Microstructure noise, realized volatility, and optimal sampling*

Federico M. Bandi† and Jeffrey R. Russell‡

December 23, 2003

Abstract

Recorded prices are known to diverge from their "efficient" values due to the presence of market microstructure contaminations. The microstructure noise creates a dichotomy in the model-free estimation of integrated volatility. While it is theoretically necessary to sum squared returns that are computed over very small intervals to better indentify the underlying quadratic variation over a period, the summing of numerous contaminated return data entails substantial accumulation of noise.

Using asymptotic arguments as in the extant theoretical literature on the subject, we argue that the realized volatility estimator diverges to infinity almost surely when noise plays a role. While realized volatility cannot be a consistent estimate of the quadratic variation of the log price process, we show that a standardized version of the realized volatility estimator can be employed to uncover the second moment of the (unobserved) noise process. More generally, we show that straightforward sample moments of the noisy return data provide consistent estimates of the moments of the noise process.

Finally, we quantify the finite sample bias/variance trade-off that is induced by the accumulation of noisy observations and provide clear and easily implementable directions for optimally sampling contaminated high frequency return data for the purpose of volatility estimation.

*We thank Tim Conley and the participants at the CIRANO conference "Realized Volatility," Montreal, November 7-8, 2003, for helpful discussions. We are especially indebted to Neil Shephard for his valuable comments.

†Graduate School of Business, University of Chicago, 1101 East 58 street, Chicago, IL 60637.

‡Graduate School of Business, University of Chicago, 1101 East 58 street, Chicago, IL 60637.

1 Introduction

A substantial amount of recent work has been devoted to the model-free measurement of volatility in the presence of high frequency return series (see the review paper by Andersen et al. (2002) and the references therein). The main idea is to aggregate intra-daily squared returns to approximate the daily quadratic variation of the semimartingale that drives the underlying log price process. The consistency result justifying this procedure is the convergence in probability of the sum of squared returns to the quadratic variation of the log price process as returns are computed over intervals that are increasingly small asymptotically. While this result is a cornerstone in semimartingale process theory (see Chung and Williams (Theorem 4.1, page 76, 1990), for instance), the availability of high frequency return data has made it possible to develop a nonparametric theory of inference for volatility estimation that heavily relies on its implications (see Andersen et al. (2003a) and Barndorff-Nielsen and Shephard (2002), BN-S hereafter).

The empirical validity of the procedure hinges on the observability of the true price process. Nonetheless, it is well-accepted that the true price process and, as a consequence, the return data are contaminated by market microstructure effects, such as discrete clustering and bid-ask spreads, among others. In other words, asset prices diverge from their "efficient values" due to a variety of market frictions. BN-S (2002) write "...The implication of this is that it is dangerous to make inference based on extremely large values of M [where M is the number of observations] for the effect of model misspecification can swamp the effects we are trying to measure. Instead it seems sensible to use moderate values of M and properly account for the fact that the realized variance error is not negligible..." In the BN-S's framework an asymptotic increase in the number of observations M translates into finer and finer sampling over time for a fixed time period of interest. Andersen et al. (2001) write "... as such it is not feasible to push the continuous record asymptotics ... beyond this level. ...Such market microstructure features ... can seriously distort the distributional properties of high frequency intra-day returns." In their review paper Andersen et al. (2002) write "...it is undesirable, and due to the presence of market microstructure frictions indeed practically infeasible, to sample returns infinitely often over infinitesimally short time intervals. Model specific calculations and simulations by [many authors[1]] illustrate the effects of finite M [number of observations, that is] and h [time span] for a variety of settings. The discrepancies in the underlying model formulation

[1] See Andersen et al. (2002) for the list of references.

and character of the assumed frictions render a general assessment of the results difficult. Moreover, the size of the measurement errors are often computed unconditionally rather than conditional on the realized volatility statistic. Nonetheless, it is evident that the measurement errors typically are non-trivial."

Using *infill* asymptotic arguments (i.e., increasingly frequent observations over a fixed time span) as in the extant theoretical literature on the subject (c.f., Andersen et al. (2003a) and BN-S (2002)) and a realistic price formation mechanism that accounts for microstructure effects (see Madhavan (2000)), we show that the quadratic variation estimates are swamped by noise as the number of squared return data increases asymptotically. The theoretical manifestation of this effect is a realized volatility estimator that fails to converge to the underlying quadratic variation of the log price process but, instead, diverges to infinity almost surely over any period of time, however small. This result provides a theoretical justification for the diverging behavior at high frequencies of the realized volatility estimates of liquid stocks as reported by Andersen et al. (2000).

Interestingly, despite the fact that realized volatility is not consistent for the conventional object of interest (quadratic variation, that is), a standardized version of the realized volatility estimator can be employed to identify a specific feature of the noise distribution (rather than a feature of the true return process, as generally believed), namely the variance of the (unobservable) noise process. More generally, we show that straightforward sample moments of the contaminated return data can be employed to identify the moments of the underlying noise process.

As stressed earlier, we are not the first ones to point out the potential impact of market microstructure frictions on volatility estimates obtained through aggregation of high frequency squared return data. Nonetheless, while previous discussions of the potential role played by microstructure contaminations are based on informal arguments, rigorous limiting results provide justification for aggregation as a means to uncover, in the limit, the true quadratic variation of the underlying log price process. To this extent, the present paper fills a gap in the existing literature by illustrating the theoretical implications of the presence of microstructure noise on the asymptotic results that are generally invoked to justify quadratic volatility estimation through aggregation of high frequency squared return data.

An important remark is worth making at this point. The use of increasingly frequent observations as a requirement for consistency in nonparametric (point-wise) continuous-time model estimation is now well-accepted. In effect, it is understood that the theoretical necessity for infill limiting

results needs to be interpreted as an asymptotic approximation. Just like the more standard "large n" requirement is meant to signify sufficient accumulation of information, the infill requirement signifies sufficient information in the vicinity of the level at which point-wise estimation is performed. In practise, while the infill approximation permits identification under mild assumptions on the properties of the process of interest (thereby not requiring often difficult stationary density-based identification procedures), its empirical validity is known *not* to hinge on the availability of high frequency observations being that daily sampling is generally sufficient for the approximation to apply (the interested reader is referred to the review paper by Bandi and Phillips (2002) and the references therein for discussions). To this extent, the issue of quadratic variation estimation is fundamentally different from the point-wise identification of continuous-time models in that the very nature of the problem makes the theoretical need for high-frequency return data a stringent empirical requirement in the former case, thereby justifying a closer investigation into more realistic limiting results. This is what the present work hopes to achieve in one of its contributions.

Having made these observations, natural remaining issues are how to formalize the finite sample loss that is induced by a realistic noise component in volatility estimation, and how to employ this information to fully exploit the identification potential of the empirically important notion of realized volatility as introduced by Andersen et al. (2003a) and BN-S (2002).

In keeping with the model-free spirit of the realized volatility literature, we tackle this issue by deriving the conditional (on the underlying volatility path) mean-squared error (MSE, henceforth) of the contaminated volatility estimator. Specifically, we show that the presence of microstructure noise induces a finite sample bias/variance trade off. The idea is simple. When the true price process is observable, as typically assumed in conventional theoretical models, the larger is the sampling frequency over a fixed period of time, the more precise is the estimation of the integrated volatility (or quadratic variation) of the log price process. When the true price process is not observable, as typically the case in practise, frequency increases provide information about the underlying integrated volatility but, necessarily, entail accumulations of noise that affect both the bias and the variance of the estimator. The optimal sampling frequency should be chosen to balance these two contrasting effects. We formalize these ideas by deriving the expression that ties the properties of the *conditional* MSE of the contaminated quadratic variation estimator to the features of the microstructure noise distribution. The idea is therefore similar in spirit to Bai et al. (2000) who consider the MSE of (*unconditional*) variance estimates in the presence of microstructure noise.

Finally, we provide a methodology to optimally choose the sampling frequency as the minimum of the conditional expected squared distance between the estimator (i.e., realized volatility) and its theoretical counterpart (i.e., quadratic variation), as summarized by the conditional MSE. The method relies on the computation of sample moments of contaminated high frequency return data as well as on the minimization of a simple nonlinear function. As such, it is straightforward to implement. We also provide a rule-of-thumb for selecting the optimal frequency without having to implement an otherwise simple minimization routine. Since the rule-of-thumb takes the familiar form of a signal-to-noise ratio (thereby highlighting the main determinants of the optimal sampling frequency), we expect it to be useful in applied work on the subject.

The paper proceeds as follows. In Section 2 we lay out the model. Section 3 is about the limiting properties of the realized volatility estimator when microstructure noise affects fair prices in a realistic manner. In Section 4 we present an expansion of the conditional MSE of the quadratic variation estimator when noise plays a role and discuss optimal sampling through minimization of the conditional MSE. In Section 5 we illustrate the implications of our findings when estimating the quadratic variation of the log price process as well as the second moment of the (unobservable) noise process in the presence of quote-to-quote IBM price changes. The analysis is Section 5 is conducted through empirical work and simulations. Section 6 concludes. Proofs and technical details are in Appendix A. Appendix B lays out the notation.

2 The model

The model we study is coherent with previous theoretical approaches to model-free volatility estimation. Specifically, we employ the same underlying set-up as in BN-S (2002, 2004) but explicitly introduce realistic microstructure effects. The notation is also consistent with BN-S (2002, 2004).

We consider a fixed time period h (a trading day, for instance) and write the observed price process as

$$\widetilde{p}_{ih} = p_{ih}\overline{\eta}_{ih} \qquad i = 1, 2, ..., n, \tag{1}$$

where p_{ih} is the true price and $\overline{\eta}_{ih}$ denotes microstructure noise. A simple log transformation gives us

$$\underbrace{\ln\left(\widetilde{p}_{ih}\right)-\ln\left(\widetilde{p}_{(i-1)h}\right)}_{\widetilde{r}_i}=\underbrace{\ln(p_{ih})-\ln(p_{(i-1)h})}_{r_i}+\underbrace{\eta_{ih}-\eta_{(i-1)h}}_{\varepsilon_i}\qquad i=1,2,...,n, \tag{2}$$

where $\eta=\ln(\overline{\eta})$.

Assumption 1. (The price process.)

(1) *The log price process* $\ln(p_{ih})$ *is a continuous local martingale. Specifically,*

$$\ln(p_{ih})=M_{ih}, \tag{3}$$

where $M_{ih}=\int_0^{ih}\sigma_s dW_s$ *and* $\{W_t:t\geq 0\}$ *is a standard Brownian motion.*

(2) *The spot volatility process* σ_t *is càdlàg and bounded away from zero.*

(3) σ_t *is independent of* $W_t\qquad\forall t.$

(4) *The integrated variance process* $V_t=\int_0^t\sigma_s^2 ds<\infty\ \forall t<\infty$.

We divide the period h into M subperiods and define the observed high frequency returns as

$$\widetilde{r}_{j,i}=\ln\left(\widetilde{p}_{(i-1)h+j\delta}\right)-\ln\left(\widetilde{p}_{(i-1)h+(j-1)\delta}\right)\qquad j=1,2,...,M, \tag{4}$$

where $\delta=h/M$. Hence, $\widetilde{r}_{j,i}$ is the j-th intra-day return for day i. Naturally then,

$$\widetilde{r}_{j,i}=r_{j,i}+\varepsilon_{j,i}, \tag{5}$$

where $r_{j,i}$ and $\varepsilon_{j,i}$ $(=\eta_{(i-1)h+j\delta}-\eta_{(i-1)h+(j-1)\delta})$ have straightforward interpretations given Eq. (2) above.

Assumption 2. (The microstructure noise.)

(1) *The random shocks* η_j *are iid mean zero with a bounded eight moment.*

(2) *The true return process* $r_{j,i}$ *is independent of* $\eta_{j,i}\ \forall i,j$.

Lemma 1 below illustrates the moments of the noises-in-returns $\varepsilon^{'}$s as a function of the moments of the price contaminations, i.e., $\eta^{'}$s.

Lemma 1. *Under the specification in Eq. (2) the following relations hold*:

(1) $\mathbf{E}(\varepsilon^2) = 2\mathbf{E}(\eta^2)$,

(2) $\mathbf{E}(\varepsilon^4) = 2\mathbf{E}(\eta^4) + 6\left(\mathbf{E}(\eta^2)\right)^2$,

(3) $\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) = 3\left(\mathbf{E}(\eta^2)\right)^2 + \mathbf{E}(\eta^4)$.

Proof. *See Appendix A.*

Lemma 2 relates the first order cross-moment of the squared noises-in-returns to the fourth moment of the noise-in-return.

Lemma 2. *Under the specification in Eq. (2) the following relation holds:*

$$\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) = \frac{1}{2}\mathbf{E}(\varepsilon^4). \tag{6}$$

Proof. *Immediate given results (2) and (3) in Lemma 1.*

Some observations on the set-up are needed. The true return process r is modelled as a local martingale with bounded variance $\mathbf{E}(r_{j,i}^2)$ equal to $\mathbf{E}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)$ over any period δ (c.f., Assumption 1(1) and 1(4)). The spot volatility σ is allowed to display jumps, diurnal effects, long-memory features,[2] and nonstationarities (c.f., Assumption 1(2)). Consistently with existing theoretical treatments (see BN-S (2002, 2004), for instance), we rule out leverage effects (c.f., Assumption 1(3)). Nonetheless, while the extant literature has pointed out that the presence of leverage generally induces second-order effects (Andersen et al. (2003b), BN-S (2003), and Meddahi (2002), among others), in the sequel (see Section 3 and Appendix A below) we show that our limiting results (as represented by Theorem 1 and 2) are robust to the existence of leverage.

The econometrician does not observe r, the true return series, but a contaminated return series $\widetilde{r}$ which is given by r plus a random shock ε that is independent of r (c.f., Assumption 2(2)). We interpret the $\varepsilon^{'}$s as being microstructure contaminations in returns. In virtue of the specification in Eq. (2) above and Assumption 2(1), the shocks $\varepsilon^{'}$s are identically distributed with a bounded eight moment. Nonetheless, they are not uncorrelated since their first-order autocovariance is negative and

[2]Long-memory is known to be an important feature of volatility series. The interested reader is referred to Bandi and Perron (2001), Ohanissian et al. (2003), and the references therein for some recent evidence.

equal to $-\mathbf{E}(\eta^2) = -\sigma_\eta^2$, i.e., the variance of the underlying shocks $\eta^{'}$s taken with a negative sign. The negative first-order autocorrelation of the microstructure contaminations in returns determines an analogous first-order autocorrelation in the contaminated return series. This feature of the noise specification captures a well-known empirical fact (see Niederhoffer and Osborne (1966), Cohen et al. (1979), and Roll (1984) for some early findings). In Section 5 we confirm this fact for IBM. Using mid-point bid ask quotes, in a companion paper we find strong negative first-order autocorrelations (and higher-order autocorrelations that are economically negligible and often statistically insignificant) for the majority of the S&P 100 stocks (Bandi and Russell (2003b)).

While being supported by a vast empirical evidence, the structural model implied by Eq. (1) appears to be a natural set-up to analyze the impact of microstructure contaminations on the realized volatility estimates. In effect, well-known canonical microstructure models with trading frictions and private information can easily be cast into our framework. One early example is Roll's implicit bid-ask model (Roll (1984)). The interested reader is referred to Campbell et al. (1996) and the review paper by Madhavan (2000) for a complete discussion of Roll's price formation mechanism and recent advances. It is noted that Aït-Sahalia and Mykland (2003) adopt a similar set-up in their analysis of the impact of microstructure noise on the parametric (i.e., maximum likelihood) estimates of the second infinitesimal moment of scalar diffusion models.

