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

Valentina Corradi[†] Walter Distaso[‡] Norman R. Swanson[§]

University of Warwick Imperial College Rutgers University

this revision: September 2006

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 and microstructure noise in prices. 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.

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

JEL classification: C22, C53, C14.

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

[‡]Imperial College, Tanaka Business School, South Kensington campus, London SW7 2AZ, UK, email: `w.distaso@imperial.ac.uk`.

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

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. Galbraith and Zinde Walsh (2006) use realized volatility measure to obtain consistent Least Square and Least Absolute Deviation deviation estimators of GARCH parameters, regardless the implied volatility measurement error. More recently, Andersen, Bollerslev and Meddahi (2006), Aït-Sahalia and Mancini (2006) and Ghysels and Sinko (2006) have considered ARMA models constructed using microstructure robust measures. 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. An important reason for paying attention to predictive densities of volatility is the recent development of numerous volatility-based derivative products. Examples include volatility options on various currencies such as the British pound and the Japanese Yen; and VOLAX futures, which are based upon the implied volatility of DAX index options.

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 density of daily volatility (based on a given realized volatility measure), conditional on recent observed values of the realized measure itself. We provide general conditions on the moments of the measurement error between the realized measure and integrated volatility. 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, 2004, 2006; and the robust subsampled realized volatility measures of (i) Zhang, Mykland and Aït-Sahalia, 2005 and (ii) Aït-Sahalia, Mykland and Zhang, 2006, Zhang, 2006 and Barndorff-Nielsen, Hansen, Lunde and Shephard, 2006a,b) 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 loss. We also compare the relative accuracy of predictive densities based on realized measures and on simulated integrated volatility. 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,

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

where $W_{1,t}$ 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

---

[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 $\mathcal{A}\mathrm{P}_i(h_t) = -\lambda_i\mathrm{P}_i(h_t)$.

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. By modeling N_t as a finite activity counting process, in this paper we consider the case of a finite number of jumps occurring 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 function for $\psi(\cdot)$ (such as, for example, the square root function), 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.[2]

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.[3]

In what follows, we assume to have MT observations for (1), consisting of M intradaily observations for T days. We allow for the possibility that proces are recorded up to a measurement error, by observing

$$X_{t+j/M} = Y_{t+j/M} + \epsilon_{t+j/M}, \quad t = 1, \ldots, T \text{ and } j = 1, \ldots, M, \qquad (3)$$

where

$$\epsilon_{t+j/M} \sim \text{i.i.d.}(0, \nu) \text{ and } \mathrm{E}(\epsilon_{t+j/M} Y_{s+i/M}) = 0 \text{ for all } t, s, j, i. \qquad (4)$$

Thus, according to (3), 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 (2005), Zhang, Mykland and Aït-Sahalia (2005), Bandi and Russell (2004, 2006).[4] Needless to say, when $\nu = 0$, then $\epsilon_{t+j/M} = 0$ (almost surely), and therefore $X_{t+j/M} = Y_{t+j/M}$.

---

[2]The a_i's are real numbers and 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 $i \neq 0$. When p is infinite, we also require $\sum_{i=1}^\infty a_i < \infty$.

[3]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.

[4]Recently, Aït-Sahalia, Mykland and Zhang (2006), and Barndorff-Nielsen, Hansen, Lunde and Shephard (2006a) allow for some dependence in the microstructure noise, while Awartani, Corradi and Distaso (2006) allow for correlation between the underlying price process and the microstructure noise.

The daily integrated volatility process at day t is defined as

$$IV_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s. \tag{5}$$

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

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

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 paper, we derive consistent estimators of predictive conditional densities of integrated volatility as follows. First, we construct functionals of kernel estimators of conditional densities, based on the observable (but noisy) realized measure $RM_{t,M}$, instead of the unobservable IV_t. Second, 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 $RM_{t,M}$ are uniformly consistent for the conditional density of IV_t. Third, 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:

$$RV_{t,M} = \sum_{j=1}^{M-1} \left(X_{t+(j+1)/M} - X_{t+j/M} \right)^2; \tag{6}$$

(b) normalized bipower variation, given by:

$$(\mu_1)^{-1} BV_{t,M} = (\mu_1)^{-1} \frac{M}{M-1} \sum_{j=2}^{M-1} \left| X_{t+(j+1)/M} - X_{t+j/M} \right| \left| X_{t+j/M} - X_{t+(j-1)/M} \right|, \tag{7}$$

where $\mu_1 = \mathrm{E}\,|Z| = 2^{1/2} \Gamma(1) / \Gamma(1/2)$ 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 (2005), defined as

$$\widehat{RV}_{t,l,M} = RV_{t,l,M}^{avg} - 2l\widehat{\nu}_{t,M}, \tag{8}$$

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

$$RV_{t,l,M}^{avg} = \frac{1}{B}\sum_{b=1}^{B} RV_{t,l}^{b} = \frac{1}{B}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(X_{t+\frac{jB+b}{M}} - X_{t+\frac{(j-1)B+b}{M}}\right)^2. \tag{9}$$

Here $Bl \cong M$, $l = O(M^{1/3})$, l denotes the subsample size and B the number of subsamples. The logic underlying (8) 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 (2006), Aït-Sahalia, Mykland and Zhang (2006). See also, Barndorff-Nielsen, Hansen, Lunde and Shephard (2006a,b) who suggest asymptotically equivalent estimators. Define,

$$
\begin{aligned}
\widetilde{RV}_{t,e,M} &= \sum_{i=1}^{e} a_i \widetilde{RV}_{t,e_i,M} + \frac{RV_{t,M}}{M} \\
&= \sum_{i=1}^{e} a_i \frac{1}{e_i}\left(\sum_{j=1}^{M-e_i}\left(X_{t+\frac{j+e_i}{M}} - X_{t+\frac{j}{M}}\right)^2\right) + \frac{RV_{t,M}}{M}, \tag{10}
\end{aligned}
$$

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 (2006) 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.

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, $\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})$, as the ratio of the estimators of the joint density, $\widehat{f}_{RM_{T+1,M},RM_{T,M}}(x, RM_{T,M})$, and of the marginal density,

$\widehat{f}_{RM_{T,M}}(RM_{T,M}),$

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

$$= \frac{\widehat{f}_{RM_{T+1,M},RM_{T,M}}(x, RM_{T,M})}{\widehat{f}_{RM_{T,M}}(RM_{T,M})} \qquad (11)$$

Notice that we use different bandwidth sequences, $\xi_{2,T}$ and $\xi_{1,T}$, for the joint and marginal density estimators.[5]

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,

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

(ii) $\mathrm{Var}(N_{t,M}) = O(b_M^{-1})$

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 λ_i is the eigenvalue associated with the $i-th$ eigenfunction $\mathrm{P}_i(h_t)$.

Assumption A4:

(i) the kernel, $\mathbf{K}$, is a symmetric, nonnegative, continuous function with bounded support $[-\Gamma, \Gamma]^2$; and is at least twice continuously differentiable on the interior of its support, satisfying:

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

Let $K_i^{(j)}$ be the $j-th$ derivative of the kernel with respect to the $i-$th variable. Then, $K_i^{(j)}(-\Gamma) = K_i^{(j)}(\Gamma) = 0$, for $i = 1, 2, j = 1, \ldots, J, J \geq 1$.

(ii) the kernel K is a symmetric, nonnegative, continuous function with bounded support $[-\Gamma, \Gamma]$, at least twice differentiable on the interior of its support, satisfying:

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

Let $K^{(j)}$ be the $j-th$ derivative of the kernel. Then, $K^{(j)}(-\Gamma) = K^{(j)}(\Gamma) = 0$, for $j = 1, \ldots, J$, $J \geq 1$.

---

[5]In the sequel, we consider one step ahead predictive densities; however, the case of $\tau-$step ahead predictive densities, for τ finite, can be treated in an analogous manner. Multistep ahead estimators are analyzed in the empirical illustration.

(iii) $f_{IV_T}(\cdot)$ 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, because of assumption A2, 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 $t = 1, 2, \ldots, T$. We can now state the following.

Theorem 1. *Let assumptions A1-A4 hold. Then, for all $RM_{T,M}$ such that $\widehat{f}_{RM_{T,M}}(RM_{T,M}) > d_T$, with $d_T = o(1)$, uniformly in $x \in \mathbb{R}_+$,*

$$
\begin{aligned}
& \left| \widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M}) \right| \\
= {} & O_P(b_M^{-1/2}\xi_{2,T}^{-3}d_T^{-1}) + O_P(b_M^{-1/2}\xi_{1,T}^{-2}d_T^{-2}) + O_P(T^{-1/2}\xi_{2,T}^{-2}d_T^{-1}) \\
& + O_P(T^{-1/2}\xi_{1,T}^{-1/2}d_T^{-2}) + O\left(\xi_{2,T}^2 d_T^{-1}\right) + O\left(\xi_{1,T}^2 d_T^{-2}\right).
\end{aligned}
\tag{12}
$$

Note that the first two terms above reflect the measurement error due to the fact that we compute densities using 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 associated with any nonparametric estimator.

In the proposition below we 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 $x \in \mathbb{R}_+$, so that we have a uniform rate; (iii) the measurement error component is of the same or smaller order than the variance and the bias components, so that that it does not "slow down" the convergence to the true conditional density.

Proposition 1. *Let $\xi_{1,T} = cT^{-\phi_1}$, $\xi_{2,T} = cT^{-\phi_2}$, $b_M = cT^{\psi}$, $d_T = cT^{-\delta}$, $\psi, \delta \geq 0$, $\phi_1 > \phi_2 > 0$, with c denoting a generic positive constant, and let assumptions of Theorem 1 hold. Then, for all $RM_{T,M}$ such that $\widehat{f}_{RM_{T,M}}(RM_{T,M}) > d_T$, uniformly in $x \in \mathbb{R}_+$,[6]*
(a) if (i) $\phi_2 < \phi_1 < 3/2\phi_2$, and (ii) $\delta < \min\{3\phi_2 - 2\phi_1, 2(\phi_1 - \phi_2)\}$,

$$
\begin{aligned}
& \left| \widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M}) \right| \\
= {} & O_P(b_M^{-1/2}\xi_{2,T}^{-3}d_T^{-1}) + O_P(T^{-1/2}\xi_{2,T}^{-2}d_T^{-1}) + O(\xi_{2,T}^2 d_T^{-1});
\end{aligned}
\tag{13}
$$

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

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

---

[6]Theorem 1 and Proposition 1 deal with the case in which we condition only on current volatility, $RM_{T,M}$. If instead we would condition on say $RM_{T,M}, \ldots, RM_{T-(d-1),M}$, then the measurement error term, the variance and the bias term would be $O_P(b_M^{-1/2}\xi_{2,T}^{-(d+2)}d_T^{-1})$, $O_P(T^{-1/2}\xi_{2,T}^{-(d+1)}d_T^{-1})$, $O(\xi_{2,T}^2 d_T^{-1})$, respectively.

$$= o_P(1);$$

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

$$\left| \widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M}) \right|$$
$$= O_P(T^{-1/4}d_T^{-1}).$$

First, from (a) above note that the error due to the variance component, i.e. the second term on the right hand side of (13), is of a larger order of probability than the typical one occurring in the pointwise case or when the supremum is taken over a bounded set (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 reason why, when taking the supremum on $\mathbb{R}_+$, we have a slower rate for the variance, 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 $\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M}) < c$, where c is an arbitrarily small positive constant .

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 we get a "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 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 ψ, as it is implied by our measure. First, it is immediate to see that, whenever $\psi < 1 + 2\phi_2$, 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 ψ, we define that value for ϕ_2 which equalizes the order of magnitude of the measurement error term and of the bias term.

Thus, we want to find ϕ_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 $T^{-\frac{2\psi}{10}}$.

Realized measures are by definition positive, thus in the neighborhood of the boundary are downward biased, as some weight would be given also to negative observations. Hence, Theorem 1 and Proposition 1 apply only when x and $RM_{T,M}$ are not too close to the boundary.

