=1}^{T-1}\widetilde{\mathbf{K}}_1'\left(\frac{\widetilde{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\left\{\left|\widehat{U}_{t+1,M}-u_M\right|\leq\xi_{2,T}\right\}}N_{t+1,M}\right|\\
&+\left|\frac{1}{T\xi_{2,T}^3}\sum_{t=1}^{T-1}\widetilde{\mathbf{K}}_2'\left(\frac{\widetilde{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\left\{\left|\widehat{U}_{t+1,M}-u_M\right|\leq\xi_{2,T}\right\}}N_{t,M}\right|\\
\leq\ &\sup_{t\leq T-1}\left|\frac{1}{\xi_{2,T}^3}\widetilde{\mathbf{K}}_1'\left(\frac{\widetilde{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\left\{\left|\widehat{U}_{t+1,M}-u_M\right|\leq\xi_{2,T}\right\}}\right|\frac{1}{T}\sum_{t=1}^{T-1}\left|N_{t+1,M}\right|\\
&+\sup_{t\leq T-1}\left|\frac{1}{\xi_{2,T}^3}\widetilde{\mathbf{K}}_2'\left(\frac{\widetilde{U}_{t+1,M}-u_M}{\xi_{2,T}}\right)\mathbf{1}_{\left\{\left|\widehat{U}_{t+1,M}-u_M\right|\leq\xi_{2,T}\right\}}\right|\frac{1}{T}\sum_{t=1}^{T-1}\left|N_{t,M}\right|\\
=\ &O_{a.s.}\left(\xi_{2,T}^{-3}\right)O_P\left(b_M^{-1/2}\right)=O_P\left(\xi_{2,T}^{-3}b_M^{-1/2}\right),
\end{aligned}$$

where $\widetilde{U}_{t+1,M}\in\left(\widehat{U}_{t+1,M},U_{t+1}\right)$ and $\widetilde{\mathbf{K}}_i'\left(\frac{\widetilde{U}_{t+1,M}-u_M}{\xi}\right)$, $i=1,2$, denotes the derivative of the function with respect to the $i-$th argument, and by noting that $\frac{1}{T}\sum_{t=1}^{T-1}|N_{t,M}|=O_P(b_M^{-1/2})$, because of A1(ii).

We now consider the term in (32). Let $b_M=T^{\psi}$. First, note that

$$\begin{aligned}
\Pr\left(b_M^{2/7}\sup_{t\leq T-1}\left|N_{t,M}\right|>\varepsilon\right)&\leq\sum_{t=1}^{T-1}\Pr\left(b_M^{2/7}\left|N_{t,M}\right|>\varepsilon\right)\leq\frac{1}{\varepsilon^{16}}Tb_M^{\frac{2}{7}16}\mathrm{E}\left(\left|N_{t,M}\right|^{16}\right)\\
&\leq\frac{1}{\varepsilon^{16}}Tb_M^{\frac{2}{7}16}b_M^{-8}=\frac{1}{\varepsilon^{16}}b_M^{1/\psi}b_M^{\frac{32}{7}}b_M^{-8}.
\end{aligned} \tag{35}$$

Now, as $\psi>1/2$, then $b_M^{1/\psi}b_M^{\frac{32}{7}}b_M^{-8}<b_M^{-6}b_M^{32/7}=b_M^{-10/7}$. Thus, the right hand side above goes to zero. Let

$$\Omega_M^+=\left\{\omega:b_M^{2/7}\sup_{t\leq T-1}\left|N_{t,M}\right|<\varepsilon\right\} \tag{36}$$

and note that, given (35) and (36), $\lim_M b_M^{5/4}\Pr(\Omega_M^+)=1$. Thus, for any $\omega\in\Omega_M^+$, $|N_{t,M}|\leq cb_M^{-2/7}$, for some $0<c<\infty$, independent of $T$ and $M$. Now, notice that

$$\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\frac{1}{\xi_{2,T}^2}\mathbf{1}_{\left\{\left|U_{t+1}-u_M\right|\leq\xi_{2,T}\right\}}\right)=\int_{-1}^{1}\int_{-1}^{1}\mathbf{K}(z)f_{U_{t+1}}\left(u_M+z\xi_{2,T}\right)\mathrm{d}z.$$

Now, when $1_{\{|RM_{t,M}-RM_{T,M}|\leq\xi_{2,T}\}}$,

$$\left|\frac{IV_t-RM_{T,M}}{\xi_{2,T}}\right| = \left|\frac{(IV_t-RM_t)+(RM_t-RM_{T,M})}{\xi_{2,T}}\right| \leq 1+\xi_{2,T}^{-1}N_{t,M},$$

and, analogously, when $1_{\{|RM_{t+1,M}-x|\leq\xi_{2,T}\}}$,

$$\left|\frac{IV_{t+1}-x}{\xi_{2,T}}\right| = \left|\frac{(IV_{t+1}-RM_{t+1})+(RM_{t+1}-x)}{\xi_{2,T}}\right| \leq 1+\xi_{2,T}^{-1}N_{t+1,M}.$$

