Microstructure noise, realized volatility, and optimal sampling*

Federico M. Bandi and Jeffrey R. Russell
Graduate School of Business, The University of Chicago

Preliminary version: comments welcome

November 3, 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 volatility over a period, the summing of numerous contaminated return data entails substantial accumulation of noise. We argue that the resulting effect is the determination of a bias/variance trade-off. We quantify the trade-off in the presence of a realistic microstructure model of price determination and provide clear and easily implementable directions for optimally sampling high frequency data for the purpose of volatility estimation.

*Paper written for the CIRANO conference "Realized Volatility," Montreal, November 7-8, 2003.

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. (2003) 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 (Revuz and Yor (1994), 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. (2003) and Barndorff-Nielsen and Shephard (2002), BS 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 are believed to diverge from their "efficient values" due to a variety of market frictions. BS (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 BS'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. (2001a) 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 (...)[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 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."

In a recent paper Bandi and Russell (2003a) (BR, henceforth) illustrate rigorously the effect of realistic microstructure noise on the limiting results that are typically employed to justify realized volatility as an estimate of quadratic variation in the presence of high-frequency return data. Specifically, we show that, when noise plays a role, the realized volatility estimator diverges to infinity almost surely as the number of observations (i.e., the sample frequency) increases without bound. Interestingly, even though the object of interest (quadratic variation) cannot be estimated consistently, a standardized version of the realized volatility estimator provides full disclosure of the second moment of the (unobservable) noise process.

Having made these observations, natural remaining issues are how to model the noise's determinants, how to formalize the finite sample loss that is induced by the neglected noise component, and how to employ this information to fully exploit the identification potential of the notion of realized volatility as introduced by Andersen et al. (2003) and BS (2002). Coherently with the model-free spirit of the realized volatility literature, the present paper tackles 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 substantial 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

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

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.[2] In addition, we provide a straightforward methodology to choose optimally the sampling frequency in order to minimize the conditional expected squared distance between the estimator and its theoretical counterpart, as summarized by the conditional MSE.

We discuss a preliminary application of the methodology to high-frequency stock return data and provide support for the optimality of frequencies that are different from those typically employed in the existing empirical work on the subject. For instance, we find that more liquid stocks necessitate higher sampling frequencies than less liquid stocks. The intuition for this finding is as follows. The ratio between the underlying quadratic variation and the variance of the noise process is typically higher for more liquid stocks than for less liquid stocks, thereby generally implying a higher signal-to-noise ratio in the former case. A higher signal leads to the optimality of relatively higher frequencies since the noise component plays a relatively inferior role.

The paper is organized as follows. In Section 2 we discuss the model. Section 3 presents an expansion of the conditional MSE of the quadratic variation estimator in the presence of realistic microstructure noise. In Section 4 we discuss optimal sampling through minimization of the conditional MSE. Section 5 contains simulations. In Section 6 we provide an empirical application to high-frequency stock return data. Section 7 concludes. Proofs and technical details are in the Appendixes.

2 The set-up

We use the same model and notation as in BS (2001b, 2002) but introduce realistic microstructure contaminations coherently with BR (2003a).

Consider a fixed time period h (a trading day, say). We write the observed price

[2] Aït-Sahalia and Mykland (2003) adopt a similar approach in a different context: the study of the implications of microstructure noise on parametric (maximum-likelihood) estimates of diffusion processes via high-frequency return data.

process as

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

where p_{ih} is the true (fair) price and η_{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 $\eta = \ln(\overline{\eta})$. Below we list the assumption that we impose on the model.

Assumption 1.

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

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

where A_{ih} *is a continuous finite variation component and* $M_{ih} = \int_0^{ih} \sigma_s dW_s$ *is a local martingale.*

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

(3) $\sigma_t \perp W_t \qquad \forall t.$

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 an obvious interpretation.

Assumption 2.

(1) *The random shocks* η_j *are i.i.d mean zero with a bounded fourth moment.*

(2) $A_{ih} = 0$. *Hence,* $\mathbf{E}_\sigma(r_{j,i}) = 0$ *and* $\mathbf{E}_\sigma(r_{j,i}^2) = \int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds$ *by the Ito's isometry* $\forall i, j$.

(3) $r_{j,i} \perp \eta_{j,i}$ $\forall i, j$.

Some remarks are in order (the interested reader is referred to BR (2003a) for additional discussions).