There is a subtle reason, which has to do with the orders of magnitude of the quantities involved, why we expect the model in Eq. (2) to capture the main effects in high-frequency data. The idea goes as follows. Different trading institutions and different price measurements potentially have different microstructure characteristics which, in turn, determine and characterize η. Generally speaking, high-frequency financial data provide both bid and ask prices as well as transaction prices. It is common practice to use the mid-point of the prevailing bid and ask prices as a (noisy) measure of the true price. In fact, actual transaction prices suffer from well-known bid ask bounce effects and are thought to be more noisy than the midpoint of the quote measurements. In agreement with this observation, the empirical work in this paper focuses on mid-points of bid and ask quotes. Specifically, our methodology exploits a fundamental difference in the nature of the true returns and the noise associated with its mid-quote observations. The efficient price is considered a continuous process. It is the price that would prevail in the absence of market frictions. Thus, the dynamics of the efficient price should be driven by a smooth process reflecting the continual updating and learning on the part of the market participants. In effect, it takes time for the market participants to react and

digest new information. Hence, with the exception of important rare public news announcements, the price will not likely jump from one level to another, but rather smoothly adjust as the market comes to grips with any new information. The characteristics of the noise are substantially different from the true price characteristics since posed quotes in a market inherently reflect different information than the efficient price. Observed prices are not permitted to vary continuously, but rather fall on a fixed grid of prices or ticks. Changes in the mid-quotes are therefore discrete in nature. Furthermore, classic microstructure theory suggests that a market maker posting quotes will take into consideration the nature of the limit order book, current inventory levels, as well as the risks associated with asymmetric information. Adjustments to these components are necessarily discrete in nature as new limit orders are submitted or a large market order consumes all of the limit orders at some given price. The transaction process is also thought to carry information regarding the likelihood of asymmetric information suggesting that this component may also adjust discretely. When one accounts for the fact that adjustments to the information used to post the quotes are not smooth coupled with the fact that observed prices must fall on a grid of tick values, it is natural to consider the departures of the observed price from the true price as a discontinuous process. Hence, provided we do not sample at a rate faster than new price information arrives (i.e., between quote updates), the noise in the observed price process should be roughly i.i.d and therefore consistent with our assumed structure.

Having made these points, we can easily generalized the noise process to more involved correlation structures without changing the main results in the paper. It will be clear that richer dependence features in the noise process simply determine a more complicated variance expression in the asymptotic distribution of the (standardized) realized volatility estimator in Section 3 as well as a more involved (but estimable) variance term in the conditional MSE expansion in Section 4.

Coherently with Andersen et al. (2003a) and BN-S (2002) we define the realized volatility estimator for the generic period i as

$$\widehat{V}_i = \sum_{j=1}^{M} \widetilde{r}_{j,i}^2. \tag{7}$$

We use $\widehat{V}_i$ to estimate $V_i = \int_{(i-1)h}^{ih} \sigma_s^2 ds$, i.e., the quadratic variation of the log price process over the same period. In the next section we discuss the asymptotic properties of $\widehat{V}_i$. As pointed out earlier, in agreement with the existing literature on the subject, our asymptotics are conducted by

increasing the number of observations M over a fixed time span h.

3 Microstructure noise and the limiting distribution of the realized volatility estimator

We can rewrite the realized volatility estimator in Eq. (7) as the sum of three components, namely

$$\widehat{V}_i = \underbrace{\sum_{j=1}^{M} r_{j,i}^2}_{A_i} + \underbrace{\sum_{j=1}^{M} \varepsilon_{j,i}^2}_{B_i} + \underbrace{2\sum_{j=1}^{M} r_{j,i}\varepsilon_{j,i}}_{C_i}. \tag{8}$$

If the true price process were observable, only the term A_i would drive the limiting properties of $\widehat{V}_i$ (as in BN-S (2002)). The presence of microstructure noise introduces two additional components, i.e., B_i and C_i. We will show that it is mainly term B_i that makes standard consistency arguments fail. Intuitively, B_i diverges to infinity almost surely as the number of observations increases asymptotically (or, equivalently, as the frequency of observations increases in the limit) since more and more noise is being accumulated for a fixed period of time h.

Theorem 1 below contains a characterization of our findings.

Theorem 1. *If Assumptions 1 and 2 are satisfied, then*

$$\widehat{V}_i \underset{M\to\infty}{\overset{a.s.}{\to}} \infty \qquad \forall i. \tag{9}$$

Furthermore,

$$\sqrt{M}\left(\frac{\widehat{V}_i}{M} - \mathbf{E}\left(\varepsilon^2\right)\right) \underset{M\to\infty}{\Rightarrow} \mathbf{N}\left(0, \mathbf{V}\left(\varepsilon^2\right) + 2\mathbf{E}\left(\left(\varepsilon^2 - \mathbf{E}\left(\varepsilon^2\right)\right)\left(\varepsilon_{-1}^2 - \mathbf{E}\left(\varepsilon^2\right)\right)\right)\right) \qquad \forall i, \tag{10}$$

which implies

$$\sqrt{M}\left(\frac{\widehat{V}_i}{M} - 2\mathbf{E}\left(\eta^2\right)\right) \underset{M\to\infty}{\Rightarrow} \mathbf{N}\left(0, 4\mathbf{E}\left(\eta^4\right)\right) \qquad \forall i, \tag{11}$$

or, equivalently,

$$\sqrt{M}\left(\frac{\frac{1}{2}\widehat{V}_i}{M} - \mathbf{E}\left(\eta^2\right)\right) \underset{M\to\infty}{\Rightarrow} \mathbf{N}\left(0, \mathbf{E}\left(\eta^4\right)\right) \qquad \forall i. \tag{12}$$

Proof. *See Appendix A.*

Some remarks are in order.

Remark 1. The asymptotic properties of term A_i are known. Specifically, A_i converges in probability to the underlying quadratic variation of the log price process over the period, namely $V_i = \int_{(i-1)h}^{ih} \sigma_s^2 ds$ (Chung and Williams (1990), for instance). Its asymptotic distribution is mixed-normal and can be expressed as

$$\sqrt{\frac{M}{h}} \left(A_i - V_i \right) \underset{M \to \infty}{\Rightarrow} \mathbf{MN} \left(0, 2Q_i \right), \tag{13}$$

where $Q_i = \int_{(i-1)h}^{ih} \sigma^4 ds$ is the so-called quartic variation (BN-S (2002)). The interested reader is referred to BN-S (2002) for an introduction to the notion of quartic variation and for a thorough discussion of the above weak convergence result. In the Appendix we provide a proof of the same result that employes different techniques borrowed from semimartingale process theory. Specifically, we generalize the BN-S (2002) findings, as represented by Eq. (13) above, to a functional central limit theorem specification while relaxing the assumption of no leverage effects.

Remark 2. The limiting features of term B_i can be studied by using standard methods for stationary mixing sequences (see Hamilton (1994), for instance). In effect,

$$\sqrt{M} \left(\frac{B_i}{M} - \mathbf{E} \left(\varepsilon^2 \right) \right) \underset{M \to \infty}{\Rightarrow} \mathbf{N} \left(0, \mathbf{V} \left(\varepsilon^2 \right) + 2\mathbf{E} \left(\left(\varepsilon^2 - \mathbf{E} \left(\varepsilon^2 \right) \right) \left(\varepsilon_{-1}^2 - \mathbf{E} \left(\varepsilon^2 \right) \right) \right) \right). \tag{14}$$

As pointed out earlier, the form of the asymptotic variance term depends on the specific correlation structure of the microstructure noise as defined by Eq. (2) and Assumption 2(1). Naturally, a richer specification would not affect the empirical significance of the limiting results reported in this section and could be easily accounted for.

It is apparent that boundedness of the fourth moment of the noise process is all that is required for the above weak convergence result to be true. We impose boundedness of the eight moment (c.f., Assumption 2(1)) for the statement in Theorem 2 below to be satisfied. Of course, this property ought to be true by virtue of the fact that the noise process is bounded in practise.

Remark 3. The asymptotic features of term C_i rely on embedding arguments for local martingales. Specifically,

$$C_i \underset{M\to\infty}{\Rightarrow} \mathbf{MN}\left(0, 2\sigma_\eta^2 V_i\right). \tag{15}$$

Intuitively, the sum that constitutes C_i does not diverge to infinity as the number of observations M increases without bound since each noise term is multiplied by increasingly smaller random returns.

Remark 4. Remark 1 through 3 imply that

$$\begin{aligned} & A_i + B_i + C_i \\ = \; & \left(V_i + O_p\left(\frac{1}{\sqrt{M}}\right)\right) + M\mathbf{E}\left(\varepsilon^2\right) + O_p\left(\sqrt{M}\right) + O_p(1) \underset{M\to\infty}{\overset{a.s.}{\to}} \infty, \end{aligned} \tag{16}$$

as $M \to \infty$, thereby justifying the first statement in Theorem 1. Divergence to infinity is induced by term B_i in that the summing of increasingly frequent squared return data causes infinite accumulation of noise.

Remark 5. We can standardize the realized volatility estimates by M and consider the estimation error decomposition given by

$$\begin{aligned} & \frac{\widehat{V}_i}{M} - \mathbf{E}\left(\varepsilon^2\right) \\ = \; & \frac{A_i + B_i + C_i}{M} - \mathbf{E}\left(\varepsilon^2\right) && (17) \\ = \; & \frac{\int_{(i-1)h}^{ih} \sigma_s^2 ds}{M} + O_p\left(\frac{1}{M^{3/2}}\right) + \left(\frac{B_i}{M} - \mathbf{E}\left(\varepsilon^2\right)\right) + O_p\left(\frac{1}{M}\right). && (18) \end{aligned}$$

Thus, the expression

$$\sqrt{M}\left(\frac{\widehat{V}_i}{M} - \mathbf{E}\left(\varepsilon^2\right)\right) = \sqrt{M}\left(\frac{B_i}{M} - \mathbf{E}\left(\varepsilon^2\right)\right) + O_p\left(\frac{1}{\sqrt{M}}\right) \tag{19}$$

justifies the second statement in Theorem 1. Interestingly, even though one cannot consistently estimate the underlying quadratic variation using $\widehat{V}_i$ when noise is present, a standardized version of the quadratic variation estimator allows us to identify the second moment of the (unobservable) noise process by exploiting the asymptotic properties of the dominating term B_i.

Thus, Remark 5 suggests the following Lemma.

Lemma 3. *While one can consistently estimate the second moment of the noise process, i.e.,* $\mathbf{E}\left(\varepsilon^2\right)$, *using* $\frac{\widehat{V}_i}{M}$ *and characterize the finite sample bias of the realized volatility estimator conditionally on the volatility path, namely*

$$bias = M\mathbf{E}\left(\varepsilon^2\right),$$

one cannot hope to consistently estimate V_i *nonparametrically using* $\widehat{V}_i$ *by controlling for the existing (increasing-in-*M*) bias term.*

Hence, any statement about the informational content of the conventional realized volatility estimator as a measurement of the quadratic variation of the underlying log price process ought to be a finite sample statement. Contrary to common intuition, consistency arguments based on limiting findings can only be invoked when estimating features of the noise distribution. The second moment of the noise process is, of course, not an exception. In Theorem 2 below we show that a simple arithmetic average of fourth powers of the contaminated return series converges to the fourth moment of the unobserved noise-in-return process. While it is clear that the procedure is general enough to be applicable to a variety of different moments (including cross-correlations), for brevity we focus on the fourth moment of the noise process in that it will be a necessary input to formulate an optimal sampling theory for $\widehat{V}_i$ as an estimator of the quadratic variation of the underlying log price process.

Theorem 2. *If Assumptions 1 and 2 are satisfied, then*

$$\frac{1}{M}\sum_{j=1}^{M}\widetilde{r}_{j,i}^4 \xrightarrow{p} \mathbf{E}\left(\varepsilon^4\right). \tag{20}$$

Proof. *See Appendix A.*

We now move from asymptotic arguments to a characterization of the finite sample bias/variance trade-off that is induced by noise accumulation. Specifically, we derive the MSE of the realized volatility estimator conditionally on the volatility path. Our strategy will be to learn about the underlying quadratic variation of the log price process through the minimization of the conditional expected squared loss of the realized volatility estimator as represented by its conditional MSE, i.e.,

$$\mathbf{E}_\sigma\left(\widehat{V}_i - V_i\right)^2 = \mathbf{E}_\sigma\left(\sum_{j=1}^{M}\widetilde{r}_{j,i}^2 - \int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2. \tag{21}$$

In the next section we show that the MSE does not converge to zero as the sampling frequency increases without bound. Specifically, the minimum MSE is achieved for a finite number of observations M^*. Naturally, M^* depends on the moments of the microstructure noise distribution as well as on the quarticity of the underlying log price process.

4 The conditional MSE and optimal sampling

The conditional MSE of the integrated volatility estimator can be represented as in Theorem 3 below.

Theorem 3. *If Assumptions 1 and 2 are satisfied, then*

$$\mathbf{E}_\sigma\left(\widehat{V}_i - \int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 = 2\frac{h}{M}\left(Q_i + o_{a.s.}(1)\right) + M\beta + M^2\alpha + \gamma, \tag{22}$$

where the parameters α, β, *and* γ *are defined as follows*:

$$\alpha = \left(\mathbf{E}(\varepsilon^2)\right)^2, \tag{23}$$

$$\beta = \mathbf{E}\left(\varepsilon^4\right) + 2\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) - 3\left(\mathbf{E}(\varepsilon^2)\right)^2, \tag{24}$$

and

$$\gamma = 4\mathbf{E}(\varepsilon^2)V_i - 2\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) + 2\left(\mathbf{E}(\varepsilon^2)\right)^2. \tag{25}$$

Proof. *See Appendix A.*

Should the return series not be affected by the microstructure noise, then the conditional MSE of the quadratic variation estimator would decrease to zero in the limit as the number of observations diverges to infinity. In effect, the MSE would reduce to the conditional variance of the sum of squared returns, i.e., $2\frac{h}{M}\left(Q_i + o_{a.s.}(1)\right)$ (see BN-S (2002) and Appendix A for a derivation).

When microstructure noise is present, Eq. (22) clarifies that the conditional MSE does not vanish as the number of observations M diverges to infinity asymptotically (or, equivalently, as the

sampling frequency increases over time). Summing up contaminated squared returns induces both an additional variance term and a bias term $\mathbf{E}_\sigma(\widehat{V}_i - V_i)$ that have the potential to affect substantially the conditional MSE decomposition. The form of the additional variance term is

$$M\mathbf{E}(\varepsilon^4) + 2(M-1)\mathbf{E}(\varepsilon^2\varepsilon^2_{-1}) + (2-3M)\left(\mathbf{E}(\varepsilon^2)\right)^2 + 4\mathbf{E}(\varepsilon^2)V_i, \tag{26}$$

where $\mathbf{E}(\varepsilon^4)$, $\mathbf{E}(\varepsilon^2\varepsilon^2_{-1})$, and $\mathbf{E}(\varepsilon^2)$ are obvious moments of the noise-in-return distribution (see Lemma 1 above) while $4\mathbf{E}(\varepsilon^2)V_i$ is an interaction term. The form of the bias is $M\mathbf{E}(\varepsilon^2)$. Apparently, both quantities diverge to infinity linearly with M, thereby inducing quadratic growth to infinity (with M) of the corresponding MSE. The coefficients β and γ depend on the correlation structure of the noise. Their expressions can be readily modified to account for higher order (up to $M-1$) correlations as follows:

$$\overline{\beta} = \mathbf{E}\left(\varepsilon^4\right) + 2\sum_{b=1}^{s}\mathbf{E}\left(\varepsilon^2\varepsilon^2_{-b}\right) - (2s+1)\left(\mathbf{E}(\varepsilon^2)\right)^2, \tag{27}$$

$$\overline{\gamma} = 4\mathbf{E}(\varepsilon^2)V_i - 2\sum_{b=1}^{s} b\mathbf{E}\left(\varepsilon^2\varepsilon^2_{-b}\right) + s(s+1)\left(\mathbf{E}(\varepsilon^2)\right)^2. \tag{28}$$

Thus, the set-up that we propose is general enough to allow for unrestricted distributional assumptions on the noise-in-returns ε as well as more involved dependence features.