Thus, for any $\omega\in\Omega_M^+$, with $\lim_{M\to\infty} b_M^{5/4}\Pr(\Omega_M^+)\to 1$,

$$\left|\frac{(IV_t-RM_{t,M})+(RM_{t,M}-RM_{T,M})}{\xi_{2,T}}\right| \leq 1+c\xi_{2,T}^{-1}b_M^{-2/7},$$

and

$$\left|\frac{(IV_{t+1}-RM_{t+1,M})+(RM_{t+1,M}-x)}{\xi_{2,T}}\right| \leq 1+c\xi_{2,T}^{-1}b_M^{-2/7},$$

with $c$ independent of $M$ and $T$. Therefore

$$\begin{aligned}
&\left|\frac{1}{T}\sum_{t=1}^{T-1}\left(\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\frac{1}{\xi_{2,T}^2}\mathbf{1}_{\{|\widehat{U}_{t+1,M}-u_M|\leq\xi_{2,T}\}}\right)\right.\right.\\
&\left.\left.-\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{t+1}-u_M}{\xi_{2,T}}\right)\frac{1}{\xi_{2,T}^2}\mathbf{1}_{\{|U_{t+1}-u_M|\leq\xi_{2,T}\}}\right)\right)\right|\\
\leq\ & \int_{-1-c\xi_{2,T}^{-1}b_M^{-2/7}}^{-1}\int_{-1-c\xi_{2,T}^{-1}b_M^{-2/7}}^{-1}\mathbf{K}(z)f_{U_{t+1}}(u_M+z\xi_{2,T})\mathrm{d}z\\
&+\int_1^{1+c\xi_{2,T}^{-1}b_M^{-2/7}}\int_1^{1+c\xi_{2,T}^{-1}b_M^{-2/7}}\mathbf{K}(z)f_{U_{t+1}}(u_M+z\xi_{2,T})\mathrm{d}z\\
=\ & O(b_M^{-2/7}\xi_{2,T}^{-1}).
\end{aligned}$$

Finally, the term in (33) is of order $O(\xi_{2,T}^2)$ and the term in (31) is of order $O_P\left(T^{-1/2}\xi_{2,T}^{-2}\right)$, by Theorem 1 in Andrews (1995), setting, in his notation, $k=2$, $\lambda=0$, $\eta=\infty$, $\sigma_{1T}=\sigma_{2T}$, and $\omega=2$. In fact, given A2-A3, $IV_t$ has an ARMA structure, and so is geometrically strong mixing, thus NP1 in Andrews holds with $\eta=\infty$, and $a(s)$ decaying at a geometric rate. Also, A4 implies that NP2 and NP4 in Andrews (1995) are satisfied. ■

## Proof of Theorem 1

A simple rearrangement of terms yields

$$\begin{aligned}
\widehat{f}_{RM_{T+1}|RM_{T,M}}(x|RM_{T,M}) &= \frac{\widehat{f}_{RM_{T+1,M},RM_{T,M}}(x,RM_{T,M})}{\widehat{f}_{RM_{t,M}}(RM_{T,M})}-\frac{f_{IV_{t+1},IV_t}(x,RM_{T,M})}{f_{IV_t}(RM_{T,M})}\\
&= \frac{\widehat{f}_{RM_{T+1,M},RM_{T,M}}(x,RM_{T,M})-f_{IV_{t+1},IV_t}(x,RM_{T,M})}{\widehat{f}_{RM_{t,M}}(RM_{T,M})}
\end{aligned}$$

$$+\frac{f_{IV_{t+1},IV_t}(x,RM_{T,M})\left(f_{IV_t}(RM_{T,M})-\widehat{f}_{RM_{t,M}}(RM_{T,M})\right)}{\widehat{f}_{RM_{t,M}}(RM_{T,M})f_{IV_t}(RM_{T,M})}.$$

Now, $\widehat{f}_{RM_{t,M}}(RM_{T,M})>d_T$, and by noting that $f_{IV_t}(RM_{T,M})=\widehat{f}_{RM_{t,M}}(RM_{T,M})+o_P(1)$, given Lemma 1 above, the statement in the theorem follows. ■

## Proof of Proposition 1

(a) We begin by showing that given (i) and (ii),

$$\frac{\xi_{1,T}^2 d_T^{-2}}{\xi_{2,T}^2 d_T^{-1}}\to 0,\quad \frac{\xi_{1,T}^{-1} d_T^{-2}}{\xi_{2,T}^{-2} d_T^{-1}}\to 0,\quad \frac{\xi_{1,T}^{-2} d_T^{-2}}{\xi_{2,T}^{-3} d_T^{-1}}\to 0. \tag{37}$$

Now, if $\delta<2\left(\phi_1-\phi_2\right)$, and $\phi_1>\phi_2$,

$$\frac{\xi_{1,T}^2 d_T^{-2}}{\xi_{2,T}^2 d_T^{-1}}=\frac{\xi_{1,T}^2}{\xi_{2,T}^2 d_T}\to 0;$$

also, if $\delta-3\phi_2+2\phi_1<0$ and $\phi_2<\phi_1<\frac{3}{2}\phi_2$, then

$$\frac{\xi_{1,T}^{-2} d_T^{-2}}{\xi_{2,T}^{-3} d_T^{-1}}=\frac{\xi_{2,T}^3}{\xi_{1,T}^2 d_T}\to 0.$$

Finally, note that given (ii), $\frac{\xi_{2,T}^3}{\xi_{1,T}^2 d_T}\to 0$ implies that

$$\frac{\xi_{1,T}^{-1} d_T^{-2}}{\xi_{2,T}^{-2} d_T^{-1}}=\frac{\xi_{2,T}^2}{\xi_{1,T} d_T}\to 0.$$

Thus, the contribution due to the estimation of the marginal density is of smaller order than that due to the estimation of the joint density.

(b) Given (37), it suffices to check that

$$\xi_{2,T}^2 d_T^{-1}\to 0,\ T^{-1/2}\xi_{2,T}^{-2}d_T^{-1}\to 0,\ b_M^{-1/2}\xi_{2,T}^{-3}d_T^{-1}\to 0. \tag{38}$$

Now, (i) and (ii) ensure that $\xi_{2,T}^2 d_T^{-1}\to 0$; also, (iii) ensures that $T^{-1/2}\xi_{2,T}^{-2}d_T^{-1}\to 0$. In fact, we need that $-\frac{1}{2}+2\phi_2+\delta<0$. Now, as $\delta<2\left(\phi_1-\phi_2\right)$ and $\phi_1<\frac{3}{2}\phi_2$,

$$-\frac{1}{2}+2\phi_2+\delta<-\frac{1}{2}+3\phi_2,$$

where the latter is negative for $\phi_2<1/6$. Finally, (iv) ensures that $b_M^{-1/2}\xi_{2,T}^{-3}d_T^{-1}\to 0$.