The true return process r_t is a local martingale with conditional (on the potentially nonstationary and long-range dependent[3] volatility path $\{\sigma_s\}_{s\in((i-1)h,ih)}$) variance $\mathbf{E}_\sigma(r_{j,i}^2)$ equal to $\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds$. The econometrician does not observe r_t but a contaminated return series $\widetilde{r}_t$ which is given by r_t plus a random shock ε_t. We interpret ε_t as being a microstructure contamination in the return series. It is noted that the shocks ε_t are not uncorrelated since their first order autocovariance is negative and equal to $-\sigma_\eta^2$, i.e., the variance of the underlying i.i.d microstructure noises $\eta' s$ taken with a negative sign. Naturally, the first order autocorrelation of the microstructure noise induces an analogous first order autocorrelation in the contaminated return series. This feature of the noise specification captures a well-known empirical fact. While our findings rely on the model in Eq. (2), the set-up can be easily generalized to allow for higher order structures in the correlation of the error terms $\varepsilon_t' s$ as discussed in the next section.

Consistently with Andersen et al. (2003) and BS (2002), we define the realized volatility estimator for the generic day i as

$$\widehat{V}_i = \sum_{j=1}^{h/\delta} \widetilde{r}_{j,i}^2, \tag{6}$$

and use $\widehat{V}_i$ to estimate $V_i = \int_{(i-1)h}^{ih} \sigma_s^2 ds$, i.e., the quadratic variation of the underlying log price process over the same day. Conditionally on the volatility path, the MSE of the estimator can be written as

[3]Long memory is a typical feature of volatility series. The interested reader is referred to Bandi and Perron (2001), Ohanissian et al. (2003) and the references therein for recent evidence and discussions.

$$\mathbf{E}_\sigma\left(\widehat{V}_i - V_i\right)^2 = \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta} \widetilde{r}_{j,i}^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds\right)^2. \tag{7}$$

In the next section we show that the MSE does not converge to zero as the sampling frequency increases without bound. In fact, 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 so-called *quarticity* of the log price process, namely $V_i^2 = \int_{(i-1)h}^{ih} \sigma_s^4 ds$ (see BS (2002, 2003) for a complete discussion of this notion).

3 The conditional MSE

The conditional MSE of the integrated volatility estimator can be represented as in Theorem 1 below.

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

$$\mathbf{E}_\sigma\left(\widehat{V}_i - \int_{(i-1)h}^{ih} \sigma_s^2 ds\right)^2 = 2\frac{h}{M}\left(V_i^2 + o_{a.s.}(1)\right) + M\beta + M^2\alpha + \gamma, \tag{8}$$

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

$$\alpha = \left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2, \tag{9}$$

$$\beta = \mathbf{E}_\sigma\left(\varepsilon^4\right) + 2\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) - 3\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2, \tag{10}$$

and

$$\gamma = 4\mathbf{E}_\sigma(\varepsilon^2)V_i - 2\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) + 2\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2. \tag{11}$$

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., $2\frac{h}{M}\left(V_i^2 + o_{a.s.}(1)\right)$ (see Appendix A for a derivation).

When the microstructure noise is present, Eq. (8) 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 (see Section 5 and 6 for empirical assessments of this result). The form of the additional variance term is

$$M\mathbf{E}_\sigma(\varepsilon^4) + 2(M-1)\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) + (2-3M)\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2 + 4\mathbf{E}_\sigma(\varepsilon^2)V_i, \tag{12}$$

where $\mathbf{E}_\sigma(\varepsilon^4)$, $\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2)$,[4] and $\mathbf{E}_\sigma(\varepsilon^2)$ are obvious moments of the noise-in-return distribution and $4\mathbf{E}_\sigma(\varepsilon^2)V_i$ is an interaction term. The form of the bias is $M\mathbf{E}_\sigma(\varepsilon^2)$. 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:

$$\overline{\beta} = \mathbf{E}_\sigma\left(\varepsilon^4\right) + 2\sum_{b=1}^{s}\mathbf{E}_\sigma\left(\varepsilon^2\varepsilon_{-b}^2\right) - (2s+1)\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2, \tag{13}$$

$$\overline{\gamma} = 4\mathbf{E}_\sigma(\varepsilon^2)V_i - 2\sum_{b=1}^{s} b\mathbf{E}_\sigma\left(\varepsilon^2\varepsilon_{-b}^2\right) + s(s+1)\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2. \tag{14}$$

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

4 Optimal sampling

Coherently with Eq. (8) above, we define the optimal number of observations (per unit interval h) as the number M^* which satisfies the following condition:

$$\left\{M^* := \arg\min\left(2\frac{h}{M}\left(V_i^2 + o_{a.s.}(1)\right) + M\beta + M^2\alpha + \gamma\right)\right\} \tag{15}$$

or, equivalently,

[4] Our model implies $\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) = \frac{1}{2}\mathbf{E}_\sigma(\varepsilon^4)$ (See Appendix B).