Our approach relates to previous work on volatility estimation. Bai et al. (2002) are the first to suggest an MSE expansion for unconditional volatility estimates in the presence of microstructure noise. Their framework, though, does not provide implications about how moments at different frequencies relate. Aït-Sahalia and Mykland (2003) and Oomen (2003) study MSE values at difference frequencies in the presence of noise. While Aït-Sahalia and Mykland (2003) derive closed-form expressions for the unconditional MSE of the constant variance estimator of a drift-less diffusion when (Gaussian) noise plays a role, Oomen (2003) uses a structural model of price formation to provide simulated MSE plots for noisy quadratic variation estimates as a function of the sampling interval in the absence of a closed-form specification for the relation between the relevant MSE and the sampling frequency. Consistently with Oomen (2003), we focus on the quadratic variation of a local martingale with time-varying stochastic variance. Coherently with Aït-Sahalia and Mykland (2003), we provide a closed-form expression for the corresponding MSE expansion as a function of the sampling frequency.

We now turn to optimal sampling. Using Eq. (22) above, we define the optimal number of observations (per unit interval h) as the number M^* which satisfies the following condition:

$$\left\{ M^* := \arg\min \left(2\frac{h}{M} \left(Q_i + o_{a.s.}(1) \right) + M\beta + M^2\alpha + \gamma \right) \right\} \tag{29}$$

or, equivalently,

$$\left\{ M^* := M : 2M^3\alpha + M^2\beta - 2h \left(Q_i + o_{a.s.}(1) \right) = 0 \right\}, \tag{30}$$

where the constant terms α, β, and γ were defined earlier.

Lemma 4. (A useful rule-of-thumb for applications.) *For high (optimal) frequencies,*

$$M^* \sim \left(\frac{hQ_i}{\left(\mathbf{E}(\varepsilon^2) \right)^2} \right)^{1/3}, \tag{31}$$

where Q_i is the realized quartic variation and $\mathbf{E}(\varepsilon^2)$ is the second moment of the noises-in-returns.

Proof. *Immediate given Eq. (29).*

Interestingly, when the quadratic term in Eq. (29) dominates the linear term (for values of M sufficiently large), the approximation in Eq. (31) provides a very good representation of the optimal number of observations M. In Section 5 we show that this property holds for a very liquid stock like IBM. Bandi and Russell (2003b) confirm the validity of this result for a large number of S&P100 stocks.

Lemma 4 is important for two reasons. First, it provides us with a very handy and immediate rule-of thumb to choose the optimal M without having to go through an otherwise rather simple minimization routine as in Eq. (29). Second, it clearly illustrates what the main determinants of the optimal frequency are, namely the underlying quarticity of the log price process and the (squared) variance of the noise-in-returns. Naturally, M^* can be regarded as a signal-to-noise ratio: the stronger the signal is, the higher the optimal frequency should be.

4.1 Estimating the optimal sampling frequency

Eq. (30) can be readily solved numerically in the presence of consistent estimates of the quarticity and the relevant terms in Eqs. (23) and (24).

BN-S (2002) provide an estimator of the quarticity that is consistent in the absence of microstructure frictions, i.e., $\widehat{Q}_i = \frac{M}{3h}\sum_{j=1}^{M} \widetilde{r}_{j,i}^4$. Inevitably, $\widehat{Q}_i$ loses its consistency features in the presence of the price formation mechanism implied by Eq. (1). Although Q_i cannot be consistently estimated using $\widehat{Q}_i$, the simulations in the next section show that the use of different estimates of it (as provided by values of $\widehat{Q}_i$ computed on the basis of frequencies that are widely employed in the existing applied work) does not have any considerable impact on the optimal sampling frequency of the realized volatility estimator.

While we can provide informative estimates of Q_i, the availability of high frequency data, along with the results reported in Theorems 1 and 2 above, allow us to consistently estimate the remaining inputs of the minimum problem in Eq. (30), i.e., the second and fourth moment of the microstructure contaminations. More precisely, since our model implies $\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) = 2\mathbf{E}\left(\varepsilon^4\right)$ (c.f., Lemma 2), Theorem 1 and 2 provide a simple strategy to identify all of the relevant moments of the noise distribution by simply averaging powers of the contaminated high frequency return data.

As pointed out earlier, even though noisy return data collected at high frequencies do not permit us to identify the object of interest, i.e., quadratic variation, using the conventional realized volatility estimator, they do allow us to estimate features of the microstructure contaminations. We use those features to learn about quadratic variation through the solution of Eq. (30) above.

One final observation is needed. The conditional MSE in Eq. (22) applies to individual periods h, thereby requiring repeated applications of the procedure. We can readily obtain an optimal (h-period) frequency M^* that is valid for the entire data set by simply working with an integrated version of the conditional MSE in Eq. (22). In other words, we can minimize the average (over i) of the individual conditional MSE's. Apparently, this procedure coincides with solving the program in Eq. (30) above with $\frac{1}{n}\sum_{i=1}^{n} Q_i$, where n denotes the number of periods h, in place of Q_i.

In Section 5 we provide an application of our methodology to quote-to-quote IBM return data.

5 The case of IBM

5.1 How big is the *unobserved* noise component of the *observed* IBM price process?

This section explores the magnitude of the unobserved noise component of the observed IBM price process. Theorem 1 implies that a rescaled version of the realized volatility estimator converges to the variance of the noise process. Hence, averaging very finely sampled squared returns should

provides a good estimate of the variance of the (unobservable) microstructure noise. Ideally, we would like to sample quote-to-quote, the highest frequency that new price information appears.

Our data consists of quote-to-quote changes in the midpoint of posted bid and ask prices. The quotes were obtained from the TAQ data set for the month of February 2002. Restricting our attention to NYSE updates, and after removing any suspicious quotes, we are left with $41,841$ quote price updates over the month. On average a new price quote arrives every 10.6 seconds. The midpoint of the price quotes are used to construct quote-to-quote returns. The smallest return is -2.9% and the largest is .9167%. The first-order autocorrelation is significantly negative and equal to -0.541. The higher order autocorrelations are generally insignificant. Naturally, the largest autocorrelation (after the first one) is the second-order autocorrelation whose value is equal to 0.048. While this estimate is statistically significant, its magnitude is virtually ten times smaller (in absolute value) than the corresponding value for the first-order autocorrelation. Thus, our model captures the main effects in the data.

The square root of the rescaled realized volatility (computed over a 6.5 hour trading period) from Theorem 1 and Remark 5 is .0278%. Notice that this estimate is essentially the sample standard deviation imposing a mean return of zero. If we instead use fixed intervals of 30 seconds, the square root of the rescaled variance increases to .0599%. To put this in dollar context, consider that the average price for IBM over the month of February in our sample was around 100 dollars. Also, recall that the variance of the noise term η in Eq. (2) is one half the variance of the return contamination ε. Thus, the standard deviation of the log noise obtained from the quote-to-quote price moves is given by $\sigma_\eta = .000278/\sqrt{2} = .000197$. Since $\sigma_\eta \sim \sigma_{\overline{\eta}}$, where $\overline{\eta} = \exp(\eta)$, then the standard deviation of the (average) IBM price over the period is about 2 cents. For added perspective, the average spread for IBM in our sample is 7.2 cents. Hence, the standard deviation is small relative to the spread with about a $+/-$ 2 standard deviation interval just about equal to the average spread. The estimated magnitude of the noise variance seems very plausible.

In finite samples, it is clearly the magnitude of the noise variance relative to the variance of the "true" return that is of interest. We therefore calculate the realized volatility for each day in the sample using 15-minute time intervals. While realized volatility remains upward biased in the presence of noise even at the 15-minute sampling frequency we still believe this statistic is useful as a benchmark. Our belief, which is consistent with conjectures that have been put forward in the extant empirical literature (see Andersen et al. (2000), for instance), is supported by the simulations

in the next subsection. The mean realized volatility over the month of February yields an estimate of the daily standard deviation of 1.652%. Thus, we find that the noise variance is .01422% of the typical (average) realized volatility. Naturally, this is a conservative assessment in that the realized volatility estimates based on 15-minute time intervals are slightly inflated by residual noise.

5.2 The bias in the realized volatility estimates: simulations based on IBM

This subsection of the paper simulates data from the model given in Eq. (2) using realistic parameter values based on IBM. From the simulated data estimates of the realized volatility can be compared to the value of the true quadratic variation. Consistently with the results in Theorem 1, we show that the realized volatility does indeed explode as the sampling interval goes to zero. We also show that the appropriately standardized realized volatility converges to the variance of the noise process. Finally, for the parameter values used, we show that a relatively small bias is present in the realized volatility estimates at sampling intervals around 15 minutes. Nonetheless, the bias can be considerable at higher frequencies. For instance, it can be substantial at the commonly-employed (in applied work) 5 minute interval.

Simulations require specifying a process for both the spot volatility σ as well as the noise term η. As in BN-S (2003), we adopt a square root specification for the evolution of the spot volatility. Specifically, the infinitesimal variation of the true price process is given by

$$d\log(p)_t = \sigma_t dW_t^1, \tag{32}$$

where

$$d\sigma_t^2 = \kappa\left(\bar{v} - \sigma_t^2\right)dt + \omega\sigma_t dW_t^2. \tag{33}$$

Further, assuming that the logged noises η are i.i.d. Gaussian completes the specification of the observed return series as given in Eq. (2) above.

We now turn to selecting parameters for the dynamics of the price and noise processes. The parameter κ dictates persistence in volatility and is set equal to .01, a value consistent with estimates obtained from simple one-factor continuous-time models. We normalize the mean volatility to unity so that $\bar{v}$ is one. The parameter ω controls the magnitude of the volatility of volatility and is set equal to .05. Finally, the logged noises η are assumed to be normally distributed with mean zero and

a variance equal to $(.000197)^2$ (as in the previous section). This implies that the noises-in-returns ε's have a variance equal to .02829% of the average daily variance . This value is equal to .000289 since the average variance is normalized to one.

We focus our simulations around a single realization of the (daily) volatility over a period of 6.5 hours. Specifically, we simulate second-by-second a volatility path given by Eq. (33). The initial value of σ^2 is set to the unconditional mean of one. Holding the volatility path fixed, we then simulate second-by-second true returns from Eq. (32) and second-by-second observed returns as in Eq. (2). The simulations are run $1,000$ times.

Fig. 1 shows the mean realized volatility across the $1,000$ simulations for various sampling intervals ranging from 1 second to 17 minutes. The horizontal line denotes the fixed (and known) quadratic variation simulated for the day (i.e., 0.9957). Consistently with the predictions of Theorem 1, the sharp spike at zero shows the realized volatility exploding as the sampling interval goes to zero. Hence, the standard realized volatility estimator *cannot* be a consistent estimate of the quadratic variation of the underlying log price process in the presence of microstructure noise.

Fig. 2 is the same graph as Fig. 1 plotted on a different scale. This plot shows that, for the parameter values used in the simulation, the bias is small at 15 minutes and, consequently, as the sampling interval exceeds 15 minutes. At the 17 minute horizon, for instance, the value of the true quadratic variation is 0.9957 whereas the value of the (average) realized volatility estimate is about 1.006 implying a bias equal to about 1% of the underlying quadratic variation. At the 5-minute horizon the bias is around 2.5%.

Two observations are in order. First, the only dimension along which Figs. 1 and 2 (which could be regarded as simulated "volatility signature plots," using the terminology in Andersen et al. (2000)) allow us to evaluate the accuracy of the quadratic variation estimates is bias. Naturally, the optimal sampling frequency should also account for the variance of the sampling error being that the trade-off between bias and variance is apparent (see Fig. 3). The optimal choice of frequency should then balance the low bias at low frequencies with the low dispersion at high frequencies as discussed in Section 4 above.

Second, the estimated bias depends on a ratio between variance of the noise and quadratic variation equal to about .0284%. In Fig. 4 we show that the in-sample variability of the estimated (daily) ratio for IBM is substantial. Specifically, the maximum value of the ratio in our sample is about 7 times as big as the minimum value. In other words, the actual bias can be larger on days

when the ratio is higher. For clarity, we also perform simulations for a value of the ratio that is 4 times as large as in Fig. 2 (i.e., 0.088%). This number is consistent with the range of values that is reported in Fig. 4 and allows us to show what are the consequences of moving from a relatively central level of the ratio to more extreme values in the upper tail of the empirical distribution. At the 17-minute interval the bias is now about 2.8% of the true quadratic variation. The new bias at the 5-minute interval is about 8% of the underlying quadratic variation.

Recall that Theorem 1 also indicates that a rescaled version of the quadratic variation estimator should converge to the variance of the noise process. Fig. 5 represents the rescaled realized volatility, namely $\frac{\widehat{V}_t}{M}$, for different sampling frequencies and a value of the ratio between variance of the noise process and underlying quadratic variation equal to 0.0284%. As earlier, the sampling interval (in minutes) is given on the horizontal axis. The horizontal line in the plot is now the true variance of the noise process. Clearly, the rescaled realized volatility estimates converge to the second moment of the contaminations-in-return $\varepsilon^{'}$s as the sampling interval goes to zero (i.e., as the number of observations M increases without bound). We perform the same exercise with sample averages of fourth powers of the contaminated return data for a variety of sample frequencies (see Fig. 6). Again, we confirm the validity of the predictions contained in Theorem 2.

5.3 The biases in the quarticity estimates and their impact on the conditional MSE of the realized volatility estimator: simulations based on IBM.

In this subsection we show that alternative (but credible) sampling frequencies used to compute the underlying quarticity (i.e., the remaining input in Eq. (30)) have little impact on the sampling distribution of the optimal number of observations needed to calculate the object of interest, i.e., the underlying quadratic variation of the log price process. In addition, we show that, when the quarticity is estimated relatively accurately, the rule-of-thumb in Lemma 4 delivers a distribution of the estimated optimal sampling frequencies that is similar to the distribution obtained from the full minimization of the conditional MSE. Should the quarticity be estimated imprecisely, then the rule-of-thumb would deliver estimates that are more biased and considerably more volatile then those delivered by the full minimization.

In Fig. 7 we plot the empirical MSE of the realized quarticity. The minimum is around 2 minutes. Going from the $2-$minute sampling frequency to the 15-minute sampling frequency implies multiplication of the MSE by a factor of 4. Interestingly, even though the loss would be considerable

should one be just interested in the estimation of the quarticity per se, we will show that the impact of the suboptimal 15-minute frequency on the sampling distribution of the minima of the conditional MSE of the realized volatility estimator is not substantial. In light of the attention that the recent empirical literature has devoted to the 15-minute sampling interval (see Andersen et al. (2000), for instance), this observation will lead us to recommend the 15-minute sampling frequency for the quarticity estimates as a valid frequency for stocks with various degrees of liquidity (see, also, Bandi and Russell (2003b)). Naturally, as shown in Fig. 7, such choice is quite conservative for highly liquid stocks like IBM. We will return to these important remarks.

In Fig. 8 we plot the distribution (across the $1,000$ simulations) of the optimal sampling frequencies obtained by minimizing the expression in Eq. (22) for values of the quarticity estimates obtained by sampling at the correct 2-minute interval. Some observations are in order. First, despite the existence of an upward bias in the estimated values (the mean and the median are equal to 2.8 minutes while the true optimal frequency is 1.7 minutes) the range of possible values is very informative about the magnitude of the optimal frequency. For instance, the obtained range does not include the 5-minute interval that has been largely used in the empirical work on the subject. Second, the bias goes in the right direction in the sense that it provides us with a conservative assessment of the optimal sampling interval while keeping us away from high frequencies corresponding to the upward spike in the MSE of the realized volatility estimator. Finally, for the range of values in Fig. 8, the incremental impact of lowering the sample frequency on the MSE of the realized volatility estimator is rather small. The value of the MSE at the optimal 1.7-minute frequency is .014. It is 0.0143 at the 2 minute interval and 0.016 at the 3.5-minute frequency. At the 5-minute interval, the MSE value is virtually twice as large as the corresponding value at the 2-minute interval (.027). Admittedly, these considerations are conditional on choosing a frequency for the quarticity that is very close to the optimal value as suggested by the simulated MSE for the quarticity in Fig. 7.