In order to overcome this issue, we can use local linear estimators (see e.g. Fan and Gijbels, 1997 and Fan and Yao, 2005).[7]Define

$$
\begin{aligned}
& \widehat{\alpha}_{T,M}\left(x|RM_{T,M}\right), \widehat{\beta}_{T,M}\left(x|RM_{T,M}\right) \\
=\ & \arg\min_{\alpha,\beta} \frac{1}{T\zeta_{1T}\zeta_{2,T}} \sum_{t=1}^{T-1}\left(K\left(\frac{RM_{t+1,M}-x}{\zeta_{2,T}}\right)-\alpha-\beta\left(RM_{t,M}-RM_{T,M}\right)\right)^2 W\left(\frac{RM_{t,M}-RM_{T,M}}{\zeta_{1,T}}\right),
\end{aligned}
$$

where K and W are two one-dimensional kernel functions satisfying A4(ii), and possibly $W \equiv K$. Note that

$$
\begin{aligned}
\widehat{\alpha}_{T,M}\left(x|RM_{T,M}\right) \ = \ & \frac{1}{T\zeta_{1,T}\zeta_{2,T}} \sum_{t=1}^{T-1} K\left(\frac{RM_{t+1,M}-x}{\zeta_{2,T}}\right) W\left(\frac{RM_{t,M}-RM_{T,M}}{\zeta_{1,T}}\right) \\
& \times \left(\frac{s_{T,M,2}(RM_{T,M})-(RM_{t,M}-RM_{T,M})\,s_{T,M,1}(RM_{T,M})}{s_{T,M,0}(RM_{T,M})s_{T,M,2}(RM_{T,M})-s_{T,M,2}^2(RM_{T,M})}\right),
\end{aligned}
$$

where

$$
s_{T,M,j}(RM_{T,M}) = \frac{1}{T\zeta_{1,T}} \sum_{t=1}^{T-1} \left(\frac{RM_{t,M}-RM_{T,M}}{\zeta_{1,T}}\right)^j W\left(\frac{RM_{t,M}-RM_{T,M}}{\zeta_{1,T}}\right),\ j=0,1,2
$$

Provided that W is a second order symmetric kernel, as $T,M \to \infty$, $s_{T,M,1}(RM_{T,M}) \xrightarrow{P} 0$, and more precisely

$$
\begin{aligned}
\widehat{\alpha}_{T,M}\left(x|RM_{T,M}\right) \ = \ & \frac{1}{T\zeta_{1,T}\xi_{2,T}} \sum_{t=1}^{T-1} K\left(\frac{RM_{t+1,M}-x}{\zeta_{2,T}}\right) W\left(\frac{RM_{t,M}-RM_{T,M}}{\zeta_{1,T}}\right)\frac{1}{s_{T,M,0}(RM_{T,M})} \\
& +O_P\left(b_M^{-1/2}\zeta_{1,T}^{-2}\right)+O_P\left(T^{-1/2}\zeta_{1,T}^{-1/2}\right)+O\left(\zeta_{1,T}^2\right). \quad\quad (14)
\end{aligned}
$$

By an analogous argument as that used in the proof of Theorem 1, for all $RM_{T,M}$ such that $s_{T,M,0}(RM_{T,M}) > d_T$, uniformly in $x \in \mathbb{R}_+$,

$$
\begin{aligned}
& \left|\widehat{\alpha}_{T,M}\left(x|RM_{T,M}\right)-f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| \\
=\ & O_P\left(b_M^{-1/2}\left(\zeta_{1,T}^{-2}\zeta_{2,T}^{-1}+\zeta_{1,T}^{-1}\zeta_{2,T}^{-2}\right)d_T^{-1}\right)+O_P\left(T^{-1/2}\zeta_{1,T}^{-1}\zeta_{2,T}^{-1}d_T^{-1}\right)+O\left(\zeta_{1,T}^2 d_T^{-1}+\zeta_{2,T}^2 d_T^{-1}\right) \\
& +O_P\left(b_M^{-1/2}\zeta_{1,T}^{-2}d_T^{-2}\right)+O_P\left(T^{-1/2}\zeta_{1,T}^{-1/2}d_T^{-2}\right)+O\left(\zeta_{1,T}^2 d_T^{-2}\right).
\end{aligned}
$$

Also, if $\zeta_{2,T}d_T^{-1} \to 0$ and $\zeta_{1,T}^2 / \left(\zeta_{2,T}^2 d_T\right) \to 0$, then, for all $RM_{T,M}$ such that $s_{T,M,0}(RM_{T,M}) > d_T$, uniformly in $x \in \mathbb{R}_+$,

$$
\begin{aligned}
& \left|\widehat{\alpha}_{T,M}\left(x|RM_{T,M}\right)-f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right| \\
=\ & O_P\left(b_M^{-1/2}\zeta_{1,T}^{-2}\zeta_{2,T}^{-1}d_T^{-1}\right)+O_P\left(T^{-1/2}\zeta_{1,T}^{-1}\zeta_{2,T}^{-1}d_T^{-1}\right)+O\left(\zeta_{2,T}^2 d_T^{-1}\right)
\end{aligned}
$$

---

[7]Alternatively, one can simply use boundary corrected kernel functions (see e.g. Müller, 1991). Local polynomial estimators are in general preferrable, as they have better finite sample properties, because of their greater flexibility.

Note that the first term on the RHS of (14) corresponds to a "standard" conditional density estimator in which the joint density is constructed using the product kernel $\mathbf{K} = K \times W$ and different bandwidths for the dependent and the conditioning variables ($\zeta_{2,T}$ and $\zeta_{1,T}$, respectively), and the marginal density is constructed using the kernel W with bandwidth $\zeta_{1,T}$.

Of course, if one constructs predictive densities for log-volatility, then the boundary problem does not arise. By Lemma 2 in Corradi and Distaso (2006), the results stated in this section also apply to the prediction of the log of volatility. However, it should be pointed out that, contrary to the marginal density case, we cannot in general recover the conditional density of volatility from the one of its logarithm.

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 (2006, Theorem 2). 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.

$$
\begin{aligned}
& h_{i,j\frac{k+1}{N}}(\boldsymbol{\theta}) \\
= \ & h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}) + \mu(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta}) - \frac{1}{2}\sigma'(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta})\sigma(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta})\frac{k+1}{N} \\
& + \sigma(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta})\left(W_{i,j\frac{k+1}{N}} - W_{i,(j-1)\frac{k+1}{N}}\right) \\
& + \frac{1}{2}\sigma'(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta})\sigma(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta})\left(W_{i,j\frac{k+1}{N}} - W_{i,(j-1)\frac{k+1}{N}}\right)^2,
\end{aligned} \tag{15}
$$

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. $N(0, \frac{k+1}{N})$ and $h_{i,0}(\boldsymbol{\theta})$ 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:

$$
IV_{i,\tau,N}(\boldsymbol{\theta}) = \frac{1}{N/(k+1)} \sum_{j=1}^{N/(k+1)} \sigma^2_{i,\tau-1+j\frac{k+1}{N}}(\boldsymbol{\theta}), \quad \tau = 1, \ldots, k+1, \tag{16}
$$

and

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

11

Also, averaging the quantity calculated in (16) over the number of simulations S and over the length of the path $k + 1$ yields respectively

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

and

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

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

$$\overline{\mathbf{g}}^{*}_{T,M} - \overline{\mathbf{g}}_{S,N}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}^{*}_{t,M} - \frac{1}{S} \sum_{i=1}^{S} \mathbf{g}_{i,N}(\boldsymbol{\theta}), \tag{17}$$

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

$$\mathbf{g}^{*}_{t,M} = \begin{pmatrix} RM_{t,M} \\ \left(RM_{t,M} - \overline{RM}_{T,M}\right)^2 \\ \left(RM_{t,M} - \overline{RM}_{T,M}\right)\left(RM_{t-1,M} - \overline{RM}_{T,M}\right) \\ \vdots \\ \left(RM_{t,M} - \overline{RM}_{T,M}\right)\left(RM_{t-k,M} - \overline{RM}_{T,M}\right) \end{pmatrix}, \tag{18}$$

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

$$\frac{1}{S} \sum_{i=1}^{S} \mathbf{g}_{i,N}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{S} \sum_{i=1}^{S} IV_{i,1,N}(\boldsymbol{\theta}) \\ \frac{1}{S} \sum_{i=1}^{S} \left(IV_{i,1,N}(\boldsymbol{\theta}) - \overline{IV}_{S,N}(\boldsymbol{\theta})\right)^2 \\ \frac{1}{S} \sum_{i=1}^{S} \left(IV_{i,1,N}(\boldsymbol{\theta}) - \overline{IV}_{S,N}(\boldsymbol{\theta})\right)\left(IV_{i,2,N}(\boldsymbol{\theta}) - \overline{IV}_{S,N}(\boldsymbol{\theta})\right) \\ \vdots \\ \frac{1}{S} \sum_{i=1}^{S} \left(IV_{i,1,N}(\boldsymbol{\theta}) - \overline{IV}_{S,N}(\boldsymbol{\theta})\right)\left(IV_{i,k+1,N}(\boldsymbol{\theta}) - \overline{IV}_{S,N}(\boldsymbol{\theta})\right) \end{pmatrix}. \tag{19}$$

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

$$\widehat{\boldsymbol{\theta}}_{T,S,M,N} = \arg\min_{\boldsymbol{\theta} \in \Theta}(\overline{\mathbf{g}}^{*}_{T,M} - \overline{\mathbf{g}}_{S,N}(\boldsymbol{\theta}))' \mathbf{W}^{-1}_{T,M}(\overline{\mathbf{g}}^{*}_{T,M} - \overline{\mathbf{g}}_{S,N}(\boldsymbol{\theta})), \tag{20}$$

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

$$\mathbf{W}_{T,M} = \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{g}^{*}_{t,M} - \overline{\mathbf{g}}^{*}_{T,M}\right)\left(\mathbf{g}^{*}_{t,M} - \overline{\mathbf{g}}^{*}_{T,M}\right)' \tag{21}$$

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

Also, define

$$\boldsymbol{\theta}^{*} = \arg\min_{\boldsymbol{\theta} \in \Theta}(\overline{\mathbf{g}}^{*}_{\infty} - \overline{\mathbf{g}}_{\infty}(\boldsymbol{\theta}))' \mathbf{W}^{-1}_{\infty}(\overline{\mathbf{g}}^{*}_{\infty} - \overline{\mathbf{g}}_{\infty}(\boldsymbol{\theta})), \tag{22}$$

where $\overline{\mathbf{g}}_\infty^*$, $\overline{\mathbf{g}}_\infty(\boldsymbol{\theta})$ and $\mathbf{W}_\infty^{-1}$ 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 $\mathbf{W}_{T,M}^{-1}$, respectively.

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

$$
\begin{aligned}
&\widehat{f}_{\Upsilon, IV_{2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M}) \\
&= \frac{\frac{1}{\Upsilon\varsigma_{2,\Upsilon}^2} \sum_{i=1}^{\Upsilon} \mathbf{K}\left(\frac{IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-x}{\varsigma_{2,\Upsilon}}, \frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\varsigma_{2,\Upsilon}}\right)}{\frac{1}{\Upsilon\varsigma_{1,\Upsilon}} \sum_{i=1}^{\Upsilon} K\left(\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\varsigma_{1,\Upsilon}}\right)} \\
&= \frac{\widehat{f}_{\Upsilon, IV_{2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N}),IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x, RM_{T,M})}{\widehat{f}_{\Upsilon, IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(RM_{T,M})}
\end{aligned}
\tag{23}
$$

Note that the S and Υ denote the number of simulation draws used in for parameter estimation and for density estimation respectively, with $\Upsilon < S$. As outlined in Proposition 2 below, we require that S grow at least as fast as T and that Υ grow slower than T.

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

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

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

$\quad$ $|\sigma(h_{r,N}(\boldsymbol{\theta}_1), \boldsymbol{\theta}_1) - \sigma(h_{r,N}(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2)| \leq K_{2,r,N} \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$, where $h_{r,N}(\boldsymbol{\theta}) = h_{\lfloor \frac{Nrh}{k+1} \rfloor}(\boldsymbol{\theta})$ 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) $|\mu(x, \boldsymbol{\theta}) - \mu(y, \boldsymbol{\theta})| \leq C_1 \|x - y\|$, $\quad |\sigma(x, \boldsymbol{\theta}) - \sigma(y, \boldsymbol{\theta})| \leq C_2 \|x - y\|$,

$\quad$ 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, where $\sigma(\cdot)$ and $\psi(\cdot)$ are defined in (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:

13

(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

$$\mathbf{g}_S(\boldsymbol{\theta}) = \frac{1}{S}\sum_{i=1}^{S}\mathbf{g}_i(\boldsymbol{\theta}),\tag{24}$$

where

$$\overline{\mathbf{g}}_S(\boldsymbol{\theta}) = \frac{1}{S}\sum_{i=1}^{S}\mathbf{g}_i(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{S}\sum_{i=1}^{S} IV_{i,1}(\boldsymbol{\theta}) \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right)^2 \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right)\left(IV_{i,2}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right) \\ \vdots \\ \frac{1}{S}\sum_{i=1}^{S}\left(IV_{i,1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right)\left(IV_{i,k+1}(\boldsymbol{\theta}) - \overline{IV}_S(\boldsymbol{\theta})\right) \end{pmatrix},\tag{25}$$

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

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

(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 $M, T, S, N, \Upsilon \to \infty$, $T/b_M^2 \to 0$, $T/N^{(1-\varkappa)} \to 0$, for $\varkappa > 0$, $S/T \to \pi \geq 1$, $T^2/S \to \infty$, $p_T \to \infty$ and $p_T/T^{1/4} \to 0$, then, for all $RM_{T,M}$ such that $\widehat{f}_{\Upsilon, IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(RM_{T,M}) > d_{\Upsilon}$, with $d_{\Upsilon} = o(1)$, uniformly in $x \in \mathbb{R}_+$,*

$$\left|\widehat{f}_{T,\Upsilon, IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right|$$
$$= O_P(T^{-1/2}\varsigma_{2,\Upsilon}^{-3}d_{\Upsilon}^{-1}) + O_P(T^{-1/2}\varsigma_{1,\Upsilon}^{-2}d_{\Upsilon}^{-2}) + O_P(\Upsilon^{-1/2}\varsigma_{2,\Upsilon}^{-2}d_{\Upsilon}^{-1})$$
$$+ O_P(\Upsilon^{-1/2}\varsigma_{1,\Upsilon}^{-1/2}d_{\Upsilon}^{-2}) + O\left(\varsigma_{2,\Upsilon}^2 d_{\Upsilon}^{-1}\right) + O\left(\varsigma_{1,\Upsilon}^2 d_{\Upsilon}^{-2}\right).$$

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.