$$\left\{M^* : 2M^3\alpha + M^2\beta - 2h\left(V_i^2 + o_{a.s.}(1)\right) = 0\right\}, \tag{16}$$

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

Lemma 1. (A useful rule-of-thumb for applications.) *At high frequencies,*

$$M^* \sim \left(\frac{hV_i^2}{\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2}\right)^{1/3}, \tag{17}$$

where V_i^2 is the realized quartic variation, i.e., $V_i^2 = \int_{(i-1)h}^{ih} \sigma_s^4 ds$, and $\mathbf{E}_\sigma(\varepsilon^2)$ is the second moment of the contaminations-in-returns.

Proof. *Immediate given Eq. (15).*

Interestingly, when the quadratic term in Eq. (15) dominates the linear term (for values of M sufficiently large), the approximation in Eq. (17) provides a very good representation of the optimal number of observations M. In the empirical application that we discuss in Section 6 below such dominance occurs in all cases.

Lemma 1 is important for two reasons. First, it provides us 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. (15). 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. This result perfectly reflects the analysis in BR (2003a). There we show that the higher is the ratio between the variance of the unobservable noise process and the underlying quadratic vatiation, the larger is the bias at high frequencies, thereby requiring a larger sampling interval or, to put it differently, a lower number of observations. Naturally, the impact of the ratio between the (squared) variance of the noise process and the quartic volatility in Eq. (17) above can be interpreted in a similar fashion. Such ratio can be regarded as a signal-to-noise ratio: the stronger the signal is, the higher the optimal sampling frequency should be.

4.1 Estimation of the optimal sampling frequency

Eq. (16) can be readily solved numerically in the presence of consistent estimates of the relevant terms in Eqs. (9) and (10).

Inevitably, the very nature of the problem makes the consistent estimation of the necessary input V_i^2 forbidding in the presence of microstructure noise. Nonetheless, the simulations in the next section show that the use of different (plausible) estimates of it does not have any considerable impact on the optimal sampling frequency of the realized volatility estimator.

While we can hope to provide a valid preliminary estimate of V_i^2, the availability of high frequency data allows us to consistently estimate the remaining inputs, i.e., the moments of the microstructure contaminations. BR (Theorem 1, 2003a) show that a (standardized) version of the realized volatility estimator converges in probability to the second moment of the unobservable noise process, i.e., $\mathbf{E}_\sigma\left(\varepsilon^2\right)$. Here we complement this unexpected result by showing that a (standardized) version of the realized quarticity converges in probability to the fourth moment of the microstructure contaminations in returns, i.e., $\mathbf{E}_\sigma\left(\varepsilon^4\right)$.

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

$$\frac{1}{M}\sum_{j=1}^{h/\delta}\widetilde{r}_{j,i}^4 \xrightarrow{p} \mathbf{E}_\sigma\left(\varepsilon^4\right). \tag{18}$$

Proof. *See Appendix C.*

Since our model implies $\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) = 2\mathbf{E}_\sigma\left(\varepsilon^4\right)$, Theorem 2 above and Theorem 1 in BR (2003a) allow us to consistently estimate *all* of the relevant moments of the noise distribution by simply averaging powers of the contaminated return data. As pointed out earlier, even though data collected at very high frequencies do not permit us to identify the conventional object of interest, be it the quadratic variation or the quartic volatility, when realistic price formation mechanisms are in place, they do permit to identify features of the microstructure contaminations. We use those features to learn about quadratic variation through the solution of Eq. (16) above.

One final observation is needed. The conditional MSE in Eq. (8) 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. (8). 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. (16) above with $\frac{1}{n}\sum_{i=1}^{n} V_i^2$, where n denotes the number of periods h, in place of V_i^2.

In Section 6 we provide an application of the methodology to quote-by-quote stock return data. The following section contains simulations.