Thus, in Fig. 9 we report the distribution of the optimal frequencies for values of the quarticity that are estimated using a 15-minute interval. The incremental bias is minimal. Additionally, while the increased variance in the quarticity estimates (as testified by the MSE in Fig. 7) translates into increased dispersion of the optimal frequencies, the array of possible values is still very informative about the range of acceptable frequencies. In other words, using an inaccurate measure of the underlying quarticity does *not* entail an uninformative characterization of the optimal sampling frequency for the object of econometric interest, namely the realized volatility estimator. As said,

employing a 15-minute frequency for the quarticity is a conservative choice. While it was shown that such choice produces informative estimates, it can certainly be improved upon. In effect, our results suggest that it believable that higher (than 15 minutes) sampling frequencies would be appropriate in the case of very liquid stocks (like IBM). Having said this, we think that a 15-minute sampling interval for the quarticity estimates provides sufficient information about the optimal sampling interval (Bandi and Russell (2003b) confirm this finding). Furthermore, it is an easy interval to use. Therefore, it deserves attention in applied work. Coherently, the 15-minute frequency is the frequency that we utilize in the empirical application in next subsection.

For completeness, we also report results for the case where the quarticity is computed using a sampling frequency equal to 30 minutes (see Fig. 10). While the increase in the bias is not considerable, the likelihood of obtaining large values is substantially higher than in the previous case. In effect, there is a non-negligible probability of obtaining optimal sampling frequencies in excess of 5 minutes. Nonetheless, the implied frequencies are still quite informative. For instance, our findings clearly rule out frequencies that have been put forward as sensible conjectures in the presence of microstructure noise in the empirical literature on the subject, namely frequencies in the vicinity of the 15-minute interval (the maximum value across the $1,000$ simulations is equal to 11.6 minutes). To conclude, even though we do not recommend using a 30-minute sampling interval for the quarticity, we find it reassuring that possibly very volatile estimates for it (as determined by very suboptimal choices of the corresponding frequency) do not cause equally suboptimal sampling frequencies for the realized volatility estimates.

In Figs. 11 through 13 we examine the impact of various quarticity measurements on the distributions on the optimal sampling frequencies for the realized volatility estimator obtained by employing the rule-of-thumb in Lemma 4. We find that the estimates are more upward biased and variable than in the case where a full minimization of the conditional MSE is performed. While these results hold across different choices of the quarticity estimates, they are particularly pronounced as we move to highly suboptimal choices of the optimal sampling frequency for the quarticity. In effect, the approximation that is provided by Lemma 4 above appears very valid when a close-to-optimal frequency for the quarticity is chosen (see Fig. 11). When using a 15-minute frequency for the quarticity, for instance, the estimates that the approximation provides are considerably more variable than in the full minimization case (since values as high as 20 minutes are possible). Nonetheless, the distribution of the resulting estimates can still be somewhat informative about the magnitude of the optimal

sampling frequency. In effect, due to the evident right skewness in the simulated distribution, the likelihood of obtaining values around 2 minutes (i.e., near the true optimal frequency) is about 50%.

In Figs. 14 and 15 we plot the true conditional MSE as implied by Eq. (22) above and corresponding 95% bands based on the simulations. In light of our previous remarks, in both cases we use the conservative 15-minute interval to estimate the underlying quarticity and quadratic variation. Coherently with the average arrival time for a new price quote for the stock IBM in the month of February 2002, in Fig. 14 we employ a 10 second sampling interval to estimate the necessary features of the noise process (i.e., the second and the fourth moment). A 1 second sampling interval for the same objects is used in Fig. 15. As expected, the graphs show that the estimated conditional MSE expansion is more accurate when using moments of the noise process that are defined on the basis of very high frequencies. This result is understandable in that higher frequencies lead to more precise estimates of the noise characteristics. Hence, stocks whose price updates occur very frequently should lead to extremely accurate MSE expansions.

5.4 Computing the optimal frequency for IBM.

In Fig. 16 we plot the estimated conditional MSE of IBM along with the minimum from the full minimization as implied by Eq. (30) and the minimum from the rule-of-thumb reported in Eq. (31). The realized quarticity is estimated conservatively (see the previous subsection) using the 15-minute sampling interval. We employ the high frequency quote-to-quote price changes to calculate the moments of the noise process.

Several observations are in order. First, the optimal sampling interval is equal to 1.5 minutes. This interval is shorter than the 5-minute interval that is used in some empirical work on the subject (see Andersen et al. (2001), for instance). Consequently, it is shorter than recent conjectures on optimal sampling based on the 15/20 minute interval (see Andersen et al. (2000)). Nonetheless, this result crucially hinges on the liquidity features of the stock. Bandi and Russell (2003b) find that less liquid stocks than IBM require lower sampling frequencies leading to optimal intervals that are in the vicinity of the 5-minute interval.

Second, the loss that is induced by suboptimal sampling depends on the slope of the conditional MSE. Going from the optimal frequency of 1.5 minutes to the 15-minute interval almost doubles the MSE. This effect is economically important since the magnitude of the MSE is large. One can easily have a feel for it by a straightforward comparison of the root MSE (as implied by the values on

the horizontal axis of Fig. 16) and the average realized volatility, namely .0273%. At the 15-minute interval the ratio between the two quantities is equal to about 36%. An alternative way to assess the importance of accounting for microstructure contaminations in returns when estimating volatility using high frequency data is to compare the magnitude of the MSE expansion to the variance that would emerge from models that do not explicitly allow for noise (BN-S (2002, 2004), for instance). In the case of IBM we find that the value of the squared bias alone is larger than the value of the variance term.

Finally, we notice that the rule-of-thumb in Lemma 4 provides an empirical answer to the optimal frequency problem that is almost indistinguishable from the answer that is provided by the full minimization of the expansion in Eq. (22). More precisely, we find that the approximate optimal frequency is 1.45. After one takes into consideration that the quality of the approximation is higher in the presence of a large number of observations M (see Section 4) and that the theoretical dispersion of the approximate estimates can be high (see the previous subsection), the rule-of-thumb can provide very useful and immediate implications for empirical work. Bandi and Russell (2003b) confirm this finding for a variety of S&P100 stocks.

6 Conclusions

Recorded prices are known to diverge from their "efficient" values due to the presence of market microstructure contaminations.

We find that the presence of market microstructure noise in high-frequency data makes the consistent estimation of the quadratic variation of the underlying log price process through the conventional realized volatility estimator unachievable. In effect, the summing of increasingly-frequent contaminated return data simply entails infinite accumulation of noise. Interestingly, we point out that a standardized version of the realized volatility estimator can be employed to consistently estimate the second moment of the unobservable noise process. More generally, we stress that sample averages of powers of contaminated returns converge to the moments of the underlying noise process.

Moving from asymptotic arguments to finite sample results, we argue that the microstructure noise creates a dichotomy in the model-free estimation of integrated volatility. While it is theoretically necessary to sum squared returns that are computed over very small intervals to better identify the underlying volatility over a period, the summing of numerous contaminated return series entails substantial accumulation of noise. Hence, the final effect is the determination of a finite sample

bias/variance trade-off. We quantify the trade-off in the presence of a realistic model of price determination and discuss optimal sampling (for the purpose of volatility estimation) on the basis of the (estimable) moments of the distribution of the noise process and the underlying quarticity of the log price process.

Specifically, in the spirit of the simplicity and generality of the conventional realized volatility estimator as a nonparametric measurement of the quadratic volatility of the underlying log price process, we derive simple implications and tools for empirical work on quadratic volatility estimation through realized volatility. We summarize our findings as follows.

(1) The optimal sampling problem can be written as the minimization of the conditional MSE expansion of the realized volatility estimator.

(2) The main ingredients of the MSE expansion are the second and the fourth moment of the (unobserved) noise process and the so-called quartic volatility of underlying log price process.

(3) We show that the moments of the underlying noise process can be estimated consistently in the presence of high frequency observations. Even though the quarticity estimator (as suggested by BN-S (2002)) does not provide a consistent nonparametric estimate of the remaining ingredient of the MSE expansion, i.e., the quartic volatility, if noise plays a role, we stress that alternative (plausible) choices of sampling frequency for it do not have a substantial impact on the optimal sampling of the realized volatility estimator.

(4) More precisely, we deem the (easy to implement) 15-minute sampling interval to be a valid (albeit conservative) choice of frequency for the quarticity estimator. Such choice can be improved upon (i.e., lowered) in the case of very liquid stocks.

(5) In addition to providing a straightforward minimization program to solve, we offer a simple rule-of-thumb to select the optimal sampling frequency on the basis of an expression that could be readily interpreted as a signal-to-noise ratio, namely the ratio between the quarticity of the log price process and the second moment of the unobserved noise.

    Despite having the property of delivering more variable estimates of the optimal frequency that the full minimization, the rule-of-thumb can provide a very useful preliminary assessment of the answer to the minimization problem. In particular, the rule-of-thumb is expected to be

very accurate when the true optimal sampling frequency is high and the quarticity is estimated accurately.

An application of our methodology to a liquid stock like IBM points to the necessity of a sampling interval that is smaller that intervals typically employed in applied work on the subject. Furthermore, our analysis provides the following conclusions:

(1) Properly accounting for market microstructure effects suggests that the MSE can be large relative to the realized volatility estimates.

(2) An MSE that fails to account for market microstructure noise can substantially understate the true MSE.

(3) Failing to sample at the optimal frequency can lead to large inefficiencies in the quality of the realized volatility estimates.

One final observation is needed. The present paper suggests a simple nonparametric technique to learn about features of the noise component in recorded high-frequency asset prices as well as about the genuine volatility properties of the efficient prices. As such, it provides tools that might prove useful in two separate strands of the finance literature, namely empirical microstructure and, of course, volatility estimation. The importance of the later is apparent. Here, we briefly expand on the former. A considerable amount of work has been devoted to understanding the determinants of the quoted bid-ask spreads. Nonetheless, it is well-known that the (average) between the bid and the ask quotes are imprecise measurements of the true cost of trade in that transactions often occur within the posted spreads. In consequence, the true quantity of interest is the so-called "effective spread," namely the difference between the transaction price and the efficient price. When using transaction prices rather than mid-point bid-ask quotes as in this paper, the methods proposed in the present piece can be used to provide nonparametric measurements of the (unconditional) distributional features of the effective spreads. Such features can be put to work to investigate the cross-sectional determinants of the implicit cost-of-trade as in Bandi and Russell (2003c) . Alternatively, one can modify the set-up that was previously discussed in order to allow for heteroskedastic structures in the noise process and study the dynamic features of the estimated microstructure frictions. The time series dynamics are currently being studied by the authors and will be reported in Bandi and Russell (2003d).

7 Appendix A: Proofs

Proof of Lemma1.

$$\begin{aligned}
\mathbf{E}(\varepsilon^2) &= \mathbf{E}(\eta-\eta_{-1})^2 && (34)\\
&= \mathbf{E}\left(\eta^2-2\eta\eta_{-1}+\eta_{-1}^2\right) && (35)\\
&= 2\sigma_\eta^2. && (36)
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\left(\varepsilon^4\right) &= \mathbf{E}\left(\eta-\eta_{-1}\right)^4 && (37)\\
&= \mathbf{E}\left(\eta^2-2\eta\eta_{-1}+\eta_{-1}^2\right)^2 && (38)\\
&= \mathbf{E}(\eta^4-2\eta^3\eta_{-1}+\eta^2\eta_{-1}^2)\\
&\quad +\mathbf{E}(-2\eta^3\eta_{-1}+4\eta^2\eta_{-1}^2-2\eta\eta_{-1}^3)\\
&\quad +\mathbf{E}(\eta^2\eta_{-1}^2-2\eta\eta_{-1}^3+\eta_{-1}^4) && (39)\\
&= 2\mathbf{E}\left(\eta^4\right)+6\sigma_\eta^4. && (40)
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) &= \mathbf{E}\left[(\eta-\eta_{-1})^2(\eta_{-1}-\eta_{-2})^2\right] && (41)\\
&= \mathbf{E}\left[(\eta^2+\eta_{-1}^2-2\eta\eta_{-1})(\eta_{-1}^2+\eta_{-2}^2-2\eta_{-1}\eta_{-2})\right] && (42)\\
&= \mathbf{E}\left[\eta^2\eta_{-1}^2+\eta^2\eta_{-2}^2-2\eta^2\eta_{-1}\eta_{-2}\right]\\
&\quad +\mathbf{E}\left[\eta_{-1}^4+\eta_{-1}^2\eta_{-2}^2-2\eta_{-1}^3\eta_{-2}\right]\\
&\quad -\mathbf{E}\left[2\eta\eta_{-1}^3+2\eta\eta_{-1}\eta_{-2}^2-4\eta\eta_{-1}^2\eta_{-2}\right] && (43)\\
&= 3\sigma_\eta^4+\mathbf{E}\left(\eta^4\right). \blacksquare && (44)
\end{aligned}$$

Proof of Theorem 1. We study the term A_i first. The interested reader is referred to BN-S (2002) for a different proof. Specifically, we generalize the classical BN-S result along two dimensions, namely we provide a functional central limit theorem-based argument and allow for leverage effects. Recall that $\frac{h}{\delta}=M$ and write

$$A_i=\sum_{j=1}^{h/\delta} r_{j,i}^2=\sum_{j=1}^{h/\delta}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s dW_s\right)^2. \tag{45}$$

and

$$L_j^i=\int_0^{(i-1)h+j\delta}\sigma_s dW_s. \tag{46}$$

Then,

$$\begin{aligned}
A_i &= \sum_{j=1}^{h/\delta}\left(L_j^i-L_{j-1}^i\right)^2 && (47)\\
&= \sum_{j=1}^{h/\delta}\left(\left(L_j^i\right)^2+\left(L_{j-1}^i\right)^2-2L_j^iL_{j-1}^i\right) && (48)\\
&= \sum_{j=1}^{h/\delta}\left(\left(L_j^i\right)^2-\left(L_{j-1}^i\right)^2-2L_{j-1}^i\left(L_j^i-L_{j-1}^i\right)\right). && (49)
\end{aligned}$$

Now, we note that

$$\left(L_j^i\right)^2=\left(L_0^i+2\int_0^{(i-1)h+j\delta}L^i dL^i\right)+\left[L^i\right]_{(i-1)h+j\delta} \tag{50}$$

by the Doob's decomposition (see Theorem 4.7 in Chung and Williams (1990), for instance). It is known that $\left(L_0^i + 2\int_0^{(i-1)h+j\delta} L^i dL^i\right)$ is a continuous local martingale whereas $\left[L^i\right]_{(i-1)h+j\delta}$ is a continuous increasing process with initial value zero, i.e., the quadratic variation of the local martingale L^i between time 0 and time $(i-1)h + j\delta$. Then,

$$A_i = \sum_{j=1}^{h/\delta} \left(2\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} L^i dL^i + \left[L^i\right]_{j-1,j} - 2L_{j-1}^i \left(L_j^i - L_{j-1}^i\right) \right), \tag{51}$$

where

$$\left[L^i\right]_{j-1,j} = \left[L^i\right]_{(i-1)h+j\delta} - \left[L^i\right]_{(i-1)h+(j-1)\delta} = \int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds. \tag{52}$$