Proposition 2. *Let the Assumptions of Theorem 2 hold. Also, let $d_{\Upsilon} = c\Upsilon^{-\varpi}$, $\varsigma_{1,\Upsilon} = c\Upsilon^{-\varphi_1}$, $\varsigma_{2,\Upsilon} = c\Upsilon^{-\varphi_2}$, $\varpi, \varphi_1, \varphi_2 > 0$, for a generic constant $c > 0$, be defined as in Proposition 1.*
Then, for all $RM_{T,M}$ such that $\widehat{f}_{\Upsilon, IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(RM_{T,M}) > d_{\Upsilon}$, uniformly in $x \in \mathbb{R}_+$,
(a) If (i) $\varphi_2 < \varphi_1 < 3/2\varphi_2$, and (ii) $\delta < \min\{3\varphi_2 - 2\varphi_1, 2(\varphi_1 - \varphi_2)\}$,

$$\left|\widehat{f}_{T,\Upsilon, IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M})\right|$$

$$= O_P(T^{-1/2}\varsigma_{2,\Upsilon}^{-3}d_\Upsilon^{-1}) + O_P(\Upsilon^{-1/2}\varsigma_{2,\Upsilon}^{-2}d_\Upsilon^{-1}) + O\left(\varsigma_{2,\Upsilon}^2 d_\Upsilon^{-1}\right). \qquad (26)$$

(b) If in addition to (i)-(ii) above, (iii) $\varsigma_{2,\Upsilon} = c\Upsilon^{-1/8}$ and (iv) $\Upsilon = cT^{4/5}$, for some constant c, all the terms on the RHS of (26) are of the same order and

$$\left| \widehat{f}_{T,\Upsilon,IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})} (x|RM_{T,M}) - f_{IV_{T+1}|IV_T}(x|RM_{T,M}) \right| = O_P(T^{-1/5}d_\Upsilon^{-1}).$$

Note that the first term on the RHS of (26) reflects the contribution of parameter estimation error, while the second and third term reflect the variance and bias term, respectively. Also, note that the "pseudo" optimal uniform rate in part (ii) of Proposition 2 requires that Υ grows at a slower rate than T. It may seem a little bit surprising that increasing the number of simulation used in the construction of the estimator could be harmful. This is due to the fact that the first term on the RHS of (26) decreases with T and increases with Υ.

Finally, it is worthwhile to point out that the statements in the theorem and proposition above hold only when the (instantaneous) volatility process has been simulated using the correct model; in fact, if volatility were simulated from the "wrong" model, then the bias components would be bounded away from zero.

By comparing the statements in Propositions 1 and 2, it appears that there is a gain in using the simulation based estimator whenever T is larger than b_M, and provided that data are simulated from the correct volatility model.

5 Applications to Specific Volatility Realized Measures

Assumption A1 states some primitive conditions on the measurement error between integrated volatility and a generic realized volatility measure. Basically, A1 requires that the first two moments of the measurement error are of order b_M^{-1}. We now specialize the conditions on the measurement error to the different estimators of integrated volatility considered in the paper.

Realized volatility has been suggested as an estimator of integrated volatility by Barndorff-Nielsen and Shephard (2002a) 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, Graversen, Jacod, Podolskij and Shephard (2006), who have shown that, as $M \longrightarrow \infty$,

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

for given $\overline{T}$.

Proposition 3. *Let $dz_t = 0$, a.s. and $\nu = 0$, where dz_t and ν are defined in (1) and in (4), 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 (2004, 2006). Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006) show that, as $M \longrightarrow \infty$,

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

Proposition 4. *Let $\nu = 0$, where ν is defined in (4). 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, it is immediate to see that bipower variation is a less efficient estimator than realized volatility.

As for the microstructure noise estimator suggested by Zhang, Mykland, and Aït-Sahalia (2005), 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 (2006), Aït-Sahalia, Mykland and Zhang (2006), and Barndorff-Nielsen, Hansen, Lunde and Shephard (2006a), 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 $M/\sqrt{e} \to \pi$, $0 < \pi < \infty$. Then Assumption A1 holds with $RM_{t,M} = \widetilde{RV}_{t,e,M}$ for $b_M = M^{1/2}$.*

The data generating process in (1) and (2) allows for jumps in the return process but not in the volatility process. Using option prices data, evidence of jumps in volatility has been supported by e.g. Eraker (2004) and Broadie, Chernov and Johannes (2006). Recently, Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2006) provided a central limit theorem for the measurement error of various realized measures, allowing for the presence of a finite number of jumps in the volatility process. While the statements in the theorems and propositions above are based on a data generating process that only allows for jumps in the return process, we nevertheless conjecture that under additional assumptions, the same (uniform) rate of convergence may hold also in the presence of jumps in volatility, provided that the integrated volatility process "retains" stationarity and the number of jumps is finite over a finite time span. This is left for future research.

6 Experimental Results: Predictive Density Estimator Accuracy

Our primary objective in this section is to assess the accuracy of the realized measured (RM) type predictive density estimator outlined in Section 3. Our secondary objective is to assess the accuracy of two alternative estimators - the simulation type estimator of Section 4 and a simple alternative

which we call the AR type estimator. The latter estimator is based on the findings of Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2002b) that the log of realized volatility is close to normally distributed. Along these lines, we consider a simple AR type estimator where an autoregressive process of order one is fitted to each realized measure discussed in Section 5, and the errors of this regression are assumed to be normally distributed.

Recall that our realized measure type estimator defined in Section 3 is:

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

Our simulation type estimator simply replaces RM data used in the formation of $\widehat{f}_{RM_{T+1,M}|RM_{T,M}}(x|RM_{T,M})$ with simulated integrated volatility data. Namely, in Section 4 we define:

$$\widehat{f}_{\Upsilon,IV_{2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M})$$

$$= \frac{\frac{1}{\Upsilon\varsigma_{2,\Upsilon}^2}\sum_{i=1}^{\Upsilon}\mathbf{K}\left(\frac{IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-x}{\varsigma_{2,\Upsilon}},\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\varsigma_{2,\Upsilon}}\right)}{\frac{1}{\Upsilon\varsigma_{1,\Upsilon}}\sum_{i=1}^{\Upsilon}K\left(\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\varsigma_{1,\Upsilon}}\right)}.$$

Note that parameters need to be estimated prior to construction of $\widehat{f}_{\Upsilon,IV_{2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M})$. Furthermore, we assume knowledge of the correct functional form of the model used for simulating volatility. As such knowledge is clearly not available in general, our experimental results based upon the simulation type estimator should be interpreted with caution. The simulation type estimator will be biased in empirical applications whenever the functional form of the volatility model is misspecified.

Our experiment is carried out as follows. Using the notation defined in Section 4, we begin by simulating S paths of the quantity $h_{i,j\frac{k+1}{N}}$, as defined in (15), 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:

$$\mu(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta}) = \kappa\left(\alpha + 1 - h_{i,(j-1)\frac{k+1}{N}}\right),$$

$$\sigma(h_{i,(j-1)\frac{k+1}{N}}(\boldsymbol{\theta}),\boldsymbol{\theta}) = \sqrt{2\kappa}\sqrt{h_{i,(j-1)\frac{k+1}{N}}},$$

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

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,

17

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:

$$IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right) = \frac{1}{N/(k+1)} \sum_{j=1}^{N/(k+1)} \sigma^2_{j\frac{k+1}{N}}(\boldsymbol{\theta}^\dagger), \tag{27}$$

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

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

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

$$\widetilde{f}_{IV_{3,N}(\boldsymbol{\theta}^\dagger)|IV_{2,N}(\boldsymbol{\theta}^\dagger)}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right)\right) = \frac{\frac{1}{S\varsigma_{2,S}^2}\sum_{i=1}^{S}\mathbf{K}\left(\frac{IV_{i,3,N}(\boldsymbol{\theta}^\dagger)-x}{\varsigma_{2,S}}, \frac{IV_{i,2,N}(\boldsymbol{\theta}^\dagger)-IV_{1,N}(\boldsymbol{\theta}^\dagger)}{\varsigma_{2,S}}\right)}{\frac{1}{S\varsigma_{1,S}}\sum_{i=1}^{S}K\left(\frac{IV_{i,2,N}(\boldsymbol{\theta}^\dagger)-IV_{1,N}(\boldsymbol{\theta}^\dagger)}{\varsigma_{1,S}}\right)}. \tag{28}$$

Note that, for $\varsigma_{2,S} = O(S^{-1/8})$,

$$\left|\widetilde{f}_{IV_{3,N}(\boldsymbol{\theta}^\dagger)|IV_{2,N}(\boldsymbol{\theta}^\dagger)}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right)\right) - f_{IV_{3,N}(\boldsymbol{\theta}^\dagger)|IV_{2,N}(\boldsymbol{\theta}^\dagger)}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^\dagger\right)\right)\right|$$
$$= \ O_P\left(\frac{1}{S^{1/4}}\right) + O_P\left(\frac{1}{N^{1/2-\delta/2}}\right), \ \text{for any } \delta > 0, \tag{29}$$

where the error on the right hand side above can be made arbitrarily small by choosing S and N sufficiently large. Note also that the first term on the right hand side of (29) can be derived in a straightforward manner using 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).[8]

In the experiment, we compare the pseudo true estimator in (28), with our three estimators discussed above. In order to facilitate the comparison, we begin 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 for simplicity we fix $\rho = 0$.

---

[8]We use Silverman (1986, equation (3.31) for univariate density estimator bandwidth selection, and the modification of Silverman (1986, equation (3.28)) due to Scott (1992, pp. 152) for multivariate density estimator bandwidth selection.

6.1 Simulation Procedure

Given the framework just outlined, we begin by generating an X_t path 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. For the RM type estimator, we then construct:

$$\widehat{f}_{RM_{T+1,M}|RM_{T,M}}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right) = \frac{\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{RM_{t+1,M}-x}{\xi_{2,T}},\frac{RM_{t,M}-IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)}{\xi_{2,T}}\right)}{\frac{1}{T\xi_{1,T}}\sum_{t=1}^{T-1}K\left(\frac{RM_{t,M}-IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)}{\xi_{1,T}}\right)}, \quad (30)$$

where $IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)$ is the quantity computed in (27); and where the kernel that we use in our experiment is the product of two Epanechnikov kernels defined as:

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

for the bivariate case, and is an Epanechnikov kernel for the univariate case. In addition, we use Silverman (1986, equation (3.31) for univariate density estimator bandwidth selection, and the modification of Silverman (1986, equation (3.28)) due to Scott (1992, pp. 152) for multivariate density estimator bandwidth selection. For the simulation type estimator, we use the X_t data to estimate the parameters of the instantaneous volatility model given above, and then use the estimated parameters to simulate Υ IV paths of length three ($\Upsilon=500$ in our experiment), where the last two observations in each path are in turn used to form:

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

Finally, for the AR type estimator, we form predictions based on the assumption that

$$(RM_{T+1,M}|RM_{T,M}) \sim N(\widehat{\theta}_{0,T,M}+\widehat{\theta}_{1,T,M}RM_{T,M},\widehat{\sigma}_{T,M}^2),$$

where $\widehat{\theta}_{0,T,M}$ and $\widehat{\theta}_{1,T,M}$ are autoregressive parameters from an AR(1) fitted to the RM data, and where $\widehat{\sigma}_{T,M}^2$ is the associated estimate of the error variance. In order to compare this estimator with our other estimators, we simply use $\widehat{\theta}_{0,T,M}+\widehat{\theta}_{1,T,M}IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)$ in place of $\widehat{\theta}_{0,T,M}+\widehat{\theta}_{1,T,M}RM_{T,M}$. (Note that while logged RM data has been reported by the authors mentioned above to be approximately normally distributed, we report results for both levels and logged data, in order to shed

further light on the advantages associated with using logged data when constructing this estimator). We repeat this simulation procedure Ξ times (i.e. we generate Ξ paths of length T for X_t and carry out Ξ Monte Carlo repetitions).

6.2 Simulation Results

In order to illustrate our approach to accuracy assessment, consider the RM type estimator, defined as $\widehat{f}_{\{RM_{T+1,M}|RM_{T,M}\},p}\left(x|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right)$, for $p = 1, ..., \Xi$. We measure the degree of accuracy of this conditional density estimator by comparing it with the pseudo true estimator as follows. For $M = \{48, 144, 360, 720\}$ and $T = \{100, 300, 500\}$, at each replication, we construct the integrated mean square error, and then we average over the number of replications. That is, we construct:

$$\frac{1}{\Xi}\sum_{p=1}^{\Xi}\left(\frac{1}{\Lambda}\sum_{j=1}^{\Lambda}\left(\widehat{f}_{\{RM_{T+1,M}|RM_{T,M}\},p}\left(x_j|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right) - \widetilde{f}_{IV_{3,N}(\boldsymbol{\theta}^{\dagger})|IV_{2,N}(\boldsymbol{\theta}^{\dagger})}\left(x_j|IV_{1,N}\left(\boldsymbol{\theta}^{\dagger}\right)\right)\right)^2\right),$$
(31)

where Λ is the number of points in the range of the data across which to evaluate the density. We set Λ equal to 100 equally spaced values across the interval $[\overline{IV} - \widehat{SE}(IV), \overline{IV} + \widehat{SE}(IV)]$, where $\overline{IV}$ and $\widehat{SE}(IV)$ are constructed from a large dataset of simulated daily IV values using the true model. Also, in order to evaluate the variability across Monte Carlo replications, we also consider the standard error of the integrated mean square error across Monte Carlo replications, that is:

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

The same approach as that outlined above is also used to assess the accuracy of the simulation type and the AR type estimators. All results are based upon 1000 Monte Carlo replications.