5 Simulations

We aim at assessing the accuracy of our theory in the presence of realistic parameter values for the model in Eq. (2) above. First, we verify that increasing the sampling frequency allows us to consistently estimate the fourth moment of the microstructure noise. As said, when paired with the identifiability of the second moment of the noise process (as discussed in BR (2003a)), this finding represents a substantial step forward in the characterization of the minimum of the conditional MSE expansion in Eq. (8). Second, we show that alternative (but credible) sampling frequencies used to compute the underlying quarticity (i.e., the remaining input in Eq. (16)) have little impact on the sampling distribution of the optimal number of observations to be used to calculate the object of interest, i.e., the underlying quadratic variation of the log price process. More precisely, we argue that we expect the 15-minute sampling interval to be a valid (albeit often conservative) interval to estimate the underlying quartic variation for a variety of stocks with different liquidity features. Finally, we show that, when the quarticity is estimated relatively accurately, the rule-of-thumb in Lemma 1 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.

We simulate a data generating process that is similar to the process in BR (2003a). Here we describe the salient aspects of the methodology but refer the reader to BR (2003a) for additional details. Specifically, the dynamics of the log price process are described by

the stochastic differential equation

$$d\log(p)_t = \sigma_t dW_t^1, \tag{19}$$

with

$$d\sigma_t^2 = \kappa(\overline{v} - \sigma_t^2)dt + \varpi\sigma_t dW_t^2. \tag{20}$$

We set the persistence parameter κ equal to .01. We normalize the mean spot volatility to 1 and hence set $\overline{v}$ equal to 1. The parameter ϖ is chosen equal to .05. As for the logged noises η, we assume that they are normally distributed with mean zero and variance equal to $(.000197)^2$. The variance of the logged noise process implies a ratio between the variance of the contaminations in returns ε and the mean daily variance equal to 0.02829%. This value is coherent with the ratio between the noise variance and the estimated average (daily) realized volatilities for the stock IBM over the month of February 2002 (further discussions are contained in BR (2003a)). In this sense, and given the important role played by the above-mentioned ratio (see BR (2003a)), our simulations are representative of a relatively liquid stock like IBM.

We simulate $1,000$ contaminated return series around a single realization of the volatility over a period of 6.5 hours. More precisely, we employ the specification in Eq. (20) to simulate second-by-second a volatility path given an initial value of σ_t^2 equal to the unconditional mean of 1. Holding the volatility path fixed, we then simulate second-by-second true returns using Eq. (19) and second-by-second observed returns as in Eq. (2) given the normality assumption on the logged noise process.

Figures 1 and 2 represents plots of the true and observed return processes for two different sampling frequencies. In Figure 1 we plot 200 second-by-second returns whereas in Figure 2 we plot 20 15-minute interval returns. The graphs show that the second-by-second returns are affected by noise substantially more than the returns computed at lower frequencies.

In Figure 3 we plot the average (across simulations) empirical fourth moment of the contaminated returns for a variety of sampling frequencies. In agreement with the statement of Theorem 2, the average of the contaminated returns raised to the fourth power

converges to the fourth moment of the (unobservable) noise process as the sampling frequency increases without bound. As pointed out earlier, this finding complements an analogous result in BR (2003a). There we find that the standardized (by the number of observations) realized volatility converges to second moment of the noise process.

While both results allow us to work towards an accurate characterization of the conditional MSE expansion in Theorem 1 for the purpose of finding the optimal sampling frequency of the quadratic variation estimator, we can only hope to provide a valid estimate of the remaining necessary input, namely the underlying quarticity. Consequently, the following discussion focuses on the estimation of the quartic volatility. In Figure 4 we plot the empirical MSE of the realized quarticity. It is apparent that 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, 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. This observation will lead us to recommend the 15 minute sampling frequency as a valid frequency for stocks with various degrees of liquidity. Naturally, as implied by Figure 4, such choice is quite conservative for highly liquid stocks like IBM. We will return to these important observations.

In Figure 5 we plot the distribution (across the $1,000$ simulations) of the optimal sampling frequencies obtained by minimizing the expression in Eq. (8) 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, it 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 low frequencies corresponding to the upward spike in the MSE of the realized volatility estimator. Finally, for the range of values in Figure 5, 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 Figure 4.

Hence, in Figure 6 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 Figure 4) 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 a very 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. Of course, one should note that using a 15 minute frequency for the quarticity is a very 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 is sufficiently informative and easy to use as to deserve attention in applied work. Coherently, it is the frequency that we utilize (for stocks with different liquidity features) in the empirical application in next section.

For completeness, we also report results for the case where the quarticity is computed using a sampling frequency equal to 30 minutes (see Figure 7). 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 around 5 minutes. All in all, albeit being biased, we believe 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 the quarticity (as determined by very suboptimal choices of the corresponding frequency) do not cause equally suboptimal sampling frequencies for the realized volatility estimates.