Hence,

$$A_i = \sum_{j=1}^{h/\delta} \left[L^i\right]_{j-1,j} + 2\sum_{j=1}^{h/\delta} \int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \left(L^i - L_{j-1}^i\right) dL^i \tag{53}$$

and

$$A_i - \int_{(i-1)h}^{ih} \sigma_s^2 ds = 2\sum_{j=1}^{h/\delta} \overline{L}_{j-1,j}, \tag{54}$$

where

$$\overline{L}_{j-1,t} = \int_{(i-1)h+(j-1)\delta}^{(i-1)h+t\delta} \left(L^i - L_{j-1}^i\right) dL^i \qquad j-1 \le t \le j \tag{55}$$

is a continuous local martingale with zero mean. Now, write

$$\Phi_t^M := 2\sqrt{\frac{M}{h}} \left(\sum_{j=1}^{h/\delta - 1} \overline{L}_{j-1,j} + \int_{(i-1)h+\left(\frac{h}{\delta}-1\right)\delta}^{(i-1)h+t\delta} \left(L^i - L_{\frac{h}{\delta}-1}^i\right) dL^i \right), \tag{56}$$

with $\frac{h}{\delta} - 1 < t \le \frac{h}{\delta}$, where Φ_t^M is a continuous local martingale with (limiting) increasing process $\left[\Phi^M\right]_t$ given by

$$4\frac{M}{h}\sum_{j=1}^{h/\delta-1}\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\left(L^i-L^i_{j-1}\right)^2 d\left[L^i\right]+4\frac{M}{h}\int_{(i-1)h+\left(\frac{h}{\delta}-1\right)\delta}^{(i-1)h+t\delta}\left(L^i-L^i_{\frac{h}{\delta}-1}\right)^2 d\left[L^i\right]$$

$$\overset{a}{\approx} 4\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\left[L^i-L^i_{j-1}\right] d\left[L^i\right]+\int_{(i-1)h+\left(\frac{h}{\delta}-1\right)\delta}^{(i-1)h+t\delta}\left[L^i-L^i_{\frac{h}{\delta}-1}\right] d\left[L^i\right]\right) \tag{57}$$

$$= 4\frac{M}{h}(\sum_{j=1}^{h/\delta-1}\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\left[L^i-L^i_{j-1}\right] d\left[L^i-L^i_{j-1}\right]$$

$$+\int_{(i-1)h+\left(\frac{h}{\delta}-1\right)\delta}^{(i-1)h+t\delta}\left[L^i-L^i_{\frac{h}{\delta}-1}\right] d\left[L^i-L^i_{\frac{h}{\delta}-1}\right]) \tag{58}$$

$$= 4\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\frac{\left[L^i-L^i_{j-1}\right]^2}{2}|^j_{j-1}+\frac{\left[L^i-L^i_{\frac{h}{\delta}-1}\right]^2}{2}|^t_{\frac{h}{\delta}-1}\right) \tag{59}$$

$$= 2\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\left[L^i_j-L^i_{j-1}\right]^2+\left[L^i_t-L^i_{\frac{h}{\delta}-1}\right]^2\right) \tag{60}$$

$$= 2\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma^2 ds\right)^2+\left(\int_{(i-1)h+\left(\frac{h}{\delta}-1\right)\delta}^{(i-1)h+t\delta}\sigma^2 ds\right)^2\right). \tag{61}$$

We now define the stopping times $\tau=\inf\left\{s:\left[\Phi^M\right]_s>t\right\}$. Hence, $B_t=\Phi^M_\tau$ is a $\Im_\tau$-Brownian motion and $\Phi^M_t=B_{\left[\Phi^M\right]_t}$ is the so-called DDS (Dambis, Dubins-Schwartz) Brownian motion of Φ_t (see Revuz and Yor (Theorem 1.6, 1994)). Finally,

$$\sqrt{\frac{M}{h}}\left[A_i-\int_{(i-1)h}^{ih}\sigma^2_s ds\right] =$$

$$\Phi^M_{h/\delta}\underset{M\to\infty}{\Rightarrow} B\left(2\frac{M}{h}\sum_{j=1}^{h/\delta}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma^2_s ds\right)^2\right) \tag{62}$$

$$\overset{d}{=} B\left(2\left(\int_{(i-1)h}^{ih}\sigma^4_s ds\right)\right), \tag{63}$$

where $Q_i=\int_{(i-1)h}^{ih}\sigma^4_s ds$ is the quartic variation (over h) as defined in BN-S (2002), by the asymptotic Knight's theorem (see Revuz and Yor (Theorem 2.3, 1994)). Under Assumption 1(3), i.e., absence of leverage effects, we can write

$$B\left(2\left(\int_{(i-1)h}^{ih}\sigma^4_s ds\right)\right)\overset{d}{=}\mathbf{MN}\left(2\left(\int_{(i-1)h}^{ih}\sigma^4_s ds\right)\right). \tag{64}$$

This proves the result in Remark 1. We now study the term B_i, that is

$$\sum_{j=1}^{h/\delta}\varepsilon^2_{j,i}. \tag{65}$$

A conventional central limit theorem for stationary mixing sequences (see Hamilton (1994), for example) allows us to write

$$\sqrt{M}\left(\frac{\sum_{j=1}^{h/\delta}\varepsilon_{j,i}^2}{M}-\mathbf{E}\left(\varepsilon^2\right)\right)\underset{M\to\infty}{\Rightarrow}\mathbf{N}\left(0,\mathbf{V}\left(\frac{1}{\sqrt{M}}\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)\right), \tag{66}$$

where

$$\begin{aligned}
&\mathbf{V}\left(\frac{1}{\sqrt{M}}\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)\\
=\;&\frac{1}{M}\mathbf{E}\left(\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)^2\right)-\frac{1}{M}\left(\mathbf{E}\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)\right)^2 &(67)\\
=\;&\frac{1}{M}\mathbf{E}\left(\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\left(\varepsilon_{g,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)\right) &(68)\\
=\;&\frac{1}{M}\left(M\mathbf{E}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)^2+2(M-1)\mathbf{E}\left(\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\left(\varepsilon_{j-1,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)\right) &(69)\\
\overset{a}{\sim}\;&\mathbf{E}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)^2+2\mathbf{E}\left(\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\left(\varepsilon_{j-1,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right), &(70)
\end{aligned}$$

with

$$\mathbf{V}(\varepsilon^2)=\mathbf{E}\left(\varepsilon^2-\mathbf{E}\left(\varepsilon^2\right)\right)^2=\mathbf{E}\left(\varepsilon^4\right)-\left(\mathbf{E}\left(\varepsilon^2\right)\right)^2 \tag{71}$$

and

$$2\mathbf{E}\left(\left(\varepsilon^2-\mathbf{E}\left(\varepsilon^2\right)\right)\left(\varepsilon_{-1}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)=2\mathbf{E}\left(\varepsilon^2\varepsilon_{-1}^2\right)-2\left(\mathbf{E}\left(\varepsilon^2\right)\right)^2. \tag{72}$$

Then,

$$\mathbf{V}\left(\frac{1}{\sqrt{M}}\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)\underset{M\to\infty}{\to}\mathbf{E}\left(\varepsilon^4\right)+2\mathbf{E}\left(\varepsilon^2\varepsilon_{-1}^2\right)-3\left(\mathbf{E}\left(\varepsilon^2\right)\right)^2. \tag{73}$$

Thus, Lemma 1 implies that

$$\begin{aligned}
&\mathbf{V}\left(\frac{1}{\sqrt{M}}\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2-\mathbf{E}\left(\varepsilon^2\right)\right)\right)\\
&\underset{M\to\infty}{\to}2\mathbf{E}\left(\eta^4\right)+6\sigma_\eta^4+6\sigma_\eta^4+2\mathbf{E}\left(\eta^4\right)-12\sigma_\eta^4=4\mathbf{E}\left(\eta^4\right). &(74)
\end{aligned}$$

This proves the result in Remark 2. We now turn to term C_i, namely

$$\sum_{j=1}^{h/\delta}r_{j,i}\varepsilon_{j,i}. \tag{75}$$

Using the same notation as earlier in the case of the quantity A_i, we write

$$\begin{aligned}
C_i\;&=\;\sum_{j=1}^{h/\delta}\left(L_j^i-L_{j-1}^i\right)\varepsilon_{j,i} &(76)\\
&=\;\sum_{j=1}^{h/\delta}\left(L_j^i-L_{j-1}^i\right)\left(\eta_{j,i}-\eta_{j-1,i}\right). &(77)
\end{aligned}$$

The objects $\left(\left(L_j^i-L_{j-1}^i\right)\varepsilon_{j,i},\Im_{j-1}\right)$ are martingale difference sequences with finite second moments. In fact,

$$\mathbf{E}\left(\left(L_j^i - L_{j-1}^i\right)\varepsilon_{j,i}|\Im_{j-1}\right) = \mathbf{E}\left(\left(L_j^i - L_{j-1}^i\right)|\Im_{j-1}\right)\mathbf{E}\left(\varepsilon_{j,i}|\Im_{j-1}\right) = 0 \qquad \forall i,j, \tag{78}$$

and

$$\begin{aligned}\mathbf{E}\left(\left(L_j^i - L_{j-1}^i\right)^2 \varepsilon_{j,i}^2|\Im_{j-1}\right) &= \mathbf{E}\left(\left(L_j^i - L_{j-1}^i\right)^2|\Im_{j-1}\right)\mathbf{E}\left(\varepsilon_{j,i}^2|\Im_{j-1}\right) &(79)\\ &= \left(\eta_{j-1,i}^2 + \sigma_\eta^2\right)\mathbf{E}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right) < \infty \qquad \forall i,j, &(80)\end{aligned}$$

given Assumptions 1(1) and 1(4). Now, write

$$\begin{aligned}&\sum_{j=1}^{k}\left(L_j^i - L_{j-1}^i\right)\left(\eta_{j,i} - \eta_{j-1,i}\right) = \sum_{j=1}^{[rh/\delta]}\left(L_j^i - L_{j-1}^i\right)\left(\eta_{j,i} - \eta_{j-1,i}\right) &(81)\\ =& \sum_{j=1}^{[rh/\delta]-1}\left(L_j^i - L_{j-1}^i\right)\left(\eta_{j,i} - \eta_{j-1,i}\right) - \eta_{[rh/\delta]-1,i}\left(L_{[rh/\delta]}^i - L_{[rh/\delta]-1}^i\right)\\ &+\eta_{[rh/\delta],i}\left(L_{[rh/\delta]}^i - L_{[rh/\delta]-1}^i\right) &(82)\\ =& \Psi^M(r) + \eta_{[rh/\delta],i}\left(\int_{(i-1)h+([rh/\delta]-1)\delta}^{(i-1)h+[rh/\delta]\delta}\sigma_s dW_s\right) &(83)\\ =& \Psi^M(r) + O_p\left(\sqrt{\delta}\right), &(84)\end{aligned}$$

with $\frac{k}{h/\delta} \leq r < \frac{k+1}{h/\delta}$. Following previous arguments (see the proof of Eq. (63) above), the local martingale $\Psi^M(r)$ can be embedded in a Brownian motion with (limiting) quadratic variation process given by

$$\begin{aligned}\left[\Psi^M\right]_r &= 2\mathbf{E}\left(\eta^2\right)\sum_{j=1}^{[rh/\delta]-1}\left(\left[L^i\right]_j - \left[L^i\right]_{j-1}\right) + \mathbf{E}\left(\eta^2\right)\left(\left[L^i\right]_{[rh/\delta]} - \left[L^i\right]_{[rh/\delta]-1}\right) &(85)\\ &\overset{a}{\sim} 2\mathbf{E}\left(\eta^2\right)\sum_{j=1}^{[rh/\delta]}\left(\left[L^i\right]_j - \left[L^i\right]_{j-1}\right) &(86)\\ &= 2\mathbf{E}\left(\eta^2\right)\left(\left[L^i\right]_{[rh/\delta]} - \left[L^i\right]_0\right) &(87)\\ &= 2\mathbf{E}\left(\eta^2\right)\left(\int_{(i-1)h}^{(i-1)h+[rh/\delta]\delta}\sigma_s^2 ds\right). &(88)\end{aligned}$$

Thus,

$$\Psi^M(r) \underset{M\to\infty}{\Rightarrow} B\left(0, \left[\Psi^M\right]_r\right), \tag{89}$$

which implies

$$C_i \overset{a}{\sim} \Psi^M(1) \underset{M\to\infty}{\Rightarrow} \mathbf{MN}\left(0, 2\mathbf{E}\left(\eta^2\right)\left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)\right). \tag{90}$$

if Assumption 1(3), i.e., absence of leverage, is satisfied. This proves the result in Remark 3 and the stated result in Theorem 1.■

Proof of Theorem 2. Recall tha $\frac{h}{\delta} = M$ and write

$$\sum_{j=1}^{h/\delta} \widetilde{r}_{j,i}^4 = \sum_{j=1}^{h/\delta} (r_{j,i} + \varepsilon_{j,i})^4 \tag{91}$$

$$= \underbrace{\sum_{j=1}^{h/\delta} r_{j,i}^4}_{A_i} + \underbrace{4\sum_{j=1}^{h/\delta} r_{j,i}^3 \varepsilon_{j,i}}_{B_i} + \underbrace{6\sum_{j=1}^{h/\delta} r_{j,i}^2 \varepsilon_{j,i}^2}_{C_i} + \underbrace{4\sum_{j=1}^{h/\delta} r_{j,i} \varepsilon_{j,i}^3}_{D_i} + \underbrace{\sum_{j=1}^{h/\delta} \varepsilon_{j,i}^4}_{E_i}. \tag{92}$$

Now, we note that

$$\frac{A_i}{M} \leq \frac{\left(\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\right)\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\right)\right)}{M} \tag{93}$$

$$= \frac{\left(\int_{(i-1)h}^{ih} \sigma_s^2 ds + 2\sum_{j=1}^{h/\delta} \overline{L}_{j-1,j}\right)^2}{M}, \tag{94}$$

where the local martingale with mean zero

$$\overline{L}_{j-1,j} = \int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \left(L^i - L^i_{j-1}\right) dL^i, \tag{95}$$

was defined in the proof of Theorem 1. Then,

$$\frac{A_i}{M} \leq \frac{\left(\int_{(i-1)h}^{ih} \sigma_s^2 ds\right)^2 + \left(2\sum_{j=1}^{h/\delta} \overline{L}_{j-1,j}\right)^2 + 2\left(\int_{(i-1)h}^{ih} \sigma_s^2 ds\right)\left(2\sum_{j=1}^{h/\delta} \overline{L}_{j-1,j}\right)}{M} \tag{96}$$

$$\leq \frac{1}{M}\left[O_p(1) + O_p\left(\sqrt{\frac{1}{M}}\right) O_p\left(\sqrt{\frac{1}{M}}\right) + O_p(1)\, O_p\left(\sqrt{\frac{1}{M}}\right)\right] \underset{M\to\infty}{\overset{p}{\to}} 0, \tag{97}$$

using Eq. (63) above and Assumption 1(4). Now write

$$\frac{C_i}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} r_{j,i}^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} \varepsilon_{j,i}^4\right)^{1/2}}{M} \tag{98}$$

$$\leq O_p\left(\sqrt{\frac{1}{M}}\right)\left[\left(\mathbf{E}\left(\varepsilon^4\right)\right)^{1/2} + O_p\left(\frac{1}{\sqrt{M}}\right)\right] \underset{M\to\infty}{\overset{p}{\to}} 0, \tag{99}$$

where the inequality in Eq. (98) follows from the Cauchy-Schwartz inequality and the inequality in Eq. (99) derives from Eq. (97) and a straightforward application of the delta method. Now consider term B_i/M.

$$\frac{B_i}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} r_{j,i}^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} r_{j,i}^2 \varepsilon_{j,i}^2\right)^{1/2}}{M} \tag{100}$$

$$\leq O_p\left(\sqrt{\frac{1}{M}}\right) O_p\left(\frac{1}{M^{1/4}}\right) \underset{M\to\infty}{\overset{p}{\to}} 0, \tag{101}$$

where the inequality in Eq. (100) follows again from the Cauchy-Schwartz inequality and the inequality in Eq. (101) derives from Eq. (99) above. We now turn to the term D_i/M and show convergence to zero in probability using a similar approach:

$$\frac{D_i}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} \varepsilon_{j,i}^4\right)^{1/2} \left(\sum_{j=1}^{h/\delta} r_{j,i}^2 \varepsilon_{j,i}^2\right)^{1/2}}{M} \tag{102}$$

$$= \left(\left(\mathbf{E}\left(\varepsilon^4\right)\right)^{1/2} + O_p\left(\frac{1}{\sqrt{M}}\right)\right) O_p\left(\frac{1}{M^{1/4}}\right) \underset{M\to\infty}{\overset{p}{\to}} 0. \tag{103}$$