The results of this experiment are gathered in four tables. Table 1 (Case I: levels data; no microstructure noise or jumps) and Table 2 (Case II: logged data; no microstructure noise or jumps) report findings for all three types of estimator, whereas Table 3 (Case III: microstructure noise) and Table 4 (Case IV: jumps) contain results only for the RM type estimator. 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 $\alpha = \{1, 3, 5\}$. In all tables, results are given for the 4 realized measures discussed above; namely for $RV_{t,M}$, $BV_{t,M}$, $\widehat{RV}_{t,l,M}$ and $\widetilde{RV}_{t,e,M}$. Recall that in the case of the RM type estimator, these four measures are used directly in the construction of the predictive density. However, in the case of the simulation type estimator, the 4 measures enter into our calculations only inasmuch as they are used to construct the historical data-based moments used in parameter estimation. Entries in the table are the MSEs defined in equation (31), and associated standard errors appear in parentheses.

A number of clear conclusions emerge upon examination of the tables. From Table 1, where results for Case I are reported, note that results based on the realized measure type estimator indicate that $RV_{t,M}$ 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 by approximately 20-40% as the number of intraday observations increases from 48 to 720; and there is a reduction of approximately 40-50% as the number of daily observations is increased from 100 to 300. As might be expected, given that we assume that the "correct" model is known, the simulation type estimator yields lower MSEs than the RM type estimator. Furthermore, again as expected, the simulation type estimator improves very little for fixed M, as T increases. Finally, notice that the AR type estimator yields inferior predictive density estimates, as they are worse by a factor of approximately 3, when compared with the RM type estimator. This, of course, is not surprising, given that there is no evidence that levels RM data are approximately normally distributed. The more interesting comparison of the AR type estimator is contained in Table 2, where results for logged data are presented.

Turning now to Table 2, three observations are worth making. First, notice that the results of our experiments are confounded slightly when data are logged. For example, the clear ordinal ranking of the MSEs as one increases M for fixed T, or as one increases T for fixed M, no longer holds when comparing the point MSEs. Second, note however that there is a clear improvement in MSE relative to the levels case under both the RM type estimator and the AR type estimator, supporting the finding of Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2002b) that the log of realized volatility is close to normally distributed. Finally, the AR type estimator does not outperform our RM type estimator, although it is relatively better, as it is now worse by a factor of about 2 rather than 3.

We now turn our discussion to microstructure noise and jumps, which are evaluated only for our RM type estimator, given the findings of Tables 1 and 2. In Table 3 (Case III), $\widehat{RV}_{t,l,M}$ 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.[9] For example, consider Panel A in Table 2. The MSE for

---

[9]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

$RV_{t,M}$ 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. Examination of Table 4 (Case IV) 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 and Methodology

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 have been 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 1st to May 20th 1997, for a total of 100 trading days.

In this exercise, we predict the conditional distribution of daily integrated volatility for Wednesday May 21st, 1997. In order to have at least a visual grasp of the effect of the conditioning set, we consider 4 cases: 1, 4, 7, and 10 day ahead prediction. Along these lines, note that the formula given in (11) can be readily generalized to allow for $(\Delta + 1)$ day ahead prediction, as follows:

$$
\widehat{f}_{RM_{T+1,M}|RM_{T-\Delta,M}}(x|RM_{T-\Delta,M}) \tag{32}
$$
$$
= \frac{\frac{1}{(T-\Delta-1)\xi_{2,T}^2} \sum_{t=1}^{T-\Delta-1} \mathbf{K}\left(\frac{RM_{t+\Delta+1,M}-x}{\xi_{2,T}}, \frac{RM_{t,M}-RM_{T-\Delta,M}}{\xi_{2,T}}\right)}{\frac{1}{(T-\Delta-1)\xi_{1,T}} \sum_{t=1}^{T-\Delta-1} K\left(\frac{RM_{t,M}-RM_{T-\Delta,M}}{\xi_{1,T}}\right)}.
$$

---

the non-robust measures exhibit increasingly dramatic performance deterioration as M increases.

In other words, we calculate (32) with $\Delta = 0, 3, 6, 9$. In our calculations, we tried Epanechnikov as well as quartic kernel functions with bandwidth parameters chosen as discussed in Section 6.1. Results are reported for estimators constructed using Epanechnikov kernels.

From the original data set, which includes prices recorded for every trade, we extracted 10 second and 5 minute interval data. Provided that there is sufficient liquidity in the market, the 5 minute frequency is generally accepted as the highest frequency at which the effect of microstructure biases are not too distorting (see Andersen, Bollerslev, Diebold and Labys, 2001 and Andersen, Bollerslev and Lang, 1999). Hence our choice of these two frequencies allows us to evaluate the effect of microstructure noise on the estimated densities.

The price figures for each 10 second and 5 minute 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, 2006); realized volatility constructed from linearly interpolated returns converges in probability to zero as the time interval shrinks to zero.

From the calculated series we have obtained 10 second and 5 minute 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 2340 (resp. 78) intraday returns calculated over an interval of ten seconds (resp. five minutes).[10]

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 and for different values of Δ. Results are shown in Figures 1 to 8 (for the raw measures) and in Figures 9 to 16 (for the logged ones), using 1000 evaluation points for the conditional densities.

---

[10]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 graphs yield some interesting findings. To begin, note that the graphs for the realized volatility and bipower variation are quite similar. This would suggest that jumps occur occasionally in the price process, or in any case that they do not affect a procedure which is based on sample sizes containing a large number of daily observations.

On the other hand, microstructure noise seems to have a 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 theory (see Aït-Sahalia, Mykland and Zhang, 2006), and confirmed by simulation results, when the time interval between successive observations becomes small, then 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 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. 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. These results thus provide further evidence in favor of a warning 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.

The effect of the conditioning set is also evident in the densities. In general, one day ahead predictive densities display more probability mass at the center of the domain, and less weight in the tails.

Finally, the effect of the logarithmic transformation supports previous findings by Andersen, Bollerslev, Diebold and Labys (2003) and Barndorff-Nielsen and Shephard (2002b). The shapes of the densities seem more symmetric and they look closer to the normal than those based on raw realized measures. Additionally, the distortionary effect of microstructure noise is even more apparent. Consider for example, Figures 13 and 14, where there is an apparent absence of mass in the left tail of the density.

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.

Appendix

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

$$
\widehat{f}_{RM_{T+1}|RM_{T,M}}(x|RM_{T,M})
$$

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

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

The proof of Theorem 1 requires the following Lemma.

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

$$
\widehat{f}_{RM_{T+1,M},RM_{T,M}}(x, RM_{T,M}) - f_{IV_{t+1},IV_t}(x, RM_{T,M}) = O_P\left(b_M^{-1/2}\xi_{2,T}^{-3}\right) + O_P\left(T^{-1/2}\xi_{2,T}^{-2}\right) + O\left(\xi_{2,T}^2\right),
$$
(33)

uniformly in $x \in \mathbb{R}_+$, and

$$
\widehat{f}_{RM_{t,M}}(RM_{T,M}) - f_{IV_t}(RM_{T,M}) = O_P\left(b_M^{-1/2}\xi_{1,T}^{-2}\right) + O_P\left(T^{-1/2}\xi_{1,T}^{-1/2}\right) + O\left(\xi_{1,T}^2\right).
$$
(34)

Proof of Lemma 1

We show the result in (33). The result in (34) follows by the same argument. Given A4(i), by a mean value expansion,

$$
\frac{1}{T\xi_{2,T}^2} \sum_{t=1}^{T-1} \mathbf{K}\left(\frac{\widehat{U}_{t+1,M}-u_{T,M}}{\xi_{2,T}}\right) - \frac{1}{T\xi_{2,T}^2} \sum_{t=1}^{T-1} \mathbf{K}\left(\frac{U_{t+1}-u_{T,M}}{\xi_{2,T}}\right)
$$

$$
= \frac{1}{T\xi_{2,T}^3} \sum_{t=1}^{T-1} \mathbf{K}_1^{(1)}\left(\frac{\overline{U}_{t+1,M}-u_{T,M}}{\xi_{2,T}}\right) N_{t+1,M} + \frac{1}{T\xi_{2,T}^3} \sum_{t=1}^{T-1} \mathbf{K}_2^{(1)}\left(\frac{\overline{U}_{t+1,M}-u_{T,M}}{\xi_{2,T}}\right) N_{t,M}, \quad (35)
$$

where $\overline{U}_{t+1,M} \in \left(\widehat{U}_{t+1,M}, U_{t+1}\right)$. The second term on the RHS of (35) is majorized by

$$
\sup_{t,x} \mathbf{K}_2^{(1)}\left(\frac{\overline{U}_{t+1,M}-u_{T,M}}{\xi_{2,T}}\right) \frac{1}{T\xi_{2,T}^3} \sum_{t=1}^{T-1} |N_{t,M}|
$$

$$
= O_P\left(\xi_{2,T}^{-3}\right)\left(\frac{1}{T}\sum_{t=1}^{T-1} \mathrm{E}\,|N_{t,M}| + \frac{1}{T}\sum_{t=1}^{T-1}\left(|N_{t,M}| - \mathrm{E}\,|N_{t,M}|\right)\right).
$$

Given A1, $\mathrm{E}\left(N_{t,M}^2\right) = O(b_M^{-1})$. By Jensen inequality,

$$
\mathrm{E}\left(|N_{t,M}|\right) = \mathrm{E}\left(\sqrt{N_{t.M}^2}\right) \le \left(\mathrm{E}\left(N_{t,M}^2\right)\right)^{1/2} = O(b_M^{-1}).
$$

26

Let $\overline{|N_{t,M}|} = |N_{t,M}| - \mathrm{E}\,|N_{t,M}|$. Now,

$$
\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T-1}\overline{|N_{t,M}|}\right) = \frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\mathrm{E}\left(\overline{|N_{t,M}||N_{s,M}|}\right)
$$

$$
\leq \frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\left(\mathrm{E}\left(\overline{|N_{t,M}|}^2\right)\right)^{1/2}\left(\mathrm{E}\left(\overline{|N_{s,M}|}^2\right)\right)^{1/2},
$$

and

$$
\mathrm{E}\left(\overline{|N_{t,M}|}^2\right) = \mathrm{E}\left(N_{t,M}^2\right) - (\mathrm{E}\left(|N_{t,M}|\right))^2 = O(b_M^{-1}), \text{ uniformly in } t.
$$

Thus, uniformly in $x \in \mathbb{R}_+$,

$$
\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{\widehat{U}_{t+1,M}-u_{T,M}}{\xi_{2,T}}\right) - \frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{U_{t+1}-u_{T,M}}{\xi_{2,T}}\right) = O_P\left(\xi_{2,T}^{-3}b_M^{-1/2}\right).
$$

Finally, uniformly in $x \in \mathbb{R}_+$,

$$
\frac{1}{T\xi_{2,T}^2}\sum_{t=1}^{T-1}\mathbf{K}\left(\frac{U_{t+1}-u_{T,M}}{\xi_{2,T}}\right) - f_{IV_{t+1},IV_t}(x, RM_{T,M})
$$

$$
= \frac{1}{T}\sum_{t=1}^{T-1}\frac{1}{\xi_{2,T}^2}\left(\mathbf{K}\left(\frac{U_{t+1}-u_{T,M}}{\xi_{2,T}}\right) - \mathrm{E}\left(\mathbf{K}\left(\frac{U_{t+1}-u_{T,M}}{\xi_{2,T}}\right)\right)\right)
$$

$$
+ \left(\frac{1}{\xi_{2,T}^2}\mathrm{E}\left(\mathbf{K}\left(\frac{U_{t+1}-u_{T,M}}{\xi_{2,T}}\right)\right) - f_{IV_{t+1},IV_t}(x, RM_{T,M})\right)
$$

$$
= O_P\left(T^{-1/2}\xi_{2,T}^{-2}\right) + O\left(\xi_{2,T}^2\right),
$$

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

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

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

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

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

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),

$$\frac{\xi_{1,T}^2 d_T^{-2}}{\xi_{2,T}^2 d_T^{-1}} \to 0, \quad \frac{\xi_{1,T}^{-1} d_T^{-2}}{\xi_{2,T}^{-2} d_T^{-1}} \to 0, \quad \frac{\xi_{1,T}^{-2} d_T^{-2}}{\xi_{2,T}^{-3} d_T^{-1}} \to 0. \tag{36}$$

Now, if $\delta < 2(\phi_1 - \phi_2)$, 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 (36), it suffices to check that