In Figure 8 through 10 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 1. 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 1 above appears very valid when a close-to-optimal frequency for the quarticity is chosen (see Figure 8). When using a 15 minute frequency for the quarticity, for instance, the estimates that the approximation provides are a lot 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 left skewness in the simulated distribution, the likelihood of obtaining values around 2 minutes (i.e., near the true optimal frequency) is about 50%.

In Figures 11 and 12 we plot the true conditional MSE as implied by Eq. (8) above and corresponding 95% bands based on the simulations. 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 (see BR (2003a)), in Figure 11 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 Figure 12. 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.

6 An empirical application

We apply our methodology to three stocks with various liquidity features. Specifically, we consider IBM, Alcoa, and Airgas (see Table 1 for descriptive statistics).

IBM is the most liquid stock in our sample, followed by Alcoa and Airgas. Naturally, the liquidity features of the stocks have an impact on the ratio between the second moment of the (unobservable) noise process and the average (daily) realized volatility. For IBM, Alcoa, and Airgas, the ratios are equal to .0283%, .0742%, and .217%, respectively.

In Figure 13 through 14 we plot the estimated conditional MSE's of the three stocks along with the minimum from the full minimization as implied by Eq. (16) and the minimum from the rule-of-thumb reported in Eq. (17). The realized quarticity is estimated conservatively (see previous section) using the 15 minute sampling interval. We employ the high frequency quote-by-quote price changes to calculate the moments of the noise process.

Several observations are in order. The optimal sampling interval is directly related to the stock's liquidity: the less liquid the stock is, the larger the optimum sampling interval is. More precisely, the minima are equal to 1.5 minutes (IBM), 2.9 minutes (Alcoa), and 5.3 minutes (Airgas). These values (c.f. Airgas) can be close to the 5 minute interval that is used in some empirical work on the subject (see Andersen et al. (2001a), for instance). Nonetheless, they are considerably smaller than recent conjectures on optimal sampling based on 15/20 minute time intervals (see Andersen et al. (2000)). Of course, the loss that is induced by suboptimal sampling varies across stocks depending on the slope of the corresponding conditional MSE. In the case of Alcoa, for example, going from the optimal frequency of 2.9 minutes to the 15 minute interval more than triples the MSE. The loss is very significant in the case of IBM and Airgas too. In the former case the MSE almost doubles while in the later case it more than doubles. These effects are economically important since the magnitude of the MSE's is large. One can easily have a feel for their magnitudes by a straightforward comparison of the root MSE's (as implied by the values on the horizontal axis of Figure 13 through 14) and the corresponding average realized volatilities, namely .0273% (IBM), .038% (Alcoa), and .0255% (Airgas). At the 15

minute interval the ratios between the two quantities are equal to about 36% (IBM), 32% (Alcoa), and 43% (Airgas). 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 our MSE's expansions to the variances that would emerge from models that do not explicitly allow for noise (BS (2002), for instance). In the case of the three stocks examined here we find that the bias squared alone can be as large as the variance and sometimes as much as two times as large.

Finally, we notice that the rule-of-thumb in Lemma 1 provides empirical answers to the optimal frequency problem that are almost indistinguishable from the answers that are provided by the full minimization of the expansion in Eq. (8). More precisely, we find that the approximate optimal frequencies are 1.45 (IBM), 2.85 (Alcoa), and 5.12 (Airgas). 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 previous Section), the rule-of-thumb can provide very useful and immediate implications for empirical work.

7 Conclusion

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 identify the underlying volatility over a period, the summing of numerous contaminated return series entails substantial accumulation of noise. We argue that the final effect is the determination of a 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 moments of the distribution of the noise process and the underlying quarticity of the log price process.

Specifically, we derive implications for empirical work on realized volatility estimation that could be summarized 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 we cannot hope to provide consistent nonparametric estimates of the remaining ingredient of the MSE expansion, i.e., the quartic volatility, we stress that alternative (plausible) choices of sampling frequency for the quarticity do not have a substantial impact on the optimal sampling of the realized volatility estimator.

(4) Specifically, we deem the (easy to implement) 15-minute sampling interval to be a valid (albeit conservative) choice of frequency for the quarticity. 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 a ratio 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 low and the quarticity is accurately estimated.

A preliminary application of our methodology to stocks with various degrees of liquidity shows the optimality of frequencies that vary with the degree of liquidity. In particular, the less liquid the stock is, the larger should be the sampling interval. Interestingly, we find sampling intervals that are either smaller than 5 minutes or very close to it, thereby contradicting recent conjectures on optimal sampling relying on the 15 minute interval.