Finally,

$$\frac{E_i}{M} \underset{M\to\infty}{\overset{p}{\to}} \mathbf{E}\left(\varepsilon^4\right),$$

using standard techniques for stationary mixing sequences (see Hamilton (1994)) given Assumption 2(1). This proves the stated result.■

Proof of Theorem 3. Recall that $\frac{h}{\delta} = M$. We expand Eq. (21) and obtain

$$\mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} (r_{j,i} + \varepsilon_{j,i})^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds\right)^2$$

$$= \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \left(r_{j,i}^2 + \varepsilon_{j,i}^2 + 2 r_{j,i}\varepsilon_{j,i}\right) - \int_{(i-1)h}^{ih} \sigma_s^2 ds\right)^2 \tag{104}$$

$$= \underbrace{\mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} r_{j,i}^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds\right)^2}_{A_i} + \underbrace{\mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \left(\varepsilon_{j,i}^2 + 2 r_{j,i}\varepsilon_{j,i}\right)\right)^2}_{B_i}$$

$$\underbrace{+2\mathbf{E}_\sigma \left(\left(\sum_{j=1}^{h/\delta} r_{j,i}^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds\right)\left(\sum_{j=1}^{h/\delta} \left(\varepsilon_{j,i}^2 + 2 r_{j,i}\varepsilon_{j,i}\right)\right)\right)}_{C_i}. \tag{105}$$

We start with B_i.

$$
\begin{aligned}
& \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \left(\varepsilon_{j,i}^2 + 2 r_{j,i} \varepsilon_{j,i} \right) \right)^2 \\
= \quad & \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \sum_{g=1}^{h/\delta} \left(\varepsilon_{j,i}^2 + 2 r_{j,i} \varepsilon_{j,i} \right) \left(\varepsilon_{g,i}^2 + 2 r_{g,i} \varepsilon_{g,i} \right) \right) && (106) \\
= \quad & \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \sum_{g=1}^{h/\delta} \varepsilon_{j,i}^2 \varepsilon_{g,i}^2 + 2 \sum_{j=1}^{h/\delta} \sum_{g=1}^{h/\delta} \varepsilon_{j,i}^2 r_{g,i} \varepsilon_{g,i} + 2 \sum_{j=1}^{h/\delta} \sum_{g=1}^{h/\delta} r_{j,i} \varepsilon_{j,i} \varepsilon_{g,i}^2 + 4 \sum_{j=1}^{h/\delta} \sum_{g=1}^{h/\delta} r_{j,i} \varepsilon_{j,i} r_{g,i} \varepsilon_{g,i} \right) \\
& && (107) \\
= \quad & \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \sum_{g=1}^{h/\delta} \varepsilon_{j,i}^2 \varepsilon_{g,i}^2 \right) + 4 \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \sum_{g=1}^{h/\delta} r_{j,i} \varepsilon_{j,i} r_{g,i} \varepsilon_{g,i} \right) && (108) \\
= \quad & \sum_{j=1}^{h/\delta} \mathbf{E} \left(\varepsilon_{j,i}^4 \right) + 2 \sum_{g=1}^{h/\delta} \sum_{j<g} \mathbf{E} \left(\varepsilon_{j,i}^2 \varepsilon_{g,i}^2 \right) \\
& + 4 \left(\sum_{j=1}^{h/\delta} \mathbf{E}_\sigma \left(r_{j,i}^2 \varepsilon_{j,i}^2 \right) \right) + 8 \sum_{g=1}^{h/\delta} \sum_{j<g} \mathbf{E}_\sigma (r_{j,i} \varepsilon_{j,i} r_{g,i} \varepsilon_{g,i}) && (109) \\
= \quad & \frac{h}{\delta} \mathbf{E}(\varepsilon^4) + 2 \left(\frac{h}{\delta} - 1 \right) \mathbf{E} \left(\varepsilon_i^2 \varepsilon_{-1,i}^2 \right) + \left(\left(\frac{h}{\delta} \right)^2 - 3 \frac{h}{\delta} + 2 \right) \left(\mathbf{E}(\varepsilon^2) \right)^2 \\
& + 4 \mathbf{E}(\varepsilon^2) V_i && (110) \\
= \quad & \left(\frac{h}{\delta} \right)^2 \left(\mathbf{E}(\varepsilon^2) \right)^2 + \frac{h}{\delta} \left(\mathbf{E} \left(\varepsilon^4 \right) + 2 \mathbf{E} \left(\varepsilon_i^2 \varepsilon_{-1,i}^2 \right) - 3 \left(\mathbf{E}(\varepsilon^2) \right)^2 \right) \\
& + 4 \mathbf{E}(\varepsilon^2) V_i - 2 \mathbf{E} \left(\varepsilon_i^2 \varepsilon_{-1,i}^2 \right) + 2 \left(\mathbf{E}(\varepsilon^2) \right)^2 && (111) \\
= \quad & \frac{h^2}{\delta^2} \alpha + \frac{h}{\delta} \beta + \gamma, && (112)
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha &= \left(\mathbf{E}(\varepsilon^2) \right)^2, && (113) \\
\beta &= \mathbf{E} \left(\varepsilon^4 \right) + 2 \mathbf{E}(\varepsilon^2 \varepsilon_{-1}^2) - 3 \left(\mathbf{E}(\varepsilon^2) \right)^2, && (114)
\end{aligned}
$$

and

$$
\gamma = 4 \mathbf{E}(\varepsilon^2) V_i - 2 \mathbf{E}(\varepsilon^2 \varepsilon_{-1}^2) + 2 \left(\mathbf{E}(\varepsilon^2) \right)^2. \tag{115}
$$

We recall that

$$
\sum_{j=1}^{h/\delta} \mathbf{E}_\sigma (r_{j,i}^2) = \sum_{j=1}^{h/\delta} \left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds \right) = V_i. \tag{116}
$$

This result is used in Eq. (110) above as well as in Eq. (118) below. Now consider C_i.

$$
\begin{aligned}
& 2\left(\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds\right)\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2 + 2r_{j,i}\varepsilon_{j,i}\right)\right)\right) \\
= \quad & 2\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\right)\mathbf{E}\left(\sum_{g=1}^{h/\delta}\varepsilon_{g,i}^2\right) + 4\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta} r_{j,i}^2 r_{g,i}\varepsilon_{g,i}\right) - 2V_i\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2 + 2r_{j,i}\varepsilon_{j,i}\right)\right) \qquad (117) \\
= \quad & 2\frac{h}{\delta}V_i\mathbf{E}(\varepsilon^2) - 2\frac{h}{\delta}V_i\mathbf{E}(\varepsilon^2) = 0. \qquad (118)
\end{aligned}
$$

We now turn to A_i. Write

$$
\begin{aligned}
& \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^2 - \int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \\
= \quad & \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\right)^2 - 2\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\right)\left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right) + \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (119) \\
= \quad & \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta} r_{j,i}^2 r_{g,i}^2\right) - \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (120) \\
= \quad & \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_\sigma\left(r_{j,i}^2 r_{g,i}^2\right) - \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (121) \\
= \quad & \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_\sigma\left(r_{j,i}^2\right)\mathbf{E}_\sigma\left(r_{g,i}^2\right) - \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (122) \\
= \quad & \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} r_{j,i}^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)\left(\int_{(i-1)h+(g-1)\delta}^{(i-1)h+g\delta}\sigma_s^2 ds\right) \\
& - \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (123) \\
= \quad & \sum_{j=1}^{h/\delta}\mathbf{E}_\sigma\left(r_{j,i}^4\right) - \sum_{j=1}^{h/\delta}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)^2 \qquad (124) \\
= \quad & \sum_{j=1}^{h/\delta}\mathbf{V}_\sigma\left(r_{j,i}^2\right) + \sum_{j=1}^{h/\delta}\left(\mathbf{E}_\sigma\left(r_{j,i}^2\right)\right)^2 - \sum_{j=1}^{h/\delta}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)^2 \qquad (125) \\
= \quad & \sum_{j=1}^{h/\delta}\mathbf{V}_\sigma\left(r_{j,i}^2\right) \qquad (126)
\end{aligned}
$$

It is noted that, conditionally on the volatility path $\{\sigma_s\}_{s\in((i-1)h,ih)}$, the quantity $r_{j,i}$ is Gaussian. Hence,

$$
\frac{r_{j,i}^2}{\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)} \qquad (127)
$$

is conditionally Chi-squared with one degree of freedom. Then,

$$\sum_{j=1}^{h/\delta} \mathbf{V}_\sigma \left(r_{j,i}^2\right) = \sum_{j=1}^{h/\delta} \mathbf{V}_\sigma \left(\frac{r_{j,i}^2}{\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds\right)} \left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds\right)\right) \tag{128}$$

$$= \sum_{j=1}^{h/\delta} 2 \left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds\right)^2 \tag{129}$$

$$= 2\frac{h}{M} \left(\frac{M}{h} \sum_{j=1}^{h/\delta} \left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds\right)^2 \right) \tag{130}$$

$$= 2\frac{h}{M} \left(Q_i + o_{a.s.}(1)\right), \tag{131}$$

where almost sure convergence to the quarticity follows from an argument in BN-S (2002). Finally,

$$\mathbf{E}_\sigma \left(\widehat{V}_i - V_i\right)^2 = 2\delta \left(Q_i + o_{a.s}(1)\right) + \frac{h}{\delta}\beta + \frac{h^2}{\delta^2}\alpha + \gamma, \tag{132}$$

where α, β, and γ where defined earlier.■

8 Appendix B: Notation

$\xrightarrow{p}$	convergence in probability
$\xrightarrow{a.s.}$	almost sure convergence
$\Rightarrow$	weak convergence
$:=$	definitional equality
$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
$o_{a.s.}(1)$	tends to zero almost surely
$O_{a.s.}(1)$	bounded almost surely
$\stackrel{d}{=}$	distributional equivalence
$\stackrel{a}{\sim}$	asymptotically equivalent to
$\sim$	approximately equivalent to
$[x]$	largest integer that is less than or equal to x
$\mathbf{MN}(0, \mathbf{V})$	mixed normal distribution with variance $\mathbf{V}$

References

[1] Aït-Sahalia, Y., and P. Mykland (2003). How often to sample a continuous-time process in the presence of market microstructure noise. *NBER working paper, March 2003.*

[2] Andersen, T.G., T. Bollerslev, and F.X. Diebold (2002). Parametric and nonparametric measurements of volatility. In Y. Aït-Sahalia and L.P. Hansen (Eds.) *Handbook of Financial Econometrics.* Elsevier North-Holland. Forthcoming.

[3] Andersen, T.G., T. Bollerslev, and N. Meddahi (2003b). Correcting the errors: volatility forecast evaluation based on high frequency data and realized volatilities. *Working paper, Northwestern University, Duke University and Universite' de Montreal.*

[4] Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics,* 61, 43-76.

[5] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2000). Great Realizations. *Risk Magazine,* 105-108.

[6] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003a). Modeling and forecasting realized volatility. *Econometrica,* 71, 579-625.

[7] Bai, X., J.R. Russell, and G. Tiao (2000). Effects on non-normality and dependence on the precision of variance estimates using high-frequency data. *Unpublished paper, GSB - University of Chicago.*

[8] Bandi, F.M., and B. Perron (2001). Long memory and the relation between realized and implied volatility. *Working paper, GSB - University of Chicago and Universite' de Montreal.*

[9] Bandi, F.M., and P.C.B. Phillips (2002). Nonstationary continuous-time models. In Y. Aït-Sahalia and L.P. Hansen (Eds.) *Handbook of Financial Econometrics.* Elsevier, North-Holland. Forthcoming.

[10] Bandi, F.M., and J.R. Russell (2003b). Separating market microstructure noise from volatility. *Working paper, GSB - University of Chicago.*

[11] Bandi, F.M., and J.R. Russell (2003c). Measuring and explaining the cross-sectional variation of the effective spreads. *Working paper, GSB - University of Chicago.*

[12] Bandi, F.M., and J.R. Russell (2003d). Measuring and forecasting liquidity using effective spreads. *Working paper, GSB - University of Chicago.*

[13] Barndorff-Nielsen, O.E., and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their use in financial economics. *Journal of the Royal Statistical Society, Series B, 63, 167-241.*

[14] Barndorff-Nielsen, O.E., and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B, 64, 253-280.*

[15] Barndorff-Nielsen, O.E., and N. Shephard (2003). How accurate is the asymptotic approximation to the distribution of realised variance? In Donald W.F. Andrews, James L. Powel, Paul L. Ruud, and James H. Stock (Eds.) *Indentification and Inference for Econometric Models (festschrift for Thomas J. Rothenberg).* Cambridge University Press. Forthcoming.

[16] Barndorff-Nielsen, O.E., and N. Shephard (2004). Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica, forthcoming.*

[17] Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997). *The econometrics of Financial Markets.* Princeton University Press.

[18] Cohen, K.J., S.F. Mayer, R.A. Schwartz, and D.K. Whitcomb (1979). On the existence of serial correlation in efficient securities markets. In E.J. Elton and M.J. Gruber (Eds.) *Portfolio theory, 25 years after: Essays in Honor of Harry Markowitz.* Elsevier North-Holland. Forthcoming.

[19] Chung, K.L., and R.J. Williams (1990). *Introduction to Stochastic Integration.* Second Edition. Birkhäuser.

[20] Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

[21] Madhavan, A. (2000). Market Microstructure: A Survey. *Working paper, Marshall School of Business, USC.*

[22] Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics,* 17, 475-508.

[23] Niederhoffer, V., and M.F.M. Osborne (1966). Market making and reversals on the stock exchange. *Journal of the American Statistical Association,* 61, 897-916.

[24] Ohanissian, A., J.R. Russell, and R. Tsay (2003). Using temporal aggregation to distinguish between true and spurious long memory. *Working paper, GSB - University of Chicago.*

[25] Oomen, R.C.A. (2002). Statistical models for high frequency security prices. *Working paper, Warwick Business School.*

[26] Roll, R. (1984). A simple measure of the effective bid-ask spread in an efficient market. *Journal of Finance,* 39, 1127-1139.

[27] Revuz, D., and M. Yor (1994). *Continuous Martingales and Brownian Motion.* Third Edition, Springer-Verlag, New York.

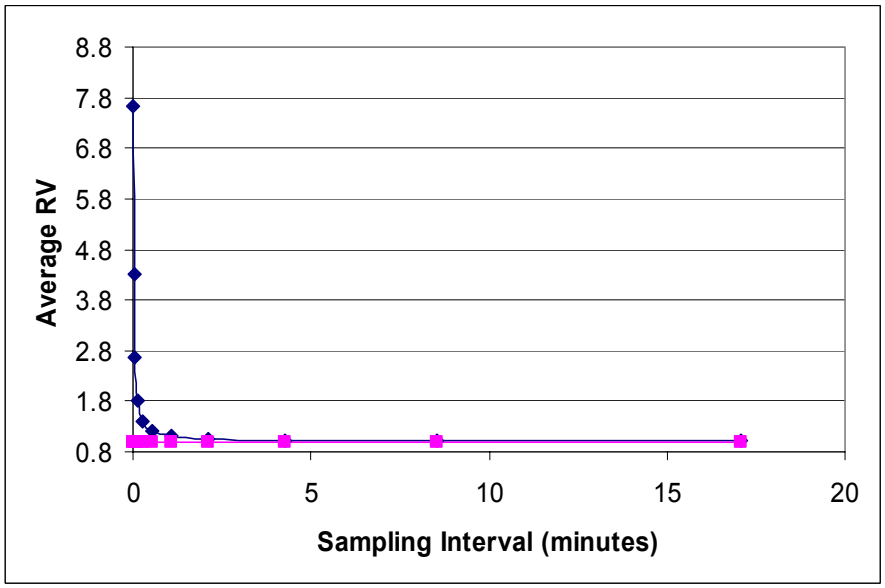

Figure 1. We plot the average realized volatility across the 1,000 simulations described in Section 5. The horizontal line denotes the known quadratic variation for the day (0.9957).