(c) We first need to find conditions under which $b_M^{-1/2}\xi_{2,T}^{-3}d_T^{-1}$, $T^{-1/2}\xi_{2,T}^{-2}d_T^{-1}$ and $\xi_{2,T}^2 d_T^{-1}$ are of the same order of magnitude. In other words, we need to find values for $\phi_2$ and $\psi$, such that

$$-2\phi_2+\delta=2\phi_2-1/2+\delta=-1/2\psi+3\phi_2+\delta$$

From the first equality, we obtain $\phi_2 = \frac{1}{8}$, and plugging in this value and solving the second equality for $\psi$, we get $\psi = \frac{5}{4}$. Thus, the rate at which all components converge is $T^{-\frac{1}{4}+\delta}$. Finally, notice that for $\psi > 5/4$, the measurement error component becomes negligible, and so the rate is determined by the variance and bias components, which are indeed of order $T^{-1/4}$. ■

Let

$$U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right) = \left(IV_{i,2,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right), IV_{i,1,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)\right)',$$

$$U_{i,N}\left(\boldsymbol{\theta}^{\dagger}\right) = \left(IV_{i,2,N}\left(\boldsymbol{\theta}^{\dagger}\right), IV_{i,1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right)', \quad U_{i}\left(\boldsymbol{\theta}^{\dagger}\right) = \left(IV_{i,2}\left(\boldsymbol{\theta}^{\dagger}\right), IV_{i,1}\left(\boldsymbol{\theta}^{\dagger}\right)\right)',$$

$$U_{i,N}\left(\boldsymbol{\theta}\right) = \left(IV_{i,2,N}\left(\boldsymbol{\theta}\right), IV_{i,1,N}\left(\boldsymbol{\theta}\right)\right)'.$$

Thus

$$\begin{aligned}
\widehat{f}_{IV_{t+1}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)|IV_t\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)}(x, RM_{T,M}) &= \frac{\widehat{f}_{IV_{t+1}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right),IV_t\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)}(x, RM_{T,M})}{\widehat{f}_{IV_{t+1}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)}(RM_{T,M})} \\
&= \frac{\frac{1}{S\zeta_{2,T}^2}\sum_{i=1}^{S}\mathbf{K}\left(\frac{IV_{i,2,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-x}{\varsigma_{2,T}}, \frac{IV_{i,1,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-RM_{T,M}}{\varsigma_{2,T}}\right)}{\frac{1}{S\varsigma_{1,T}}\sum_{i=1}^{S}K\left(\frac{IV_{i,1,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-RM_{T,M}}{\varsigma_{1,T}}\right)} \\
&= \frac{\frac{1}{S\varsigma_{2,T}^2}\sum_{i=1}^{S}\mathbf{K}\left(\frac{U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M}{\varsigma_{2,T}}\right)}{\frac{1}{S\varsigma_{1,T}}\sum_{i=1}^{S}K\left(\frac{IV_{i,1,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-RM_{T,M}}{\varsigma_{1,T}}\right)}.
\end{aligned}$$

The proof of Theorem 2 requires the following Lemma.

**Lemma 2.** *Let A1-A7 hold. If $T/b_M^2 \to 0$, $T/N^{(1-\delta)} \to 0$, $T/S \to 0$, $p_T \to \infty$ and $p_T/T^{1/4} \to 0$, and $\varphi_2 > 1/12$, $3\varphi_2 - \varphi_1 > 1/6$, then*
*(i)*

$$\widehat{f}_{IV_{t+1}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right),IV_t\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)}(x, RM_{T,M}) - \widehat{f}_{IV_{t+1}\left(\boldsymbol{\theta}^{\dagger}\right),IV_t\left(\boldsymbol{\theta}^{\dagger}\right)}(x, RM_{T,M}) = O_P\left(T^{-1/2}\varsigma_{2,T}^{-3}\right) + O\left(\varsigma_{2,T}^2\right),$$

*uniformly in $x$;*
*(ii)*

$$\widehat{f}_{IV_t\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)}(RM_{T,M}) - \widehat{f}_{IV_t\left(\boldsymbol{\theta}^{\dagger}\right)}(RM_{T,M}) = O_P\left(T^{-1/2}\varsigma_{1,T}^{-2}\right) + O\left(\varsigma_{1,T}^2\right).$$

## Proof of Lemma 2

(i)

$$\frac{1}{S\varsigma_{2,T}^2}\sum_{i=1}^{S}\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M}{\varsigma_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M\right|\leq\varsigma_{2,T}\right\}} - f_{U_i\left(\boldsymbol{\theta}^{\dagger}\right)}(u_M)$$