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

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 $\delta < 2(\phi_1 - \phi_2)$ 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 ϕ_2 and ψ, 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 ψ, 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 $T^{-1/4}$. ∎

Before proving Theorem 2, we need to introduce some convenient notation. 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

$$\widehat{f}_{\Upsilon,IV_{2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})|IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x|RM_{T,M})$$

$$= \frac{\widehat{f}_{\Upsilon,IV_{2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N}),IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x, RM_{T,M})}{\widehat{f}_{\Upsilon,IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(RM_{T,M})} \tag{38}$$

$$= \frac{\frac{1}{\Upsilon \varsigma_{2,\Upsilon}^2} \sum_{i=1}^{\Upsilon} \mathbf{K}\left(\frac{IV_{i,2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-x}{\varsigma_{2,\Upsilon}}, \frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\varsigma_{2,\Upsilon}}\right)}{\frac{1}{\Upsilon \varsigma_{1,\Upsilon}} \sum_{i=1}^{\Upsilon} K\left(\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\varsigma_{1,\Upsilon}}\right)}$$

$$= \frac{\frac{1}{\Upsilon \varsigma_{2,\Upsilon}^2} \sum_{i=1}^{\Upsilon} \mathbf{K}\left(\frac{U_{i,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-u_{T,M}}{\varsigma_{2,T}}\right)}{\frac{1}{\Upsilon \varsigma_{1,\Upsilon}} \sum_{i=1}^{\Upsilon} K\left(\frac{IV_{i,1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})-RM_{T,M}}{\varsigma_{1,T}}\right)}. \tag{39}$$

Also, let

$$\widehat{f}_{\Upsilon,IV_{2,N}(\boldsymbol{\theta}^{\dagger})|IV_{1,N}(\boldsymbol{\theta}^{\dagger})}(x, RM_{T,M}) = \frac{\frac{1}{\Upsilon \varsigma_{2,\Upsilon}^2} \sum_{i=1}^{\Upsilon} \mathbf{K}\left(\frac{U_{i,N}(\boldsymbol{\theta}^{\dagger})-u_{T,M}}{\varsigma_{2,T}}\right)}{\frac{1}{\Upsilon \varsigma_{1,\Upsilon}} \sum_{i=1}^{\Upsilon} K\left(\frac{IV_{i,1,N}(\boldsymbol{\theta}^{\dagger})-RM_{T,M}}{\varsigma_{1,T}}\right)} \tag{40}$$

and

$$\widehat{f}_{\Upsilon,IV_{2}(\boldsymbol{\theta}^{\dagger})|IV_{1}(\boldsymbol{\theta}^{\dagger})}(x, RM_{T,M}) = \frac{\frac{1}{\Upsilon \varsigma_{2,\Upsilon}^2} \sum_{i=1}^{\Upsilon} \mathbf{K}\left(\frac{U_{i}(\boldsymbol{\theta}^{\dagger})-u_{T,M}}{\varsigma_{2,T}}\right)}{\frac{1}{\Upsilon \varsigma_{1,\Upsilon}} \sum_{i=1}^{\Upsilon} K\left(\frac{IV_{i,1}(\boldsymbol{\theta}^{\dagger})-RM_{T,M}}{\varsigma_{1,T}}\right)}. \tag{41}$$

The proof of Theorem 2 requires the following Lemma.

Lemma 2. *Let A1-A7 hold. If* $T/b_M^2 \to 0$, $T/N^{(1-\varkappa)} \to 0$, *for a positive constant* $\varkappa$, $T/S \to \pi \geq 1$, $T^2/S \to \infty$, $p_T \to \infty$ *and* $p_T/T^{1/4} \to 0$, *then*
(i)

$$\widehat{f}_{\Upsilon,IV_{2,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N}),IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(x, RM_{T,M}) - f_{IV_{2}(\boldsymbol{\theta}^{\dagger}),IV_{1}(\boldsymbol{\theta}^{\dagger})}(x, RM_{T,M})$$

$$= O_P\left(T^{-1/2}\varsigma_{2,\Upsilon}^{-3}\right) + O_P\left(\Upsilon^{-1/2}\varsigma_{2,\Upsilon}^{-2}\right) + O\left(\varsigma_{2,\Upsilon}^2\right),$$

uniformly in $x \in \mathbb{R}_+$;
(ii)

$$\widehat{f}_{\Upsilon,IV_{1,N}(\widehat{\boldsymbol{\theta}}_{T,S,M,N})}(RM_{T,M}) - f_{IV_{1}(\boldsymbol{\theta}^{\dagger})}(RM_{T,M})$$

$$= O_P\left(T^{-1/2}\varsigma_{1,\Upsilon}^{-2}\right) + O_P\left(\Upsilon^{-1/2}\varsigma_{1,\Upsilon}^{-1/2}\right) + O\left(\varsigma_{1,\Upsilon}^2\right).$$

29

Proof of Lemma 2

(i) First note that

$$
\sup_x \left| \widehat{f}_{\Upsilon, IV_{2,N}(\widehat{\theta}_{T,S,M,N}), IV_{1,N}(\widehat{\theta}_{T,S,M,N})}(x, RM_{T,M}) - f_{IV_2(\theta^\dagger), IV_1(\theta^\dagger)}(x, RM_{T,M}) \right|
$$

$$
\leq \sup_x \left| \widehat{f}_{\Upsilon, IV_{2,N}(\widehat{\theta}_{T,S,M,N}), IV_{1,N}(\widehat{\theta}_{T,S,M,N})}(x, RM_{T,M}) - \widehat{f}_{\Upsilon, IV_{2,N}(\theta^\dagger), IV_{1,N}(\theta^\dagger)}(x, RM_{T,M}) \right|
$$

$$
+ \sup_x \left| \widehat{f}_{\Upsilon, IV_{2,N}(\theta^\dagger), IV_{1,N}(\theta^\dagger)}(x, RM_{T,M}) - \widehat{f}_{\Upsilon, IV_2(\theta^\dagger), IV_1(\theta^\dagger)}(x, RM_{T,M}) \right|
$$

$$
+ \sup_x \left| \widehat{f}_{\Upsilon, IV_2(\theta^\dagger), IV_1(\theta^\dagger)}(x, RM_{T,M}) - f_{IV_2(\theta^\dagger), IV_1(\theta^\dagger)}(x, RM_{T,M}) \right|. \tag{42}
$$

We begin by considering the first term on the RHS of (42). Via a mean value expansion around $\theta^\dagger$,

$$
\widehat{f}_{\Upsilon, IV_{2,N}(\widehat{\theta}_{T,S,M,N}), IV_{1,N}(\widehat{\theta}_{T,S,M,N})}(x, RM_{T,M}) - f_{IV_2(\theta^\dagger), IV_1(\theta^\dagger)}(x, RM_{T,M})
$$

$$
= \frac{1}{\Upsilon \varsigma_{2,\Upsilon}^3} \sum_{i=1}^{\Upsilon} \mathbf{K}_1^{(1)} \left(\frac{U_{i,N}\left(\overline{\theta}_{T,S,M,N}\right) - u_{T,M}}{\varsigma_{2,T}} \right) \frac{\partial IV_{i,2,N}(\theta)}{\partial \theta} \Big|_{\theta = \overline{\theta}_{T,S,M,N}} \left(\widehat{\theta}_{T,S,M,N} - \theta^\dagger \right)
$$

$$
+ \frac{1}{\Upsilon \varsigma_{2,\Upsilon}^3} \sum_{i=1}^{\Upsilon} \mathbf{K}_2^{(1)} \left(\frac{U_{i,N}\left(\overline{\theta}_{T,S,M,N}\right) - u_{T,M}}{\varsigma_{2,T}} \right) \frac{\partial IV_{i,1,N}(\theta)}{\partial \theta} \Big|_{\theta = \overline{\theta}_{T,S,M,N}} \left(\widehat{\theta}_{T,S,M,N} - \theta^\dagger \right).
$$

Given A2,A3,A5 and A7, by the uniform law of large numbers,

$$
\frac{1}{\Upsilon} \sum_{i=1}^{\Upsilon} \mathbf{K}_2^{(1)} \left(\frac{U_{i,N}(\theta_{T,S,M,N}) - u_{T,M}}{\varsigma_{2,T}} \right) \frac{\partial IV_{i,1,N}(\theta)}{\partial \theta} = O_P(1),
$$

uniformly in $x \in \mathbb{R}_+$ and in θ. Also, given A1-A3, A5-A7, and under the regularity conditions set out in the statement of the Lemma, then

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

because of Theorem 2 in Corradi and Distaso (2006).[11] Thus, the first term on the RHS of (42) is $O_P\left(T^{-1/2}\varsigma_{2,T}^{-3}\right)$.

The second term on the RHS of (42) captures the discretization error incurred in the simulation of the volatility path. Now,

$$
\widehat{f}_{\Upsilon, IV_{2,N}(\theta^\dagger)|IV_{1,N}(\theta^\dagger)}(x, RM_{T,M}) - \widehat{f}_{\Upsilon, IV_2(\theta^\dagger)|IV_1(\theta^\dagger)}(x, RM_{T,M})
$$

$$
= \frac{1}{\Upsilon \varsigma_{2,\Upsilon}^3} \sum_{i=1}^{\Upsilon} \mathbf{K}_1^{(1)} \left(\frac{\overline{U_{i,N}(\theta^\dagger)} - u_{T,M}}{\varsigma_{2,T}} \right) \left(IV_{i,2,N}\left(\theta^\dagger\right) - IV_{i,2}\left(\theta^\dagger\right) \right)
$$

$$
+ \frac{1}{\Upsilon \varsigma_{2,\Upsilon}^3} \sum_{i=1}^{\Upsilon} \mathbf{K}_2^{(1)} \left(\frac{\overline{U_{i,N}(\theta^\dagger)} - u_{T,M}}{\varsigma_{2,T}} \right) \left(IV_{i,1,N}\left(\theta^\dagger\right) - IV_{i,1}\left(\theta^\dagger\right) \right). \tag{43}
$$

---

[11] Note that for $\sqrt{T}$-consistency it suffices that S grows at least as fast as T.

We just analyze the first term on the RHS of (43), as the second term can the dealt with in the same manner. Note that, given A4(i),

$$\frac{1}{\Upsilon\varsigma_{2,\Upsilon}^3}\sum_{i=1}^{\Upsilon}\mathbf{K}_1^{(1)}\left(\frac{\overline{U_{i,N}}\left(\boldsymbol{\theta}^\dagger\right)-u_{T,M}}{\varsigma_{2,T}}\right)\left(IV_{i,2,N}\left(\boldsymbol{\theta}^\dagger\right)-IV_{i,2}\left(\boldsymbol{\theta}^\dagger\right)\right)$$

$$\leq \sup_{i,x}\left|\mathbf{K}_1^{(1)}\left(\frac{\overline{U_{i,N}}\left(\boldsymbol{\theta}^\dagger\right)-u_{T,M}}{\varsigma_{2,T}}\right)\right|\frac{1}{\varsigma_{2,\Upsilon}^3}\frac{1}{\Upsilon}\sum_{i=1}^{\Upsilon}\left|IV_{i,1,N}\left(\boldsymbol{\theta}^\dagger\right)-IV_{i,1}\left(\boldsymbol{\theta}^\dagger\right)\right|$$

$$= O_P(1)\frac{1}{\varsigma_{2,\Upsilon}^3}\frac{1}{\Upsilon}\sum_{i=1}^{\Upsilon}\left|IV_{i,1,N}\left(\boldsymbol{\theta}^\dagger\right)-IV_{i,1}\left(\boldsymbol{\theta}^\dagger\right)\right|.$$

Now, by Corollary 1.8 in Pardoux and Talay (1985), for $\varkappa > 0$,

$$\frac{1}{\Upsilon}\sum_{i=1}^{\Upsilon}\left|IV_{i,1,N}\left(\boldsymbol{\theta}^\dagger\right)-IV_{i,1}\left(\boldsymbol{\theta}^\dagger\right)\right|$$

$$\leq \sup_{i}\left|IV_{i,N}\left(\boldsymbol{\theta}^\dagger\right)-IV_{i,1}\left(\boldsymbol{\theta}^\dagger\right)\right|=O_P\left(N^{-1/2(1-\varkappa)}\right).$$

Thus,

$$\sup_{i,x}\mathbf{K}_1^{(1)}\left(\frac{\overline{U_{i,N}}\left(\boldsymbol{\theta}^\dagger\right)-u_{T,M}}{\varsigma_{2,T}}\right)\frac{1}{\varsigma_{2,\Upsilon}^3}\frac{1}{\Upsilon}\sum_{i=1}^{\Upsilon}\left|IV_{i,1,N}\left(\boldsymbol{\theta}^\dagger\right)-IV_{i,1}\left(\boldsymbol{\theta}^\dagger\right)\right|$$

$$= O_P\left(\varsigma_{2,\Upsilon}^{-3}\right)\frac{1}{\Upsilon}\sum_{i=1}^{\Upsilon}\left|IV_{i,1,N}\left(\boldsymbol{\theta}^\dagger\right)-IV_{i,1}\left(\boldsymbol{\theta}^\dagger\right)\right|=O_P\left(N^{-1/2(1-\varkappa)}\varsigma_{2,\Upsilon}^{-3}\right).$$

As $T/N^{(1-\varkappa)}\to 0$, the second term on the RHS of (42) is of a smaller order than the first one. Finally, as for the third term on the RHS of (42), from Theorem 1 in Andrews,

$$\sup_{x}\left|\widehat{f}_{\Upsilon,IV_2(\boldsymbol{\theta}^\dagger),IV_1(\boldsymbol{\theta}^\dagger)}(x,RM_{T,M})-f_{IV_2(\boldsymbol{\theta}^\dagger),IV_1(\boldsymbol{\theta}^\dagger)}(x,RM_{T,M})\right|$$

$$= O_P\left(T^{-1/2}\varsigma_{2,T}^{-2}\right)+O\left(\varsigma_{2,T}^2\right).$$

(ii) By the same argument used in the proof of (i). ∎

Proof of Theorem 2

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

Proof of Proposition 2

(a) Given the rate conditions in (i) and (ii), the contribution of the marginal density estimator is negligible, and thus the rate is determined only by the estimator of the joint density. The statement then follow from Lemma 2, part (i).