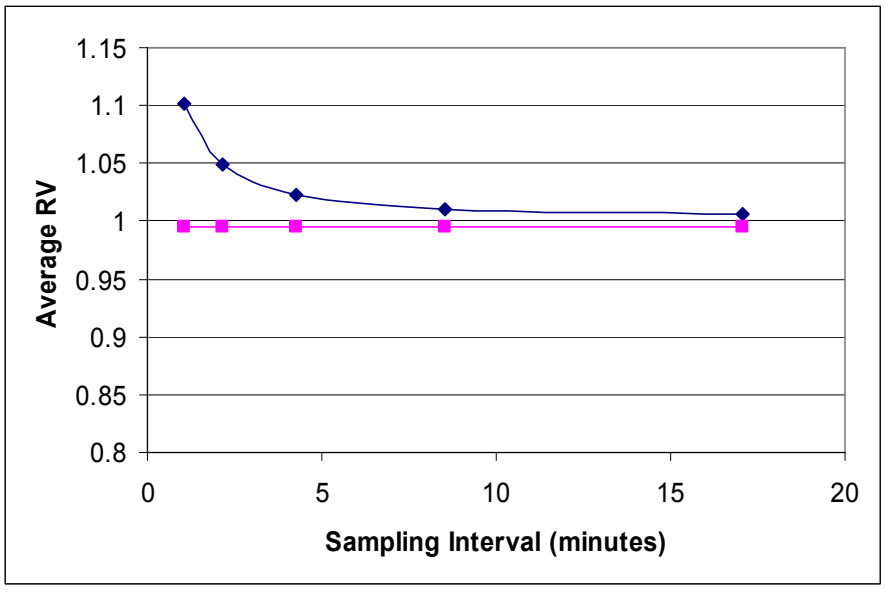

Figure 2. We plot the average realized volatility across the 1,000 simulations described in Section 5. The horizontal line denotes the known quadratic variation for the day (0.9957). This plot is the same as the plot in Figure 1 but on a different scale.

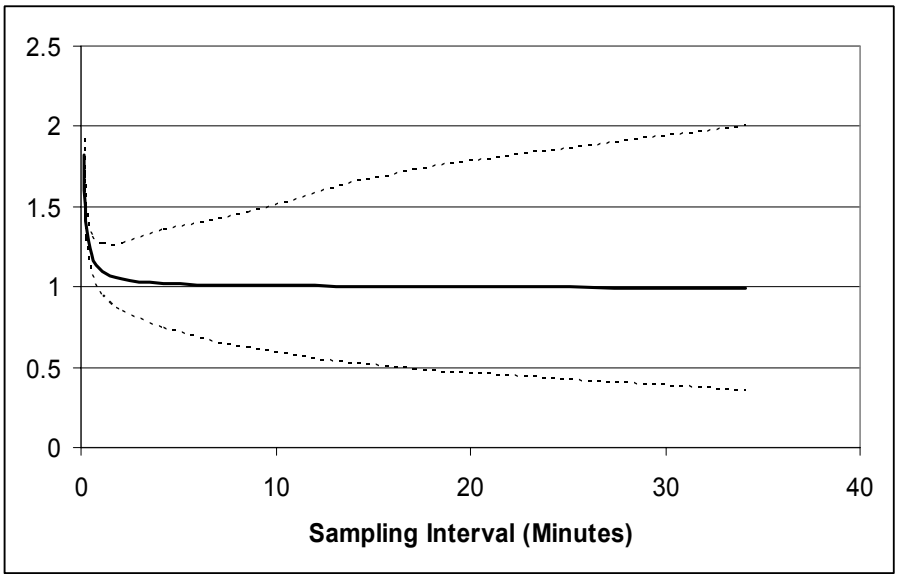

Figure 3. We plot the average realized volatility and 95% empirical bands computed across the 1,000 simulations described in Section 5.

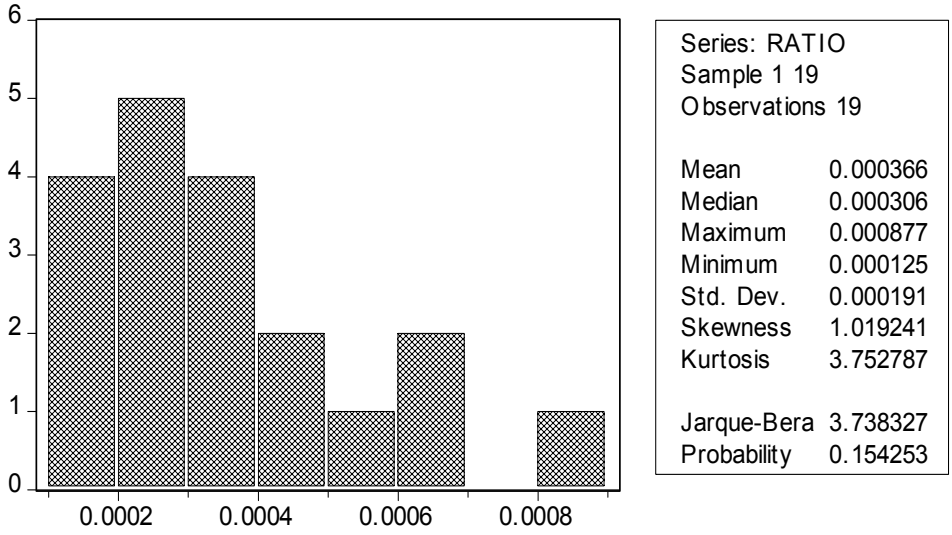

Figure 4. We plot the histogram of the ratio between the estimated second moment of the noise process and the daily realized volatilities (computed over a 6.5 hour period). The table contains the corresponding descriptive statistics.

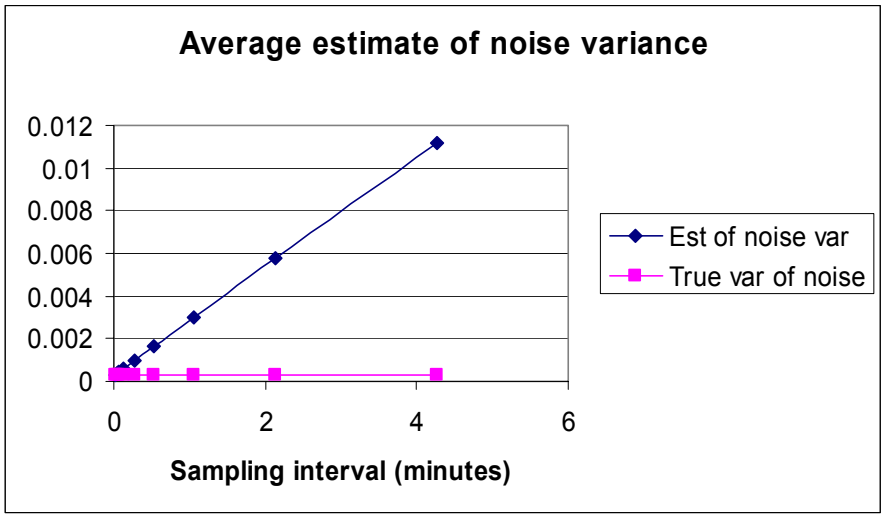

Figure 5. We plot the average of the standardized (by the number of observations) realized volatility estimates across the 1,000 simulations described in Section 5. The horizontal line denotes the known second moment of the noise process.

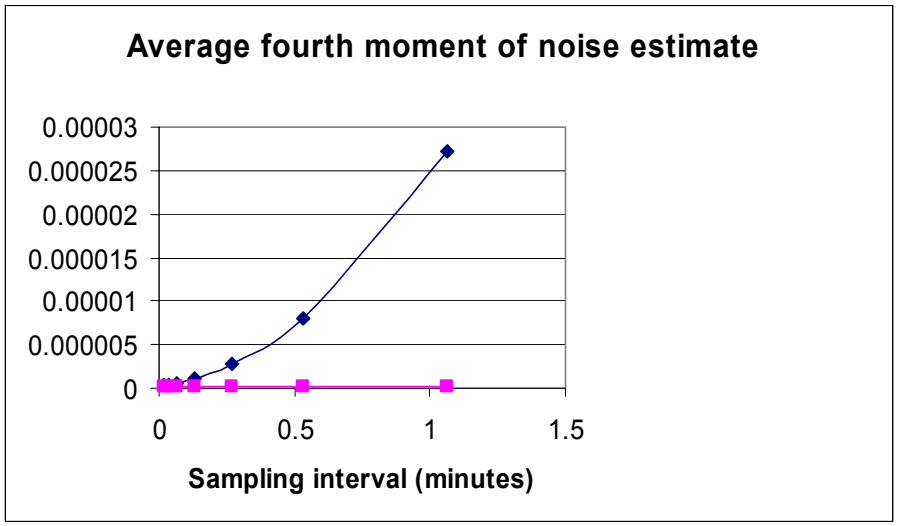

Figure 6. We plot the average of the fourth moment estimates across the 1,000 simulations described in Section 5. The horizontal line denotes the known fourth moment of the noise process.

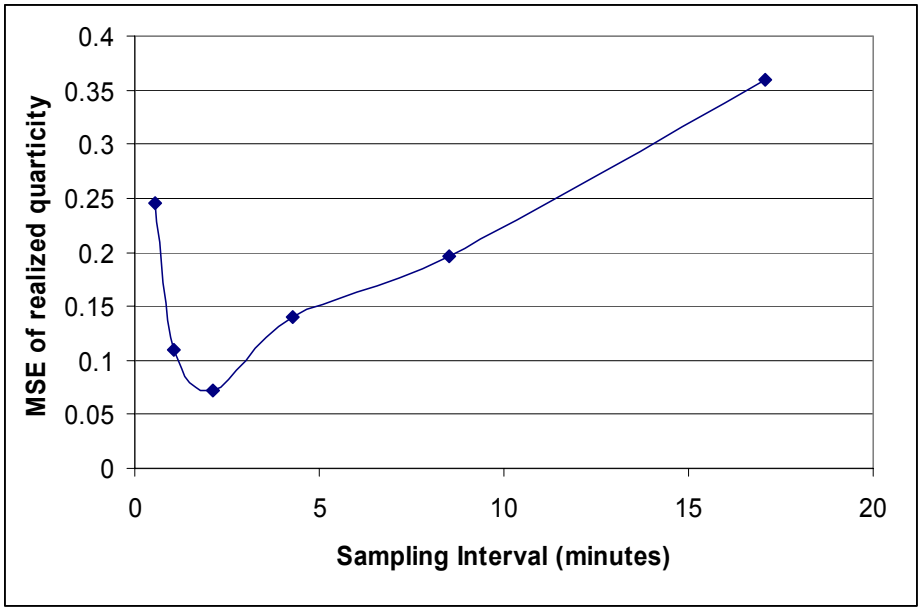

Figure 7. We plot the simulated conditional mean-squared error of the realized quarticity. We use 1,000 simulations as described in Section 5.

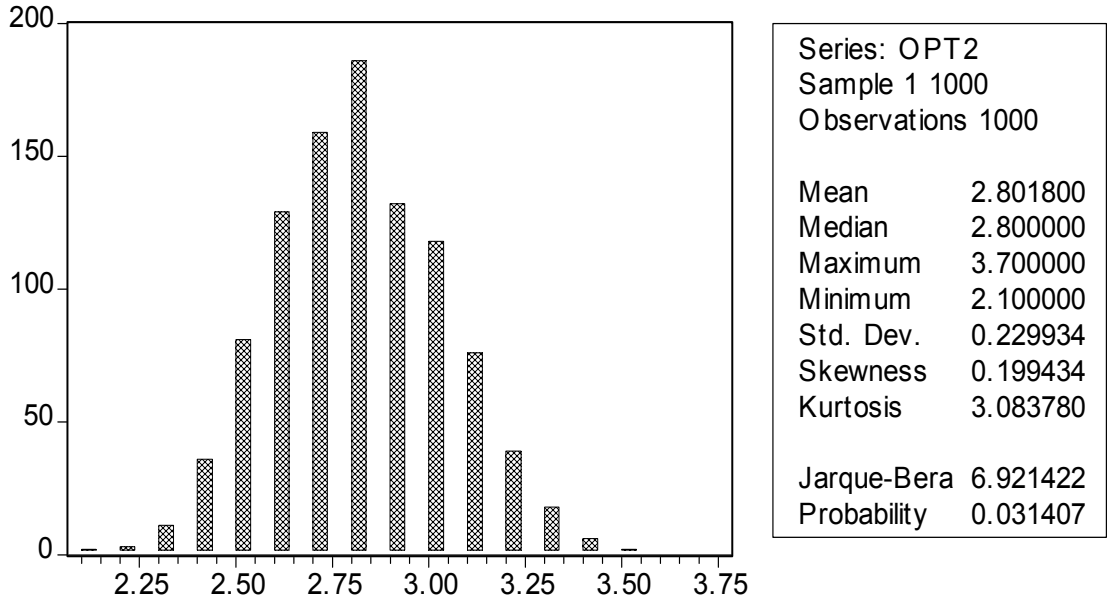

Figure 8. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 2 minute sampling interval. The table contains the corresponding descriptive statistics.

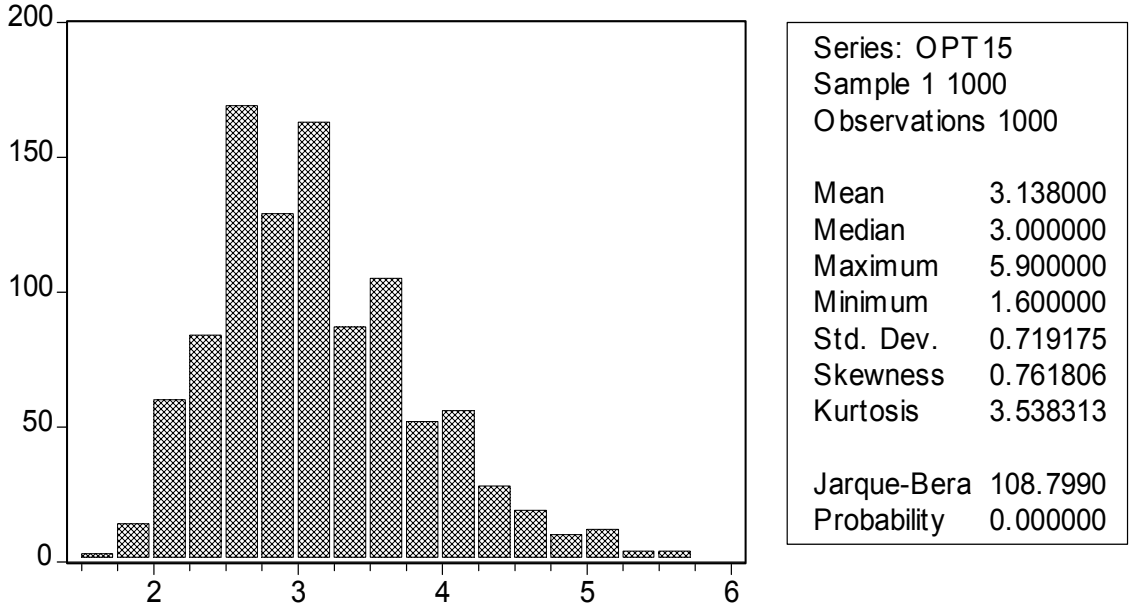

Figure 9. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The table contains the corresponding descriptive statistics.