$$
\begin{aligned}
&= \frac{1}{S\varsigma_{2,T}^2}\sum_{i=1}^{S}\left(\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M}{\varsigma_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M\right|\leq\varsigma_{2,T}\right\}}\right.\\
&\quad\left.-\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M}{\varsigma_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M\right|\leq\varsigma_{2,T}\right\}}\right)\right) \qquad (39)\\
&\quad+\frac{1}{S\varsigma_{2,T}^2}\sum_{i=1}^{S}\left(\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M}{\varsigma_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M\right|\leq\varsigma_{2,T}\right\}}\right)\right.\\
&\quad\left.-\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M}{\varsigma_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M\right|\leq\varsigma_{2,T}\right\}}\right)\right) \qquad (40)\\
&\quad+\frac{1}{S\varsigma_{2,T}^2}\sum_{i=1}^{S}\mathrm{E}\left(\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M}{\varsigma_{2,T}}\right)\mathbf{1}_{\left\{\left|U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M\right|\leq\varsigma_{2,T}\right\}}\right)-f_{U_i\left(\boldsymbol{\theta}^\dagger\right)}(u_M) \qquad (41)\\
&\quad+\frac{1}{S\varsigma_{2,T}^2}\sum_{i=1}^{S}\left(\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M}{\varsigma_{2,T}}\right)-\widetilde{\mathbf{K}}\left(\frac{U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-u_M}{\varsigma_{2,T}}\right)\right)\\
&\quad\times\mathbf{1}_{\left\{\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M\right|\leq\varsigma_{2,T}\right\}}. \qquad (42)
\end{aligned}
$$

Via an intermediate value expansion around $\boldsymbol{\theta}^\dagger$, (42) is of smaller order in probability than

$$
\begin{aligned}
&\frac{1}{S}\sum_{i=1}^{S}\left|\widetilde{\mathbf{K}}_1'\left(\frac{U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M}{\varsigma_{2,T}}\right)\frac{\partial IV_{i,2,N}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}_{T,S,M,N}}\mathbf{1}_{\left\{\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M\right|\leq\varsigma_{2,T}\right\}}\right|\\
&\times\frac{\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}-\boldsymbol{\theta}^\dagger\right)}{\varsigma_{2,T}^3}\\
+\;&\frac{1}{S}\sum_{i=1}^{S}\left|\widetilde{\mathbf{K}}_2'\left(\frac{U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M}{\varsigma_{2,T}}\right)\frac{\partial IV_{i,1,N}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}_{T,S,M,N}}\mathbf{1}_{\left\{\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-u_M\right|\leq\varsigma_{2,T}\right\}}\right|\\
&\times\frac{\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}-\boldsymbol{\theta}^\dagger\right)}{\varsigma_{2,T}^3}\\
=\;&O_P(1)O_P\left(\frac{T^{-1/2}}{\varsigma_{2,T}^3}\right),\ \text{ with }\ \overline{\boldsymbol{\theta}}_{T,S,M,N}\in\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N},\boldsymbol{\theta}^\dagger\right),
\end{aligned}
$$

uniformly in $u_M$, because of the uniform law of large numbers and because, given A5-A7, $\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}-\boldsymbol{\theta}^\dagger\right)=O_P(T^{-1/2})$, by Lemma 4 in Corradi and Distaso (2004).

We now need to consider the term in (40). Now,

$$
U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)=\frac{\partial U_{i,N}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}_{T,S,M,N}}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}-\boldsymbol{\theta}^\dagger\right).
$$

As $\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}-\boldsymbol{\theta}^\dagger\right)=O_P(T^{-1/2})$,

$$
\Pr\left(T^{1/3}\sup_i\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-U_{i,N}\left(\boldsymbol{\theta}^\dagger\right)\right|<\varepsilon\right)
$$

$$
\begin{aligned}
&< \Pr\left(T^{-1/6}\sup_i\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial U_{i,N}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}_{T,S,M,N}}\right|<\varepsilon\right)\\
&\leq \sum_{i=1}^{S}\Pr\left(T^{-1/6}\sup_i\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial U_{i,N}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}_{T,S,M,N}}\right|<\varepsilon\right)\\
&\leq \frac{1}{\varepsilon^{14}}ST^{-\frac{7}{3}}\mathrm{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial U_{i,N}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}_{T,S,M,N}}\right|^{14}\right)<\frac{1}{\varepsilon^{14}}T^{-\frac{1}{3}}
\end{aligned}
\tag{43}
$$

for $S/T^2 \to 0$. Define

$$
\Omega_T^{+}=\left\{\omega : T^{1/3}\sup_i\left|U_{i,N}\left(\widehat{\boldsymbol{\theta}}_{T,S,M,N}\right)-U_{i,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right|<\varepsilon\right\}
$$

and, given (43), $T^{1/3}\lim_{T\to\infty}\Pr(1-\Omega_T^{+})=0$. Thus, by the same argument as the one used in the proof of Lemma 1, the term in (40) is of order $O\left(T^{-1/3}\varsigma_2^T\right)$. If $\varsigma_{2,T}T^{1/12}\to 0$, then this term is of a smaller order than the term in (42) and so can be ignored.

Finally, the terms in (39) and (41) can be treated as in Lemma 1.

(ii) By the same argument used in the proof of (i), by noting that the $O\left(T^{-1/3}\varsigma_1^T\right)$ term (due to the fact that we are using realized measures in the indicator function) is of smaller order than the term in (42), given $3\varphi_2-\varphi_1>1/6$. ■

## Proof of Theorem 2

Given Lemma 2, by the same argument used in the proof of Theorem 1. ■

## Proof of Proposition 2

By a similar argument as the one used in the Proof of Proposition 1. ■

## Proof of Proposition 3

By the same argument used in the proof of Proposition 1 in Corradi and Distaso (2004). ■

## Proof of Propositions 4, 5, 6

From Propositions 1,2 and 3 in Corradi, Distaso and Swanson (2005). ■

## References


Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2005a). 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 (2005b). 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, 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.

Andrews, D.W.K. (1990). Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity and Nonparametric Kernel Estimation. Cowles Foundation Discussion Paper 909R.

Andrews, D.W.K. (1995). Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, 11, 560-596.

Awartani, B., V. Corradi and W. Distaso (2004). Testing and Modelling Market Microstructure Effects with an Application to the Dow Jones Industrial Average. FERC Working Paper, WP04-06.
`http://www2.warwick.ac.uk/fac/soc/wbs/research/wfri/ferc/wrkingpaprseries/`

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. 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). A Feasible Central Limit Theory for Realized Volatility under Leverage. Manuscript, University of Oxford.