Furthermore, our analysis suggests the following conclusions:

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

(2) MSE's that fail to account for market microstructure 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. Since volatility appears to be better revealed by optimal frequencies that vary across assets, it seems valuable to revisit the properties of the nonparametric quadratic variation estimates in the presence of individually selected optimal sampling frequencies. Work on the subject is being conducted by the authors and will be reported in later work.

8 Appendix A

Proof of Theorem 1. We expand Eq. (7) and obtain

$$
\begin{aligned}
& \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} (r_{j,i} + \varepsilon_{j,i})^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds \right)^2 \\
= \; & \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \left(r_{j,i}^2 + \varepsilon_{j,i}^2 + 2 r_{j,i} \varepsilon_{j,i} \right) - \int_{(i-1)h}^{ih} \sigma_s^2 ds \right)^2 \qquad (21) \\
= \; & \underbrace{\mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} r_{j,i}^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds \right)^2}_{A} + \underbrace{\mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} \left(\varepsilon_{j,i}^2 + 2 r_{j,i} \varepsilon_{j,i} \right) \right)^2}_{B} \\
& \underbrace{+ \mathbf{E}_\sigma \left(\left(\sum_{j=1}^{h/\delta} r_{j,i}^2 - \int_{(i-1)h}^{ih} \sigma_s^2 ds \right) \left(\sum_{j=1}^{h/\delta} \left(\varepsilon_{j,i}^2 + 2 r_{j,i} \varepsilon_{j,i} \right) \right) \right)}_{C} . \qquad (22)
\end{aligned}
$$

We start with B.

$$
\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 \\
= \; & \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) \qquad (23) \\
= \; & \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) \qquad (24) \\
= \; & \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) \qquad (25) \\
= \; & \sum_{j=1}^{h/\delta} \mathbf{E}_\sigma \left(\varepsilon_{j,i}^4 \right) + 2 \sum_{g=1}^{h/\delta} \sum_{j<g} \mathbf{E}_\sigma \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}) \qquad (26) \\
= \; & \frac{h}{\delta} \mathbf{E}_\sigma (\varepsilon^4) + 2 \left(\frac{h}{\delta} - 1 \right) \mathbf{E}_\sigma \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}_\sigma (\varepsilon^2) \right)^2 \\
& + 4 \mathbf{E}_\sigma (\varepsilon^2) V_i \qquad (27) \\
= \; & \left(\frac{h}{\delta} \right)^2 \left(\mathbf{E}_\sigma (\varepsilon^2) \right)^2 + \frac{h}{\delta} \left(\mathbf{E}_\sigma \left(\varepsilon^4 \right) + 2 \mathbf{E}_\sigma \left(\varepsilon_i^2 \varepsilon_{-1,i}^2 \right) - 3 \left(\mathbf{E}_\sigma (\varepsilon^2) \right)^2 \right) \\
& + 4 \mathbf{E}_\sigma (\varepsilon^2) V_i - 2 \mathbf{E}_\sigma \left(\varepsilon_i^2 \varepsilon_{-1,i}^2 \right) + 2 \left(\mathbf{E}_\sigma (\varepsilon^2) \right)^2 \qquad (28) \\
= \; & \frac{h^2}{\delta^2} \alpha + \frac{h}{\delta} \beta + \gamma, \qquad (29)
\end{aligned}
$$

where

$$\alpha = \left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2, \tag{30}$$

$$\beta = \mathbf{E}_\sigma\left(\varepsilon^4\right) + 2\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) - 3\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2, \tag{31}$$

and

$$\gamma = 4\mathbf{E}_\sigma(\varepsilon^2)V_i - 2\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) + 2\left(\mathbf{E}_\sigma(\varepsilon^2)\right)^2. \tag{32}$$

We recall that

$$\sum_{j=1}^{h/\delta}\mathbf{E}_\sigma(r_{j,i}^2) = \sum_{j=1}^{h/\delta}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right) = V_i. \tag{33}$$

This result is used in Eq. (27) above as well as in Eq. (35) below. We now consider C.

$$\begin{aligned}
&\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^2 - \int_{(i-1)h}^{ih}\sigma_s^2 ds\right)\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2 + 2r_{j,i}\varepsilon_{j,i}\right)\right) \\
= \;& \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^2\right)\mathbf{E}_\sigma\left(\sum_{g=1}^{h/\delta}\varepsilon_{g,i}^2\right) + 2\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}r_{j,i}^2 r_{g,i}\varepsilon_{g,i}\right) - V_i\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_{j,i}^2 + 2r_{j,i}\varepsilon_{j,i}\right)\right) \qquad (34)\\
= \;& \frac{h}{\delta}V_i\mathbf{E}_\sigma(\varepsilon^2) - \frac{h}{\delta}V_i\mathbf{E}_\sigma(\varepsilon^2) = 0. \qquad (35)
\end{aligned}$$