(b) Given (iii) and (iv), $\varsigma_{2,\Upsilon}=\Upsilon^{-1/8}=T^{-1/10}$, and all the three terms on the RHS of (26) are of the same probability order $T^{-1/5}d_\Upsilon^{-1}$. ∎

Proof of Proposition 3,4,5

By the same argument used in the proof of Propositions 1 and 2 in Corradi and Distaso (2006). ∎

Proof of Propositions 6

From Part (iv) in the proof of Lemma 1 in Corradi, Distaso and Swanson (2006). ∎

References

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

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

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

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

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

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

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

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

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

Andersen, T.G., T. Bollerslev and N. Meddahi (2006). Realized Volatility Forecasting and Market Microstructure Noise. Manuscript, Imperial College.

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 (2006). Testing Market Microstructure Effects with an Application to the Dow Jones Industrial Average. Manuscript, University of Warwick.

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 (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79, 655-692.

Barndorff-Nielsen, O.E. and N. Shephard (2002a). 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 (2002b). Estimating Quadratic Variation using Realised Variance. *Journal of Applied Econometrics*, 17, 457-477.

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

Barndorff-Nielsen, O.E. and N. Shephard (2006). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics*, 4, 1-30.

Barndorff-Nielsen O.E., S.E. Graversen, J. Jacod, M. Podolskij and N. Shephard (2006). A Central Limit Theorem for Realized Power and Bipower Variations of Continuous Semimartingales. In Y. Kabanov, J. Stoyanov and R. Lipster (eds.), *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*. Springer and Verlag, New York, 33-68.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2006a). Designing Realised Kernels to measure ex-post variation of equity prices in the presence of noise. Manuscript, Oxford University.

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

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.

Broadie, M., M. Chernov and M. Johannes (2006). Model Specification and Risk Premia: Evidence from S&P500 Future Option Market. *Journal of Finance,* forthcoming.

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

Corradi, V., W. Distaso and N.R. Swanson (2006). Predictive Inference for Integrated Volatility. Manuscript, University of Warwick.

Eraker, B. (2004). Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance,* 59, 1367-1403.

Fan, J. and I. Gijbels (1997). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

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

Galbraith, J. and V. Zinde Walsh (2006). GARCH model Estimation Using Auxiliary Volatility Information. Manuscript, McGill University.

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

Hansen, P.R. and A. Lunde (2006). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics*, 24, 127-218.

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.

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

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

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley and Sons, New York.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

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

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

Zhang, L. (2004). Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. *Bernoulli*, forthcoming.

Table 1: Predictive Integrated Volatility Density Estimator Accuracy Experiment - Mean Square Error

Case I: Levels Data - No Microstructure Noise or Jumps

	Realized Measure Type Estimator				Simulation Type Estimator				Autoregression Type Estimator			
M	RV	Bipower	SubSample1	SubSample2	RV	Bipower	SubSample1	SubSample2	RV	Bipower	SubSample1	SubSample2
100 Daily Observations Used in Density Construction												
48	0.063380	0.063931	0.080448	0.108170	0.016702	0.017588	0.040949	0.014980	0.199255	0.197918	0.184532	0.187755
	(0.030270)	(0.030687)	(0.050367)	(0.050683)	(0.010950)	(0.011951)	(0.136217)	(0.011968)	(0.007873)	(0.007270)	(0.008266)	(0.005350)
144	0.053374	0.053810	0.052412	0.088510	0.018499	0.017182	0.010973	0.012421	0.200937	0.200402	0.193105	0.192044
	(0.028465)	(0.029127)	(0.030904)	(0.040846)	(0.012088)	(0.011893)	(0.007472)	(0.009059)	(0.007511)	(0.007059)	(0.006536)	(0.006218)
360	0.049495	0.049516	0.040359	0.073545	0.012923	0.011969	0.011678	0.014458	0.202239	0.201897	0.198842	0.195479
	(0.028077)	(0.028305)	(0.026044)	(0.035423)	(0.009616)	(0.007586)	(0.007357)	(0.008203)	(0.008141)	(0.008042)	(0.007600)	(0.006991)
720	0.048862	0.049083	0.034974	0.066255	0.012489	0.012341	0.009589	0.012991	0.202647	0.202511	0.201803	0.197215
	(0.027863)	(0.028135)	(0.023958)	(0.032855)	(0.011230)	(0.010169)	(0.006278)	(0.008128)	(0.008351)	(0.008299)	(0.008538)	(0.007160)
300 Daily Observations Used in Density Construction												
48	0.037318	0.037624	0.050876	0.071175	0.011939	0.011811	0.015378	0.014869	0.196566	0.195318	0.181446	0.186011
	(0.015493)	(0.015969)	(0.018006)	(0.026487)	(0.007375)	(0.006782)	(0.009736)	(0.010518)	(0.004493)	(0.004290)	(0.004200)	(0.003138)
144	0.030571	0.030752	0.027656	0.056256	0.012662	0.012035	0.011177	0.012141	0.198933	0.198419	0.190444	0.190171
	(0.013112)	(0.013048)	(0.013456)	(0.021183)	(0.008183)	(0.007516)	(0.005718)	(0.006891)	(0.004463)	(0.004395)	(0.003459)	(0.003808)
360	0.028750	0.028956	0.019378	0.043908	0.008164	0.007976	0.009148	0.013365	0.199872	0.199717	0.195944	0.193547
	(0.012698)	(0.012544)	(0.010730)	(0.017503)	(0.006030)	(0.005197)	(0.004915)	(0.007739)	(0.004345)	(0.004386)	(0.003731)	(0.004049)
720	0.028281	0.028809	0.010862	0.038592	0.008253	0.008595	0.008859	0.012256	0.200124	0.200061	0.198749	0.195172
	(0.012399)	(0.012230)	(0.009240)	(0.015772)	(0.006373)	(0.006889)	(0.004804)	(0.008333)	(0.004414)	(0.004373)	(0.004059)	(0.004179)
500 Daily Observations Used in Density Construction												
48	0.026397	0.026229	0.042915	0.055208	0.009645	0.010868	0.012856	0.012165	0.196212	0.195012	0.180781	0.185809
	(0.010435)	(0.010240)	(0.018703)	(0.026379)	(0.005879)	(0.006237)	(0.006731)	(0.008106)	(0.003386)	(0.003381)	(0.002774)	(0.002529)
144	0.020435	0.020715	0.020264	0.042832	0.011301	0.010936	0.011026	0.011116	0.198394	0.197944	0.190065	0.190007
	(0.009229)	(0.009731)	(0.007601)	(0.014037)	(0.007058)	(0.006273)	(0.006096)	(0.006517)	(0.003546)	(0.003441)	(0.002755)	(0.003175)
360	0.018822	0.018677	0.011882	0.032450	0.007246	0.006565	0.009166	0.013609	0.199299	0.199135	0.195876	0.193413
	(0.008892)	(0.008613)	(0.006494)	(0.011278)	(0.005233)	(0.003767)	(0.006031)	(0.006336)	(0.003498)	(0.003498)	(0.003084)	(0.003369)
720	0.018485	0.018500	0.008988	0.027769	0.007138	0.007125	0.006688	0.010743	0.199557	0.199428	0.198553	0.194978
	(0.008867)	(0.008818)	(0.005280)	(0.010628)	(0.004742)	(0.005127)	(0.004946)	(0.005515)	(0.003520)	(0.003441)	(0.003422)	(0.003411)

* Notes: Entries report average mean square errors (MSEs) and associated standard errors from 1000 Monte Carlo iterations, where each MSE is constructed based upon comparison of a pseudo true predictive integrated volatility density with estimated densities constructed using four different realized measures (RV, $Bipower$, $SubSample1$, and $SubSample2$) and three density estimators (realized measure type, simulation type and AR type - see above for further details). Density construction is based upon use of M intraday simulated observable state variable observations, for various daily sample of size T. The range of the data across which estimated densities were evaluated was set to 100 equally spaced values from a large simulated sample of integrated volatility (IV) data, for the interval $[\overline{IV} - \widehat{SE}(IV), \overline{IV} + \widehat{SE}(IV)]$. See Section 6 for further details.

Table 2: Predictive Integrated Volatility Density Estimator Accuracy Experiment - Mean Square Error
Case II: Logged Data - No Microstructure Noise or Jumps

M	Realized Measure Type Estimator				Simulation Type Estimator				Autoregression Type Estimator			
	RV	Bipower	SubSample1	SubSample2	RV	Bipower	SubSample1	SubSample2	RV	Bipower	SubSample1	SubSample2
100 Daily Observations Used in Density Construction												
48	0.048392	0.043335	0.062733	0.039921	0.117834	0.141155	0.114618	0.095271	0.097443	0.089888	0.060163	0.053952
	0.029511	0.017966	0.023845	0.019822	0.055365	0.168737	0.131532	0.070783	0.017451	0.016609	0.013888	0.005820
144	0.061156	0.058670	0.027972	0.033538	0.177223	0.141916	0.086466	0.073507	0.112545	0.108467	0.063456	0.065366
	0.030485	0.028158	0.015505	0.016574	0.301966	0.110988	0.046442	0.023308	0.019612	0.019208	0.009936	0.011718
360	0.066711	0.063014	0.039128	0.038519	0.126503	0.120801	0.097811	0.077744	0.118525	0.116894	0.086522	0.080325
	0.049213	0.030008	0.019649	0.016344	0.046559	0.046588	0.050183	0.028516	0.019201	0.018855	0.015856	0.014789
720	0.064295	0.062779	0.048142	0.043377	0.123585	0.132855	0.104779	0.084115	0.119866	0.118786	0.097915	0.089470
	0.023666	0.021131	0.024254	0.019706	0.074434	0.069017	0.031843	0.034381	0.019634	0.019609	0.017541	0.016261
300 Daily Observations Used in Density Construction												
48	0.052311	0.045615	0.054521	0.033382	0.125713	0.103276	0.080283	0.080969	0.105559	0.092252	0.055699	0.051278
	0.012707	0.011453	0.016435	0.011059	0.097779	0.065338	0.035289	0.034427	0.011321	0.010091	0.006887	0.002720
144	0.042298	0.059979	0.030599	0.029849	0.286725	0.147900	0.073195	0.067814	0.115377	0.110091	0.062447	0.064604
	0.014347	0.013902	0.020849	0.016266	0.124960	0.174496	0.028276	0.022914	0.011883	0.011783	0.005849	0.006626
360	0.067555	0.065137	0.043104	0.038272	0.126808	0.131396	0.090341	0.064934	0.120812	0.119162	0.088854	0.082221
	0.014670	0.014003	0.013635	0.010934	0.036654	0.046135	0.028792	0.021220	0.011772	0.011662	0.010319	0.009593
720	0.069277	0.068735	0.053709	0.045128	0.123266	0.126728	0.111504	0.080115	0.122585	0.121783	0.100337	0.092094
	0.014819	0.014919	0.014989	0.011961	0.040452	0.050374	0.043985	0.028056	0.011700	0.011833	0.011145	0.010491
500 Daily Observations Used in Density Construction												
48	0.050817	0.044888	0.050428	0.030920	0.103168	0.091822	0.081936	0.077020	0.097506	0.089252	0.055612	0.050709
	0.011416	0.010994	0.013578	0.009767	0.042948	0.037774	0.030888	0.026369	0.009035	0.008323	0.005143	0.002010
144	0.064792	0.061609	0.023186	0.029585	0.164110	0.142157	0.069156	0.070648	0.113031	0.109473	0.061246	0.063454
	0.012227	0.011891	0.008253	0.008744	0.192193	0.089447	0.023626	0.024031	0.009181	0.009153	0.004432	0.005035
360	0.069484	0.067564	0.043414	0.038394	0.126607	0.118665	0.082908	0.062560	0.118010	0.116278	0.086316	0.079978
	0.012554	0.011990	0.012327	0.009498	0.051275	0.039298	0.030183	0.019356	0.009245	0.009145	0.008272	0.007510
720	0.070574	0.069734	0.054636	0.045624	0.125353	0.120986	0.105384	0.073423	0.120007	0.119146	0.097885	0.089492
	0.012557	0.012833	0.013231	0.011154	0.039414	0.038302	0.030640	0.022347	0.009316	0.009321	0.008987	0.008481

* Notes: See notes to Table 1.