Barndorff-Nielsen, O.E. and N. Shephard (2004b). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics*, forthcoming.

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

Barndorff-Nielsen O.E., S.E. Graversen, J. Jacod, M. Podolskij and N. Shephard (2004). A Central Limit Theorem for Realized Power and Bipower Variations of Continuous Semimartingales. To appear in *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev* (ed. Y. Kabanov and R. Lipster), Springer and Verlag, New York.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2004). Regular and Modified Kernel-Based Estimators of Integrated Variance: The Case with Independent Noise. Manuscript, Oxford University.

Bierens, H.J. (1983). Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions. *Journal of the American Statistical Association*, 77, 699-707.

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

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Lecture Notes in Statistics #110, Springer and Verlag, New York.

Corradi, V. and W. Distaso (2004). Estimating and Testing Stochastic Volatility Models via Realized Measures. FERC Working Paper, WP04-12.
`http://www2.warwick.ac.uk/fac/soc/wbs/research/wfri/ferc/wrkingpaprseries/`

Corradi, V., W. Distaso and N.R. Swanson (2005). Predictive Confidence Intervals for Integrated Volatility. Manuscript, Queen Mary, University of London.

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

Hansen, P.R. and A. Lunde (2004). An Unbiased Version of Realized Variance. Working Paper, Stanford University.

Karatzas, I. and S.E. Shreve (1988). *Brownian Motion and Stochastic Calculus.* Springer and Verlag, New York.

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

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

Pardoux, E. and D. Talay (1985). Discretization and Simulation of Stochastic Differential Equations. *Acta Applicandae Mathematicae*, 3, 23-47.

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

Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2004). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data. *Journal of the American Statistical Association*, forthcoming.

Zhang, L. (2004). Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. Manuscript, Carnegie Mellon University.

| $M$ | *Realized Volatility* | *Bipower Variation* | *Sub − Sample 1* | *Sub − Sample 2* |
|---|---|---|---|---|
| | | *Panel A: Mean Square Error* | | |
| | | *Sample Size = 100 Daily Realized Measure Observations* | | |
| 48 | 0.053870 | 0.055665 | 0.073754 | 0.095998 |
| | (0.028011) | (0.029723) | (0.034950) | (0.044536) |
| 144 | 0.045777 | 0.045448 | 0.043273 | 0.077079 |
| | (0.024918) | (0.026034) | (0.028945) | (0.036589) |
| 360 | 0.042844 | 0.042479 | 0.033216 | 0.062194 |
| | (0.023178) | (0.023066) | (0.021217) | (0.030070) |
| 720 | 0.041658 | 0.042036 | 0.028346 | 0.056442 |
| | (0.022664) | (0.022796) | (0.019153) | (0.026939) |
| | | *Sample Size = 300 Daily Realized Measure Observations* | | |
| 48 | 0.028822 | 0.029235 | 0.046452 | 0.063649 |
| | (0.012953) | (0.013955) | (0.013735) | (0.023882) |
| 144 | 0.021065 | 0.021299 | 0.023060 | 0.048501 |
| | (0.010467) | (0.011077) | (0.011813) | (0.019617) |
| 360 | 0.019180 | 0.019231 | 0.014329 | 0.036681 |
| | (0.009712) | (0.009702) | (0.009103) | (0.016350) |
| 720 | 0.018690 | 0.018789 | 0.010195 | 0.031490 |
| | (0.009918) | (0.010040) | (0.007151) | (0.014383) |
| | | *Sample Size = 500 Daily Realized Measure Observations* | | |