We now turn to A. Write

$$\begin{aligned}
&\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^2 - \int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \\
= \;& \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^2\right)^2 - 2\mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^2\right)\left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right) + \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (36)\\
= \;& \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}r_{j,i}^2 r_{g,i}^2\right) - \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (37)\\
= \;& \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_\sigma\left(r_{j,i}^2 r_{g,i}^2\right) - \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (38)\\
= \;& \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_\sigma\left(r_{j,i}^2\right)\mathbf{E}_\sigma\left(r_{g,i}^2\right) - \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (39)\\
= \;& \mathbf{E}_\sigma\left(\sum_{j=1}^{h/\delta}r_{j,i}^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)\left(\int_{(i-1)h+(g-1)\delta}^{(i-1)h+g\delta}\sigma_s^2 ds\right) \\
&- \left(\int_{(i-1)h}^{ih}\sigma_s^2 ds\right)^2 \qquad (40)\\
= \;& \sum_{j=1}^{h/\delta}\mathbf{E}_\sigma\left(r_{j,i}^4\right) - \sum_{j=1}^{h/\delta}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)^2 \qquad (41)\\
= \;& \sum_{j=1}^{h/\delta}\mathbf{V}_\sigma\left(r_{j,i}^2\right) + \sum_{j=1}^{h/\delta}\left(\mathbf{E}_\sigma\left(r_{j,i}^2\right)\right)^2 - \sum_{j=1}^{h/\delta}\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta}\sigma_s^2 ds\right)^2 \qquad (42)
\end{aligned}$$

$$= \sum_{j=1}^{h/\delta} \mathbf{V}_\sigma \left(r_{j,i}^2 \right) \tag{43}$$

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,

$$\frac{r_{j,i}^2}{\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds \right)} \tag{44}$$

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

$$\sum_{j=1}^{h/\delta} \mathbf{V}_\sigma \left(r_{j,i}^2 \right) = \sum_{j=1}^{h/\delta} \mathbf{V}_\sigma \left(\frac{r_{j,i}^2}{\left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds \right)} \left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds \right) \right) \tag{45}$$

$$= \sum_{j=1}^{h/\delta} 2 \left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds \right)^2 \tag{46}$$

$$= 2\frac{h}{M} \left(\frac{M}{h} \sum_{j=1}^{h/\delta} 2 \left(\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \sigma_s^2 ds \right)^2 \right) \tag{47}$$

$$= 2\frac{h}{M} \left(V_i^2 + o_{a.s.}(1) \right), \tag{48}$$

where almost sure converges to the realized quarticity, that is $V_i^2 = \int_{(i-1)h}^{ih} \sigma_s^4 ds$, follows from a result in BS (2002). Finally,

$$\mathbf{E}_\sigma \left(\widehat{V}_i - V_i \right)^2 = \delta 2 \left(V_i^2 + o_{a.s}(1) \right) + \frac{h}{\delta}\beta + \frac{h^2}{\delta^2}\alpha + \gamma. \tag{49}$$

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

9 Appendix B: The moments of the microstructure contaminations in returns

Simple calculations permit to show that:

(1)

$$\mathbf{E}_\sigma(\varepsilon^2 \varepsilon_{-1}^2) = 3\sigma_\eta^4 + \mathbf{E}_\sigma \left(\eta^4 \right), \tag{50}$$

(2)

$$\mathbf{E}_\sigma \left(\varepsilon^4 \right) = 2\mathbf{E}_\sigma \left(\eta^4 \right) + 6\sigma_\eta^4, \tag{51}$$

and

(3)

$$\mathbf{E}_\sigma(\varepsilon^2) = 2\sigma_\eta^2, \tag{52}$$

(see BR (2003a)).