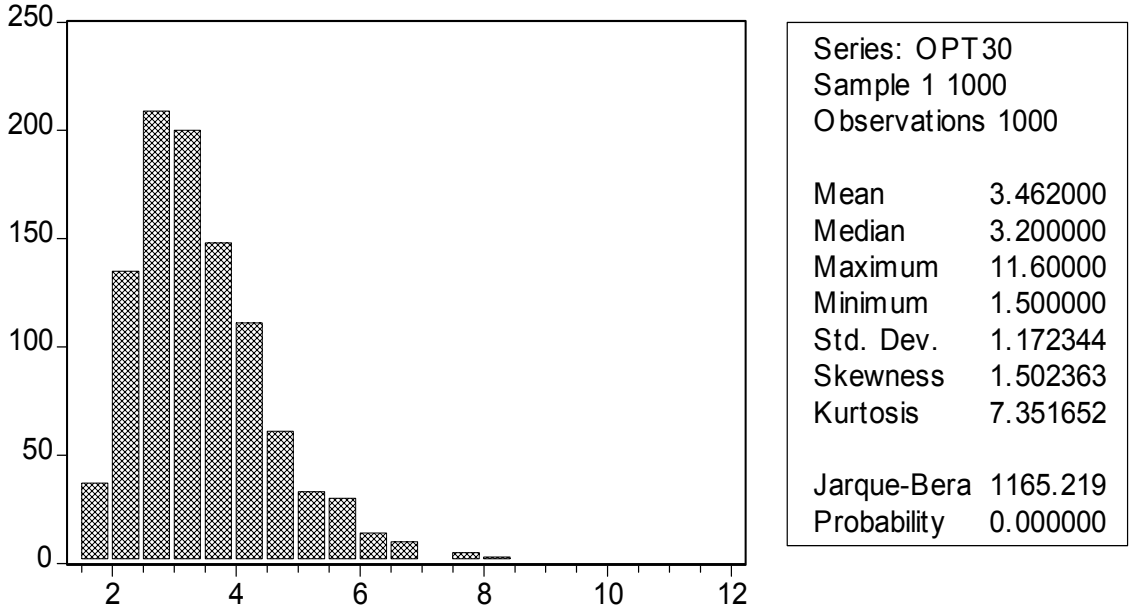

Figure 10. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 30 minute sampling interval. The table contains the corresponding descriptive statistics.

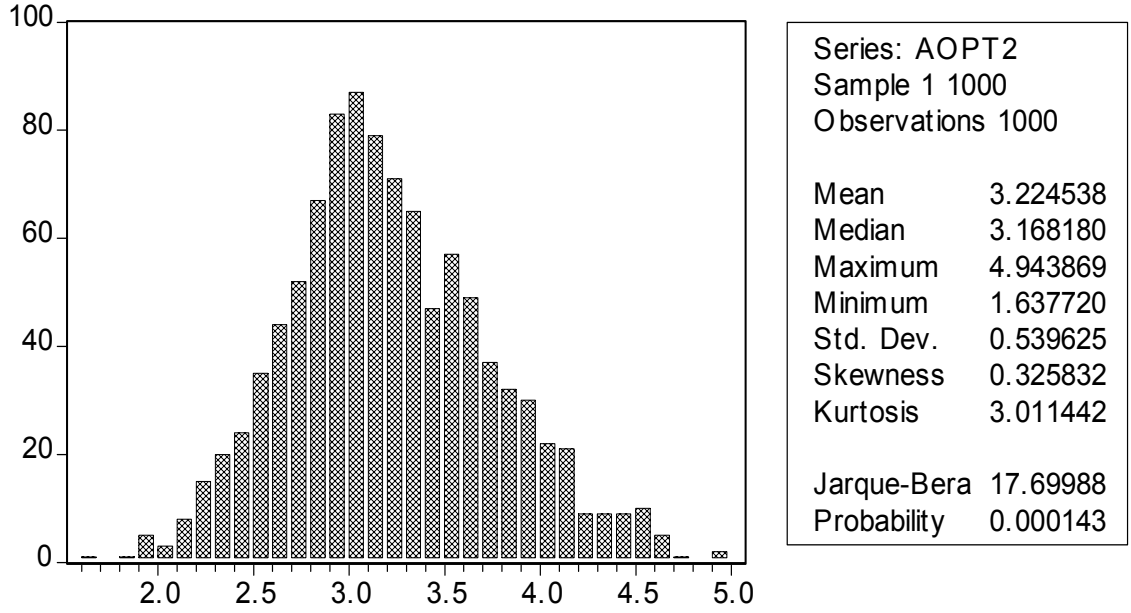

Figure 11. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Lemma 4 across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 2 minute sampling interval. The table contains the corresponding descriptive statistics.

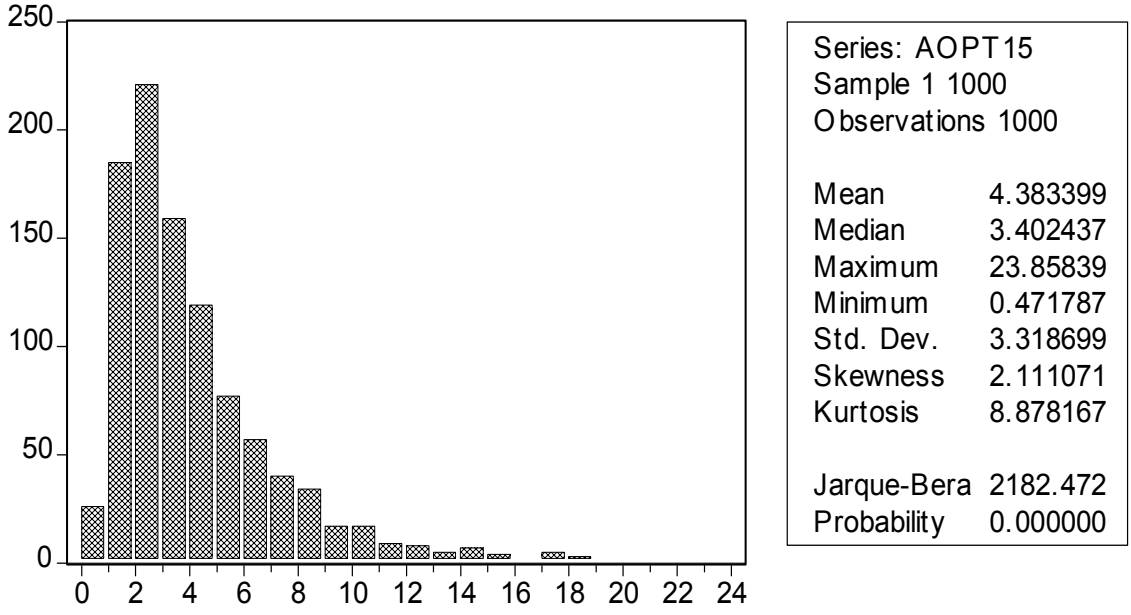

Figure 12. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Lemma 4 across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The table contains the corresponding descriptive statistics.

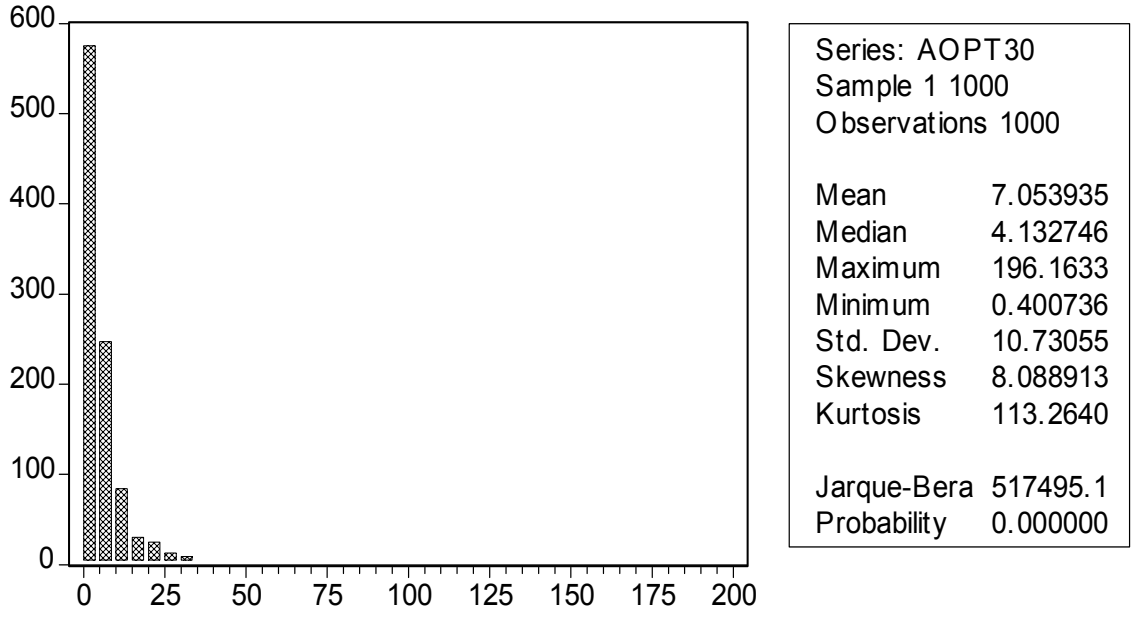

Figure 13. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Lemma 4 across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 30 minute sampling interval. The table contains the corresponding descriptive statistics.

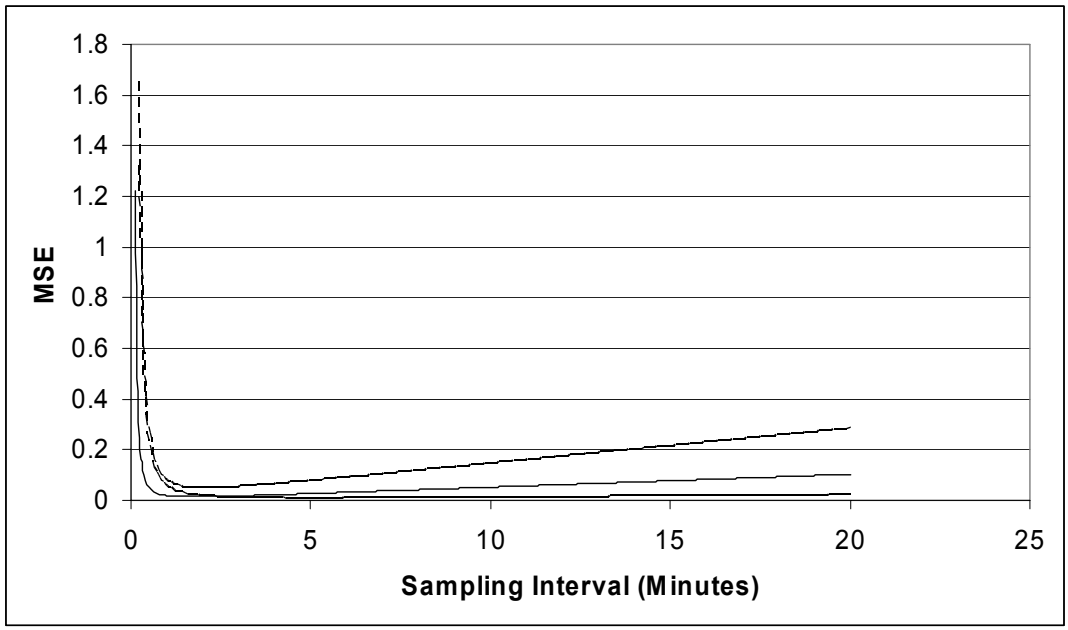

Figure 14. We plot the true MSE expansion in Section 4 and the corresponding 95% bands obtained by implementing the 1,000 simulations in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The moments of the noise process are computed using a 10 second sampling interval.

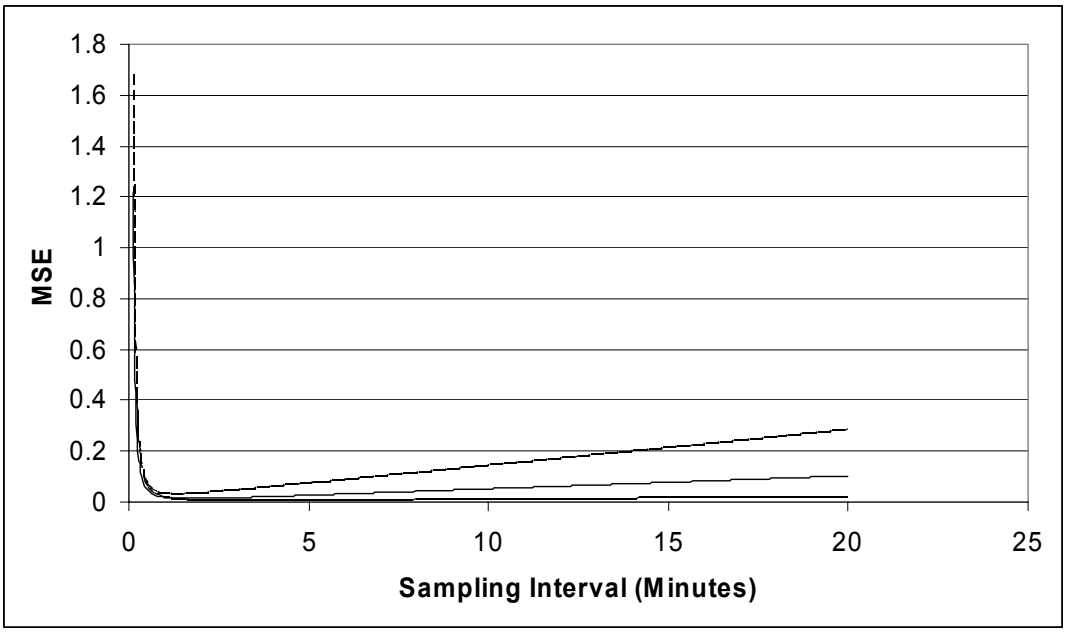

Figure 15. We plot the true MSE expansion in Section 4 and the corresponding 95% bands obtained by implementing the 1,000 simulations in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The moments of the noise process are computed using a 1 second sampling interval.

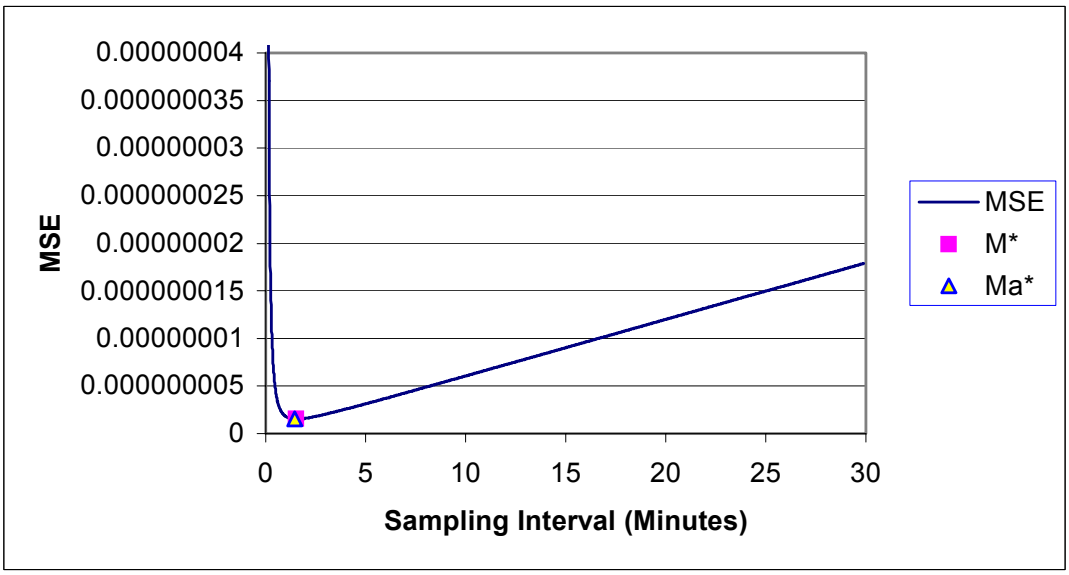

Figure 16. We plot the estimated conditional MSE expansion in Section 4 for the stock IBM. The realized quarticity is computed using a 15 minute sampling interval. M* and Ma* stand for the values corresponding to the true and approximate minima, respectively.