| 48 | 0.022632 | 0.022739 | 0.043977 | 0.050315 |
| | (0.008907) | (0.009246) | (0.008954) | (0.013776) |
| 144 | 0.016424 | 0.016745 | 0.018705 | 0.038360 |
| | (0.007479) | (0.007768) | (0.007018) | (0.012277) |
| 360 | 0.014854 | 0.015086 | 0.010255 | 0.028555 |
| | (0.006648) | (0.006645) | (0.005631) | (0.009963) |
| 720 | 0.014333 | 0.014320 | 0.007144 | 0.024457 |
| | (0.006551) | (0.006665) | (0.004324) | (0.009894) |
| | | *Panel B: Absolute Deviation* | | |
| | | *Sample Size = 100 Daily Realized Measure Observations* | | |
| 48 | 0.188981 | 0.192932 | 0.240104 | 0.265386 |
| | (0.049611) | (0.052319) | (0.060206) | (0.065314) |
| 144 | 0.173871 | 0.172523 | 0.173078 | 0.232526 |
| | (0.046587) | (0.049388) | (0.058981) | (0.056756) |
| 360 | 0.167663 | 0.167202 | 0.149160 | 0.205365 |
| | (0.045167) | (0.045220) | (0.049198) | (0.050795) |
| 720 | 0.165309 | 0.166226 | 0.136624 | 0.195169 |
| | (0.045418) | (0.045401) | (0.048126) | (0.046704) |
| | | *Sample Size = 300 Daily Realized Measure Observations* | | |
| 48 | 0.139506 | 0.140570 | 0.191801 | 0.219076 |
| | (0.031400) | (0.034158) | (0.030577) | (0.042557) |
| 144 | 0.117934 | 0.118605 | 0.129916 | 0.186882 |
| | (0.030252) | (0.031963) | (0.035176) | (0.039041) |
| 360 | 0.112833 | 0.113064 | 0.098203 | 0.158514 |
| | (0.029319) | (0.029121) | (0.033662) | (0.035900) |
| 720 | 0.111152 | 0.111202 | 0.082443 | 0.146051 |
| | (0.030273) | (0.030432) | (0.030654) | (0.033170) |
| | | *Sample Size = 500 Daily Realized Measure Observations* | | |
| 48 | 0.124125 | 0.125183 | 0.187133 | 0.195883 |
| | (0.025349) | (0.025932) | (0.020353) | (0.029999) |
| 144 | 0.105380 | 0.106080 | 0.117595 | 0.166842 |
| | (0.025125) | (0.025693) | (0.025684) | (0.029706) |
| 360 | 0.100509 | 0.101381 | 0.084131 | 0.140632 |
| | (0.024462) | (0.024481) | (0.025054) | (0.026286) |
| 720 | 0.098552 | 0.098384 | 0.069731 | 0.129232 |
| | (0.025041) | (0.025325) | (0.022508) | (0.027355) |
| | | *Mean Absolute Percentage Error* | | |
| | | *Sample Size = 100 Daily Realized Measure Observations* | | |
| 48 | 0.236586 | 0.241727 | 0.335532 | 0.338964 |
| | (0.058507) | (0.062586) | (0.078656) | (0.086477) |
| 144 | 0.222331 | 0.219939 | 0.228168 | 0.290530 |
| | (0.056065) | (0.059807) | (0.076758) | (0.073018) |
| 360 | 0.214781 | 0.214612 | 0.193540 | 0.255875 |
| | (0.056244) | (0.055699) | (0.062155) | (0.059796) |
| 720 | 0.212774 | 0.213677 | 0.179604 | 0.244154 |
| | (0.057475) | (0.057360) | (0.059741) | (0.054898) |
| | | *Sample Size = 300 Daily Realized Measure Observations* | | |
| 48 | 0.173948 | 0.174956 | 0.277997 | 0.281771 |
| | (0.039793) | (0.042811) | (0.047112) | (0.058474) |
| 144 | 0.151525 | 0.151716 | 0.177804 | 0.234718 |
| | (0.041707) | (0.042811) | (0.049256) | (0.051303) |
| 360 | 0.147569 | 0.147146 | 0.128852 | 0.195912 |
| | (0.040791) | (0.040118) | (0.044552) | (0.045072) |
| 720 | 0.146160 | 0.145520 | 0.111250 | 0.181255 |
| | (0.042977) | (0.042500) | (0.040404) | (0.040681) |
| | | *Sample Size = 500 Daily Realized Measure Observations* | | |
| 48 | 0.154497 | 0.156075 | 0.273693 | 0.253704 |
| | (0.033575) | (0.033007) | (0.033506) | (0.042301) |
| 144 | 0.137003 | 0.136042 | 0.163027 | 0.209215 |
| | (0.034882) | (0.034338) | (0.039367) | (0.040504) |
| 360 | 0.132830 | 0.133188 | 0.112601 | 0.173071 |
| | (0.035757) | (0.035239) | (0.033780) | (0.034227) |
| 720 | 0.131184 | 0.130644 | 0.095137 | 0.160254 |
| | (0.036891) | (0.037310) | (0.029174) | (0.035133) |