M	$Realized\ Volatility$	$Bipower\ Variation$	$Sub-Sample\ 1$	$Sub-Sample\ 2$
		$Panel\ A:\ Noise =$ i.i.d.$N(0,(0.5\times1440)^{-1})$		
		$Mean\ Square\ Error$		
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	0.016920	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)	(0.099653)	(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.$N(0,(2\times1440)^{-1})$		
		$Mean\ Square\ Error$		
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)	(0.012597)	(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.

Table 4: Predictive Integrated Volatility Density Estimator Accuracy Experiment*
Case IV: Levels Data - Jumps

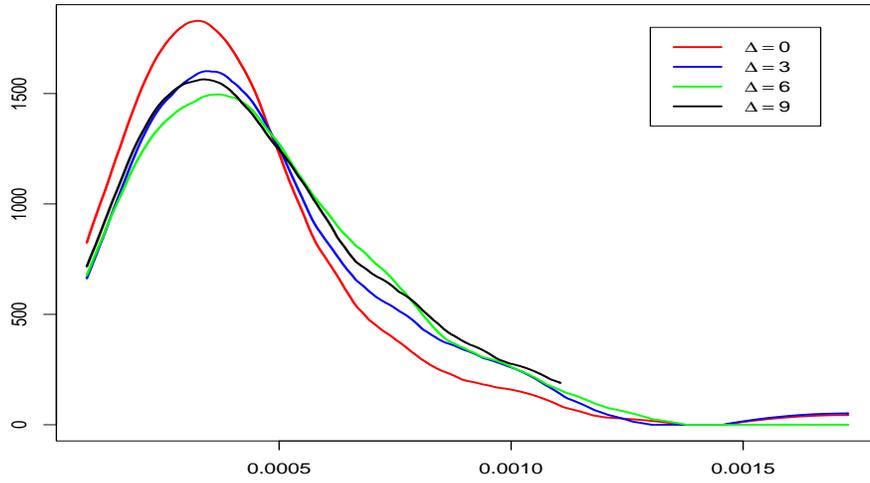| M	Realized Volatility	Bipower Variation	Sub − Sample 1	Sub − Sample 2
	Panel A: One i.i.d.$N(0, 0.64 \times \mathrm{E}(IV_t))$ *Jump Every 5 Days*			
	Mean Square Error			
48	0.144282	0.133275	0.111008	0.149304
	(0.023653)	(0.027431)	(0.053430)	(0.051297)
144	0.139575	0.142283	0.092840	0.133265
	(0.022013)	(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.

Figure 1: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on realized volatility. $M = 78$, $T = 100$.
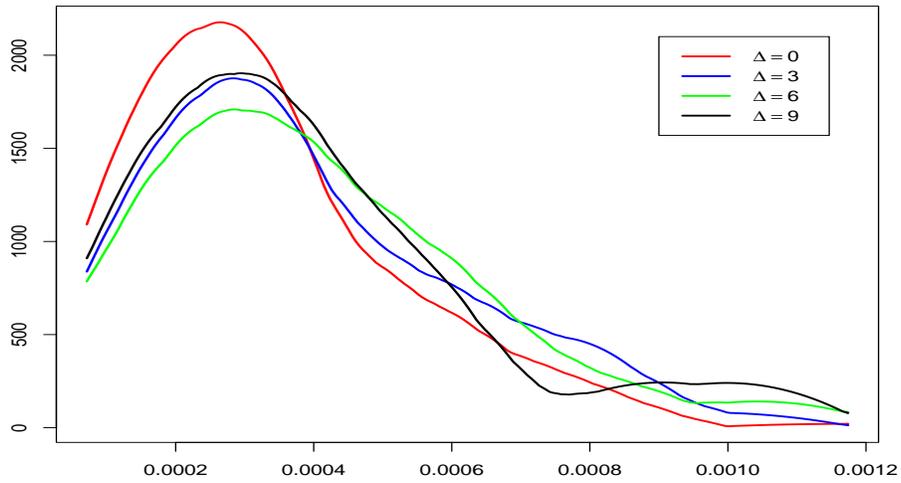

Figure 2: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on bipower variation. $M = 78$, $T = 100$.
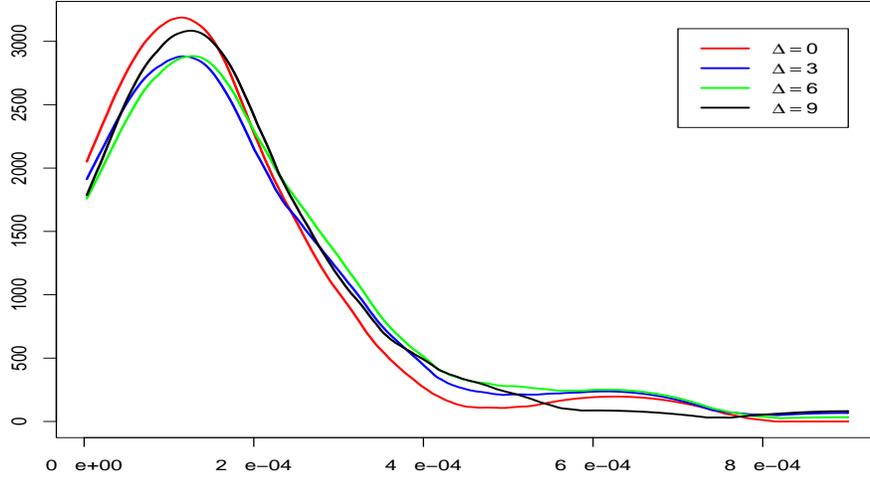
Figure 3: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on $\widehat{RV}_{t,l,M}$. $M = 78$, $T = 100$.
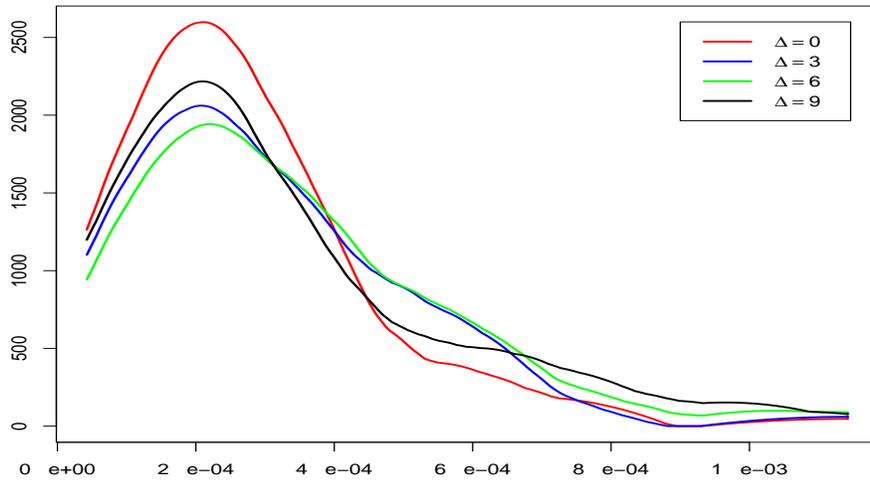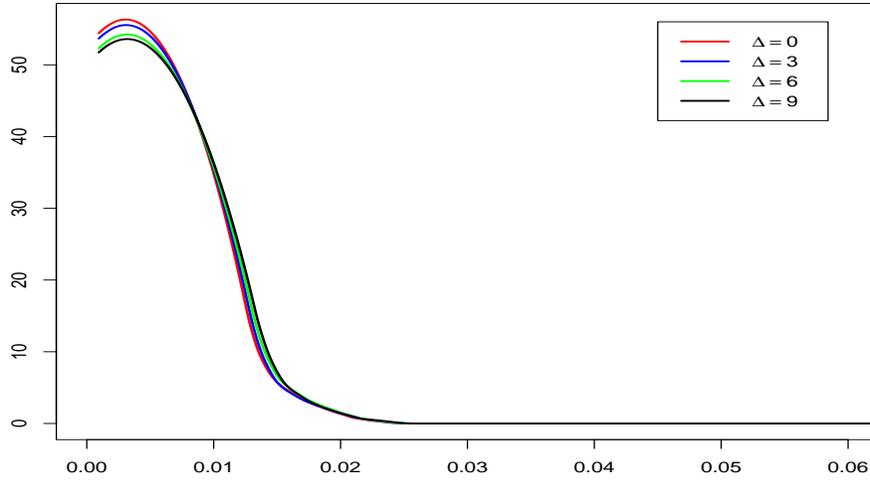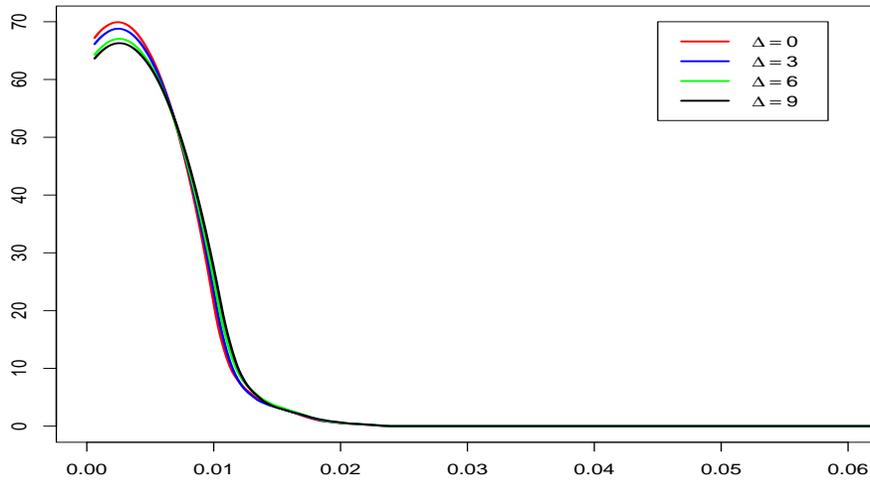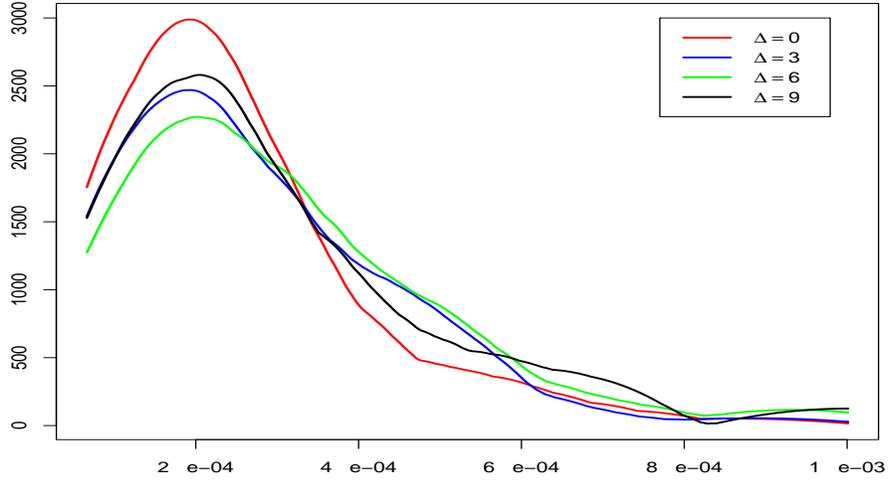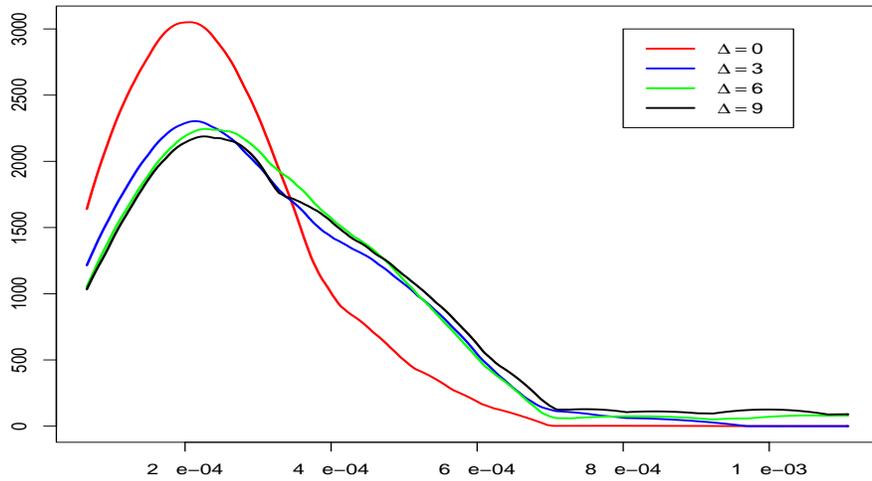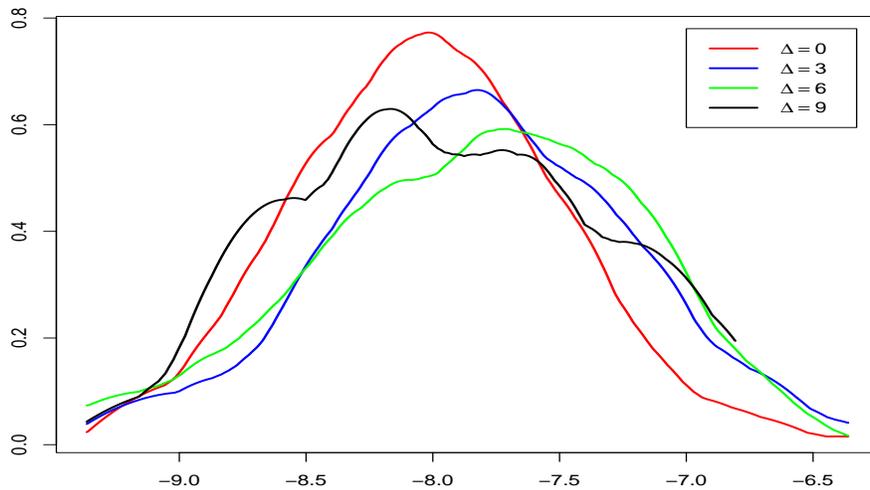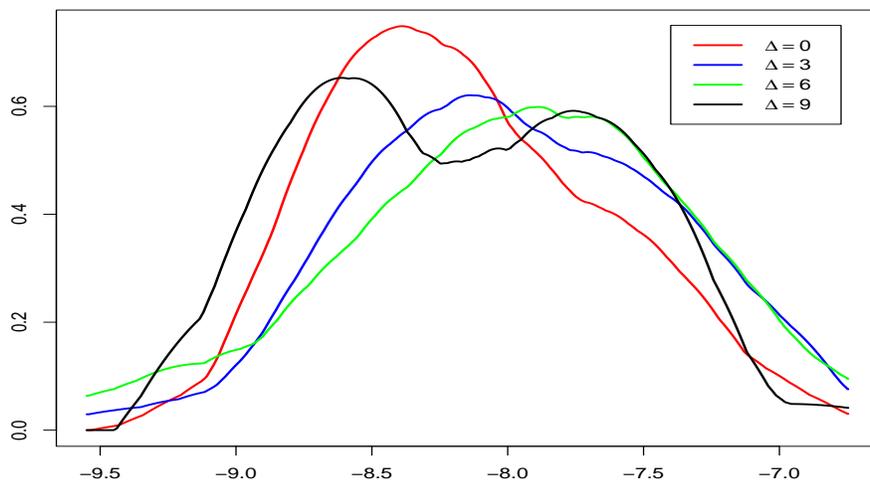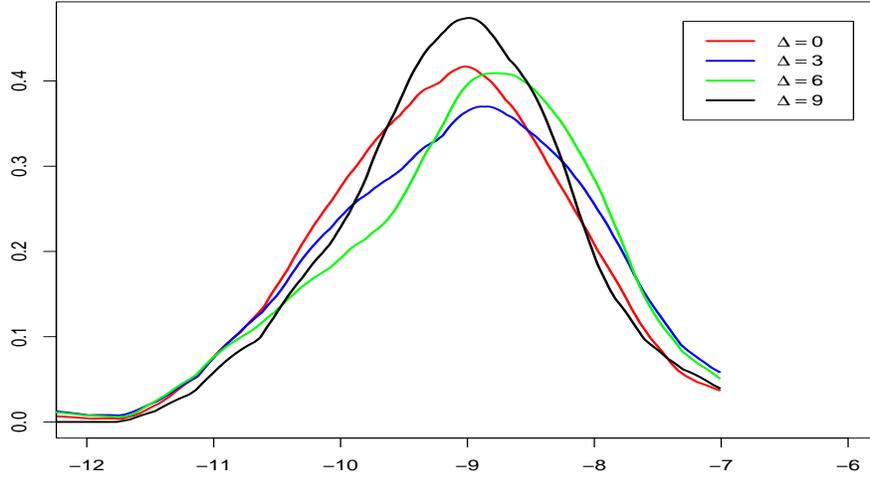