* Notes: Entries report mean square error and related predictive error loss measures that are based upon comparison of pseudo true predictive integrated volatility density values with those constructed using various realized measures formed using sets of $M$ intraday simulated observable state variable data, for various daily sample sizes. The range of the data across which to evaluate the estimated densities was set to 100 equally spaced values from a large simulated sample of $IV$ data, for the interval $[\overline{IV} - \widehat{\mathrm{SE}}(IV), \overline{IV} + \widehat{\mathrm{SE}}(IV)]$. See Section 6 for further details.

| $M$ | *Realized Volatility* | *Bipower Variation* | *Sub − Sample 1* | *Sub − Sample 2* |
|---|---|---|---|---|
| | | *Panel A: Noise* = i.i.d.$N(0,(0.5\times 1440)^{-1})$ | | |
| | | *Mean Square Error* | | |
| 48 | 0.023634 | 0.021947 | 0.098536 | 0.029185 |
| | (0.018776) | (0.023383) | (0.458070) | (0.028641) |
| 144 | 0.064163 | 0.065729 | 0.034548 | 0.023347 |
| | (0.023403) | (0.023196) | (0.062030) | (0.029805) |
| 360 | 0.263250 | 0.290226 | 0.022895 | 0.023367 |
| | (0.025255) | (0.024631) | (0.033408) | (0.039845) |
| 720 | 0.481822 | NA | 0.023558 | 0.021236 |
| | (0.328727) | NA | (0.041435) | (0.024881) |
| | | *Mean Absolute Deviation* | | |
| 48 | 0.125476 | 0.119375 | 0.230040 | 0.141292 |
| | (0.051502) | (0.052888) | (0.157586) | (0.056501) |
| 144 | 0.222514 | 0.225302 | 0.135356 | 0.121413 |
| | (0.040080) | (0.039833) | (0.087225) | (0.060361) |
| 360 | 0.454759 | 0.481319 | 0.115923 | 0.118471 |
| | (0.024858) | (0.024268) | (0.067095) | (0.067344) |
| 720 | 0.650182 | NA | 0.116885 | 0.115182 |
| | (0.067010) | NA | (0.069086) | (0.059355) |
| | | *Mean Absolute Percentage Error* | | |
| 48 | 0.193329 | 0.182913 | 0.376544 | 0.222798 |
| | (0.078471) | (0.079412) | (0.252663) | (0.089189) |
| 144 | 0.355870 | 0.360370 | 0.215683 | 0.188564 |
| | (0.063781) | (0.063299) | (0.138491) | (0.093522) |
| 360 | 0.678564 | 0.715640 | 0.180751 | 0.181184 |
| | (0.036460) | (0.035842) | (0.102403) | (0.102182) |
| 720 | 1.014426 | NA | 0.182902 | 0.175388 |
| | (0.142956) | NA | (0.105107) | (0.089049) |
| | | *Panel B: Noise* = i.i.d.$N(0,(1440)^{-1})$ | | |
| | | *Mean Square Error* | | |
| 48 | 0.022364 | 0.020362 | 0.049526 | 0.026017 |
| | (0.015850) | (0.015011) | (0.029492) | (0.016029) |
| 144 | 0.034303 | 0.033039 | 0.017617 | 0.020784 |
| | (0.018292) | (0.017939) | (0.014197) | (0.014413) |
| 360 | 0.092704 | 0.100991 | 0.015277 | 0.019279 |
| | (0.024913) | (0.024982) | (0.012597) | (0.014351) |
| 720 | 0.266881 | 0.297374 | 0.016220 | 0.019797 |
| | (0.024044) | (0.022913) | (0.012469) | (0.014826) |
| | | *Mean Absolute Deviation* | | |
| 48 | 0.122231 | 0.116203 | 0.191830 | 0.137637 |
| | (0.048673) | (0.046974) | (0.062965) | (0.047308) |
| 144 | 0.158116 | 0.154575 | 0.106559 | 0.119460 |
| | (0.044247) | (0.044540) | (0.047763) | (0.045933) |
| 360 | 0.268330 | 0.280626 | 0.099561 | 0.112874 |
| | (0.035813) | (0.034774) | (0.043763) | (0.047031) |
| 720 | 0.457690 | 0.487453 | 0.103695 | 0.114165 |
| | (0.023591) | (0.022716) | (0.043095) | (0.047574) |
| | | *Mean Absolute Percentage Error* | | |
| 48 | 0.185577 | 0.175721 | 0.311322 | 0.212097 |
| | (0.072429) | (0.068862) | (0.099865) | (0.071829) |
| 144 | 0.248965 | 0.243625 | 0.167605 | 0.180836 |
| | (0.069022) | (0.069458) | (0.073303) | (0.066432) |
| 360 | 0.424465 | 0.444162 | 0.154502 | 0.169177 |
| | (0.055356) | (0.054443) | (0.065684) | (0.067548) |
| 720 | 0.682359 | 0.724261 | 0.162286 | 0.171107 |
| | (0.034150) | (0.033646) | (0.065899) | (0.068964) |
| | | *Panel C: Noise* = i.i.d.$N(0,(2\times 1440)^{-1})$ | | |
| | | *Mean Square Error* | | |
| 48 | 0.021719 | 0.019801 | 0.049092 | 0.026016 |
| | (0.015822) | (0.015106) | (0.029676) | (0.016129) |
| 144 | 0.026526 | 0.025115 | 0.017488 | 0.021031 |
| | (0.016575) | (0.015719) | (0.014223) | (0.014577) |
| 360 | 0.043186 | 0.043747 | 0.015396 | 0.019621 |
| | (0.020275) | (0.020104) | (0.012785) | (0.014348) |
| 720 | 0.094308 | 0.104114 | 0.016668 | 0.020113 |
| | (0.025476) | (0.025880) | (0.012915) | (0.015034) |
| | | *Mean Absolute Deviation* | | |
| 48 | 0.120176 | 0.114174 | 0.190779 | 0.137585 |
| | (0.048803) | (0.048095) | (0.063387) | (0.047790) |
| 144 | 0.136212 | 0.132317 | 0.105977 | 0.120308 |
| | (0.045642) | (0.044667) | (0.047744) | (0.046114) |
| 360 | 0.179547 | 0.181078 | 0.099895 | 0.114161 |
| | (0.043652) | (0.043078) | (0.044144) | (0.046737) |
| 720 | 0.271028 | 0.285306 | 0.105120 | 0.115111 |
| | (0.036244) | (0.035354) | (0.043999) | (0.048060) |
| | | *Mean Absolute Percentage Error* | | |
| 48 | 0.181875 | 0.172265 | 0.309536 | 0.211937 |
| | (0.072050) | (0.070549) | (0.100485) | (0.072448) |
| 144 | 0.211320 | 0.204913 | 0.166340 | 0.181935 |
| | (0.069706) | (0.068446) | (0.073093) | (0.066678) |
| 360 | 0.285055 | 0.288008 | 0.154824 | 0.170972 |
| | (0.068402) | (0.067658) | (0.066069) | (0.066966) |
| 720 | 0.429704 | 0.452715 | 0.164250 | 0.172409 |
| | (0.056172) | (0.055448) | (0.067260) | (0.069574) |