Figure 4: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on $\widetilde{RV}_{t,e,M}$. $M = 78$, $T = 100$.

Figure 5: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on realized volatility. $M = 2340$, $T = 100$.

Figure 6: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on bipower variation. $M = 2340$, $T = 100$.

Figure 7: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on $\widehat{RV}_{t,l,M}$. $M = 2340$, $T = 100$.
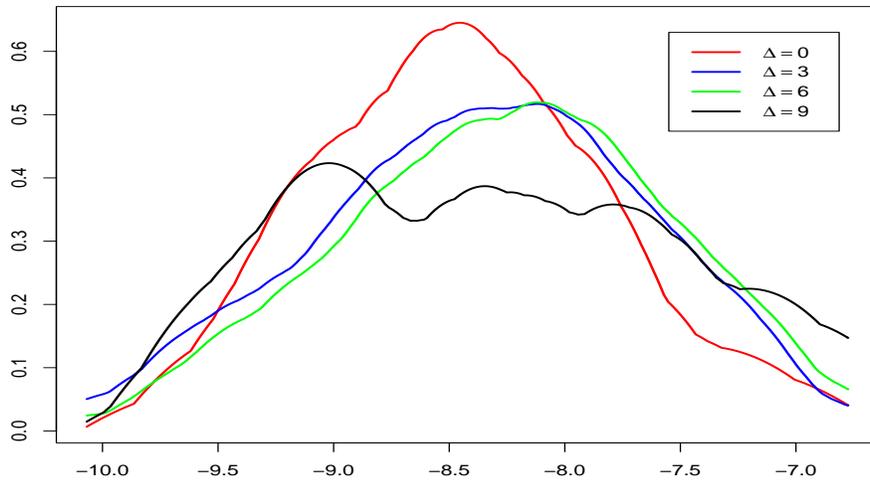

Figure 8: Predictive density of IV_{T+1}, conditional on $IV_{T-\Delta}$, for the Intel stock, based on $\widetilde{RV}_{t,e,M}$. $M = 2340$, $T = 100$.
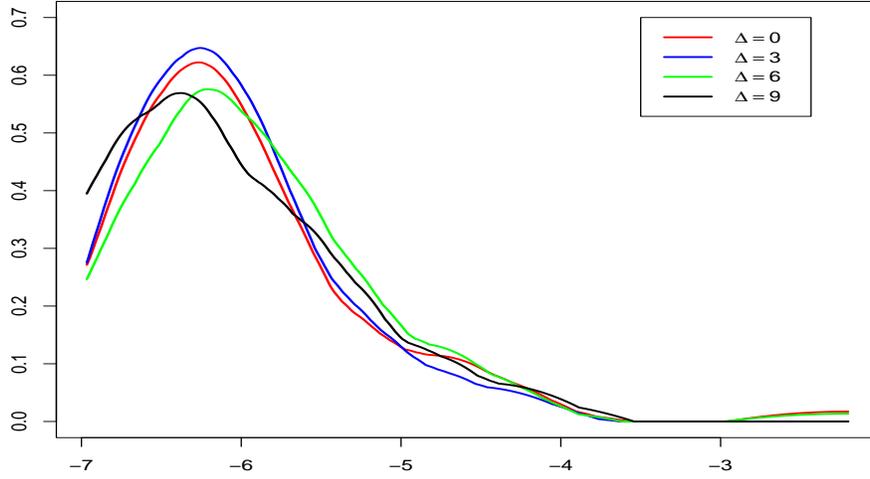
Figure 9: Predictive density of $\log{(IV_{T+1})}$, conditional on $\log{(IV_{T-\Delta})}$, for the Intel stock, based on realized volatility. $M = 78$, $T = 100$.
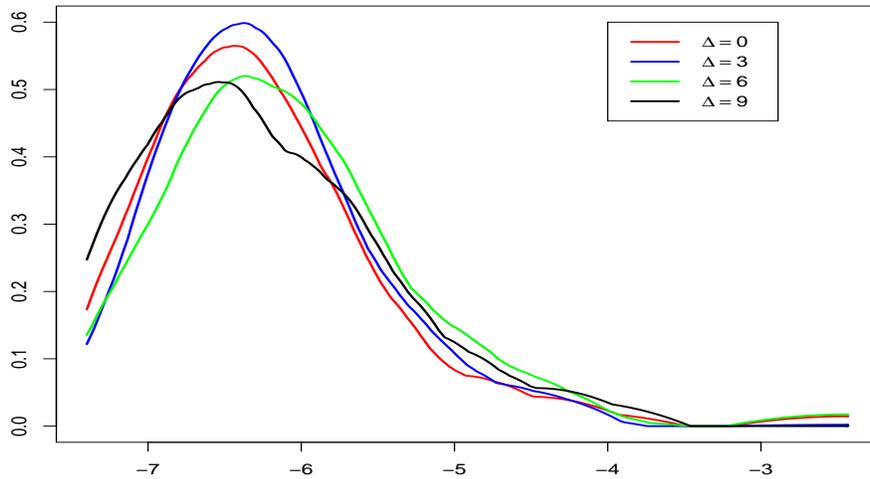

Figure 10: Predictive density of $\log{(IV_{T+1})}$, conditional on $\log{(IV_{T-\Delta})}$, for the Intel stock, based on bipower variation. $M = 78$, $T = 100$.
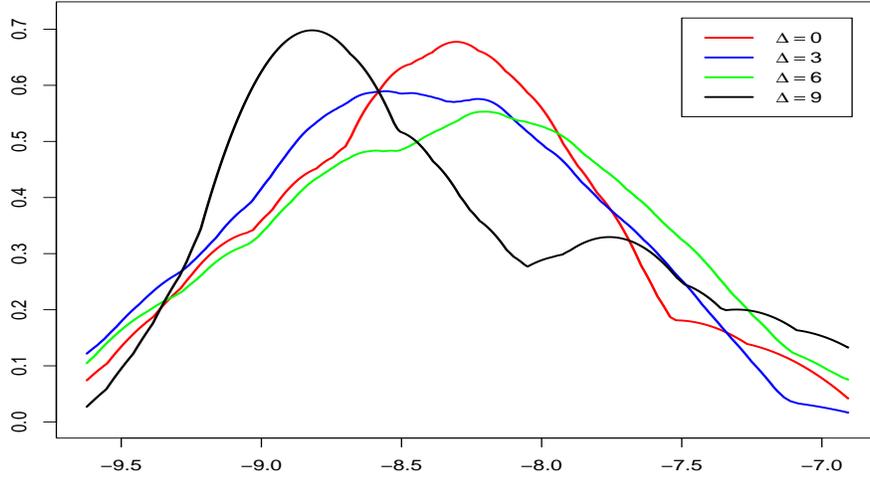
Figure 11: Predictive density of $\log{(IV_{T+1})}$, conditional on $\log{(IV_{T-\Delta})}$, for the Intel stock, based on $\widehat{RV}_{t,l,M}$. $M = 78$, $T = 100$.

Figure 12: Predictive density of $\log{(IV_{T+1})}$, conditional on $\log{(IV_{T-\Delta})}$, for the Intel stock, based on $\widetilde{RV}_{t,e,M}$. $M = 78$, $T = 100$.

Figure 13: Predictive density of $\log\left(IV_{T+1}\right)$, conditional on $\log\left(IV_{T-\Delta}\right)$, for the Intel stock, based on realized volatility. $M = 2340$, $T = 100$.

Figure 14: Predictive density of $\log\left(IV_{T+1}\right)$, conditional on $\log\left(IV_{T-\Delta}\right)$, for the Intel stock, based on bipower variation. $M = 2340$, $T = 100$.

Figure 15: Predictive density of $\log\left(IV_{T+1}\right)$, conditional on $\log\left(IV_{T-\Delta}\right)$, for the Intel stock, based on $\widehat{RV}_{t,l,M}$. $M = 2340$, $T = 100$.
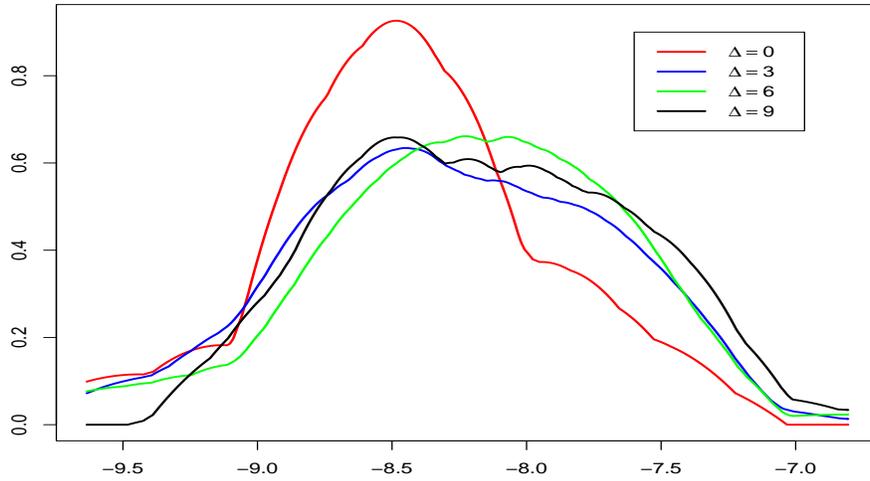

Figure 16: Predictive density of $\log\left(IV_{T+1}\right)$, conditional on $\log\left(IV_{T-\Delta}\right)$, for the Intel stock, based on $\widetilde{RV}_{t,e,M}$. $M = 2340$, $T = 100$.