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

| $M$ | *Realized Volatility* | *Bipower Variation* | *Sub − Sample 1* | *Sub − Sample 2* |
|---|---|---|---|---|
| *Panel A: One* i.i.d.$N(0, 0.64 \times \mathrm{E}(IV_t))$ *Jump Every 5 Days* | | | | |
| *Mean Square Error* | | | | |
| 48 | 0.144282 | 0.133275 | 0.111008 | 0.149304 |
| | (0.023653) | (0.027431) | (0.053430) | (0.051297) |
| 144 | 0.139575 | 0.142283 | 0.092840 | 0.133265 |
| | (0.022013) | (0.026829) | (0.042424) | (0.045331) |
| 360 | 0.138015 | 0.150057 | 0.081050 | 0.119391 |
| | (0.022929) | (0.030385) | (0.035938) | (0.041793) |
| 720 | 0.137182 | 0.150156 | 0.072408 | 0.111794 |
| | (0.022836) | (0.029449) | (0.032069) | (0.039420) |
| *Mean Absolute Deviation* | | | | |
| 48 | 0.325440 | 0.315792 | 0.297938 | 0.339042 |
| | (0.026261) | (0.032255) | (0.073026) | (0.062056) |
| 144 | 0.327073 | 0.337826 | 0.261602 | 0.314090 |
| | (0.026458) | (0.032870) | (0.064234) | (0.057707) |
| 360 | 0.326980 | 0.349090 | 0.238300 | 0.292410 |
| | (0.027335) | (0.035934) | (0.054750) | (0.054195) |
| 720 | 0.326562 | 0.349791 | 0.222487 | 0.280613 |
| | (0.027339) | (0.035073) | (0.051296) | (0.052627) |
| *Mean Absolute Percentage Error* | | | | |
| 48 | 0.451091 | 0.458656 | 0.402300 | 0.434432 |
| | (0.044360) | (0.060193) | (0.093534) | (0.083214) |
| 144 | 0.485407 | 0.542330 | 0.334005 | 0.394789 |
| | (0.055779) | (0.066547) | (0.084554) | (0.075377) |
| 360 | 0.492771 | 0.576933 | 0.298630 | 0.363185 |
| | (0.055249) | (0.069065) | (0.067737) | (0.068037) |
| 720 | 0.495093 | 0.584380 | 0.277939 | 0.347582 |
| | (0.054141) | (0.068187) | (0.062904) | (0.065759) |
| *Panel B: One* i.i.d.$N(0, 0.64 \times 3 \times \mathrm{E}(IV_t))$ *Jump Every 5 Days* | | | | |
| *Mean Square Error* | | | | |
| 48 | 0.223485 | 0.139572 | 0.211826 | 0.265271 |
| | (0.051631) | (0.027181) | (0.089160) | (0.079232) |
| 144 | 0.200653 | 0.143716 | 0.226750 | 0.260806 |
| | (0.045961) | (0.027634) | (0.082941) | (0.076183) |
| 360 | 0.196553 | 0.150798 | 0.227445 | 0.257420 |
| | (0.043549) | (0.032258) | (0.077804) | (0.074407) |
| 720 | 0.194077 | 0.151596 | 0.222624 | 0.254306 |
| | (0.042500) | (0.029691) | (0.075733) | (0.073837) |
| *Mean Absolute Deviation* | | | | |
| 48 | 0.407790 | 0.320696 | 0.412592 | 0.462877 |
| | (0.056069) | (0.029645) | (0.094558) | (0.079086) |
| 144 | 0.384516 | 0.337640 | 0.422375 | 0.456395 |
| | (0.048356) | (0.032754) | (0.087711) | (0.077892) |
| 360 | 0.380547 | 0.349315 | 0.419720 | 0.451196 |
| | (0.045522) | (0.037239) | (0.084946) | (0.077381) |
| 720 | 0.377833 | 0.351351 | 0.413667 | 0.447094 |
| | (0.044201) | (0.034673) | (0.084104) | (0.078405) |
| *Mean Absolute Percentage Error* | | | | |
| 48 | 0.517496 | 0.452164 | 0.548307 | 0.605975 |
| | (0.070684) | (0.044784) | (0.127808) | (0.118153) |
| 144 | 0.494314 | 0.532773 | 0.548666 | 0.593166 |
| | (0.050939) | (0.066360) | (0.127268) | (0.117351) |
| 360 | 0.491051 | 0.572192 | 0.539174 | 0.582774 |
| | (0.046094) | (0.071289) | (0.124280) | (0.117380) |
| 720 | 0.487681 | 0.585180 | 0.529439 | 0.575835 |
| | (0.043446) | (0.067117) | (0.122332) | (0.118434) |
| *Panel C: One* i.i.d.$N(0, 0.64 \times 5 \times \mathrm{E}(IV_t))$ *Jump Every 5 Days* | | | | |
| *Mean Square Error* | | | | |
| 48 | 0.305656 | 0.144657 | 0.297220 | 0.351849 |
| | (0.070500) | (0.027574) | (0.098630) | (0.074695) |
| 144 | 0.279918 | 0.148138 | 0.328315 | 0.353022 |
| | (0.068412) | (0.029469) | (0.081995) | (0.072312) |
| 360 | 0.275542 | 0.154826 | 0.335798 | 0.354874 |
| | (0.068330) | (0.033539) | (0.079966) | (0.072359) |
| 720 | 0.272757 | 0.157227 | 0.333624 | 0.354244 |
| | (0.067245) | (0.032349) | (0.079103) | (0.074101) |
| *Mean Absolute Deviation* | | | | |
| 48 | 0.494553 | 0.326322 | 0.496872 | 0.546059 |
| | (0.073259) | (0.030058) | (0.093704) | (0.067478) |
| 144 | 0.468434 | 0.342204 | 0.524310 | 0.546529 |
| | (0.072569) | (0.034736) | (0.076079) | (0.065959) |
| 360 | 0.464017 | 0.353723 | 0.529770 | 0.547601 |
| | (0.072597) | (0.038517) | (0.075786) | (0.066556) |
| 720 | 0.461241 | 0.357720 | 0.527285 | 0.546460 |
| | (0.071641) | (0.037297) | (0.076073) | (0.069101) |
| *Mean Absolute Percentage Error* | | | | |
| 48 | 0.642378 | 0.458218 | 0.666059 | 0.731700 |
| | (0.115489) | (0.051364) | (0.133621) | (0.104542) |
| 144 | 0.603734 | 0.537271 | 0.699370 | 0.731113 |
| | (0.109214) | (0.070743) | (0.115908) | (0.103420) |
| 360 | 0.597831 | 0.578814 | 0.705339 | 0.731934 |
| | (0.107915) | (0.074945) | (0.117669) | (0.105038) |
| 720 | 0.593559 | 0.595093 | 0.700783 | 0.729717 |
| | (0.105708) | (0.071941) | (0.118685) | (0.109426) |

* Notes: See notes to Table 2.