(4) We plug (52) into (51) and rearrange to obtain

$$\begin{aligned}\mathbf{E}_\sigma(\eta^4) &= \frac{1}{2}\mathbf{E}_\sigma\left(\varepsilon^4\right) - 3\sigma_\eta^4 \\ &= \frac{1}{2}\mathbf{E}_\sigma\left(\varepsilon^4\right) - \frac{3}{4}\sigma_\varepsilon^4. \end{aligned} \tag{53}$$

(5) Plugging (53) and (52) into (50) yields

$$\mathbf{E}_\sigma(\varepsilon^2\varepsilon_{-1}^2) = 3\sigma_\eta^4 + \mathbf{E}_\sigma\left(\eta^4\right) \tag{54}$$

$$= \frac{3}{4}\sigma_\varepsilon^4 + \frac{1}{2}\mathbf{E}_\sigma\left(\varepsilon^4\right) - \frac{3}{4}\sigma_\varepsilon^4 \tag{55}$$

$$= \frac{1}{2}\mathbf{E}_\sigma\left(\varepsilon^4\right). \tag{56}$$

Hence, consistent estimates of $\mathbf{E}_\sigma\left(\varepsilon^4\right)$ can be readily used to provide consistent estimates of the first-order autocovariance of the squared contaminations in returns.■

10 Appendix C

Proof of Theorem 2. Write

$$\sum_{j=1}^{h/\delta} \widetilde{r}_{j,i}^4 = \sum_{j=1}^{h/\delta} \left(r_{j,i} + \varepsilon_{j,i}\right)^4 \tag{57}$$

$$= \underbrace{\sum_{j=1}^{h/\delta} r_{j,i}^4}_{A} + 4\underbrace{\sum_{j=1}^{h/\delta} r_{j,i}^3\varepsilon_{j,i}}_{B} + 6\underbrace{\sum_{j=1}^{h/\delta} r_{j,i}^2\varepsilon_{j,i}^2}_{C} + 4\underbrace{\sum_{j=1}^{h/\delta} r_{j,i}\varepsilon_{j,i}^3}_{D} + \underbrace{\sum_{j=1}^{h/\delta} \varepsilon_{j,i}^4}_{E}. \tag{58}$$

Now we note that

$$\frac{A}{M} \leq \frac{\left(\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\right)\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\right)\right)}{M} \tag{59}$$

$$= \frac{\left(\int_{(i-1)h}^{ih} \sigma_s^2 ds + 2\sum_{j=1}^{h/\delta} \overline{M}_{j-1,j}\right)^2}{M}, \tag{60}$$

where

$$\overline{M}_{j-1,j} = \int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \left(M^i - M_{j-1}^i\right) dM^i, \tag{61}$$

from BR (2003a). Thus,

$$\frac{A}{M} \leq \frac{\left(\int_{(i-1)h}^{ih} \sigma_s^2 ds\right)^2 + \left(2\sum_{j=1}^{h/\delta} \overline{M}_{j-1,j}\right)^2 + 2\left(\int_{(i-1)h}^{ih} \sigma_s^2 ds\right)\left(2\sum_{j=1}^{h/\delta} \overline{M}_{j-1,j}\right)}{M} \tag{62}$$

$$\leq \frac{1}{M}\left[O_p\left(1\right) + O_p\left(\sqrt{\frac{1}{M}}\right) O_p\left(\sqrt{\frac{1}{M}}\right) + O_p\left(1\right) O_p\left(\sqrt{\frac{1}{M}}\right)\right] \xrightarrow{p} 0, \tag{63}$$

using results in BR (2003a). Now write

$$\frac{C}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} r_{j,i}^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} \varepsilon_{j,i}^4\right)^{1/2}}{M} \tag{64}$$

$$\leq O_p\left(\sqrt{\frac{1}{M}}\right)\left[\left(\mathbf{E}\left(\varepsilon^4\right)\right)^{1/2} + O_p\left(\frac{1}{\sqrt{M}}\right)\right] \xrightarrow{p} 0, \tag{65}$$

where the inequality in Eq. (64) follows from the Cauchy-Schwartz inequality and the inequality in Eq. (65) derives from Eq. (63) above and a straightforward application of the delta method. Now consider term B.

$$\frac{B}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} r_{j,i}^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\varepsilon_{j,i}^2\right)^{1/2}}{M} \tag{66}$$

$$\leq O_p\left(\sqrt{\frac{1}{M}}\right) O_p\left(\frac{1}{M^{1/4}}\right) \xrightarrow{p} 0, \tag{67}$$

where the inequality in Eq. (66) follows again from the Cauchy-Schwartz inequality and the inequality in Eq. (67) derives from Eq. (65) above. We now turn to term D/M and show convergence to zero in probability using a similar approach:

$$\frac{D}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} \varepsilon_{j,i}^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} r_{j,i}^2\varepsilon_{j,i}^2\right)^{1/2}}{M} \tag{68}$$

$$= \left(\left(\mathbf{E}\left(\varepsilon^4\right)\right)^{1/2} + O_p\left(\frac{1}{\sqrt{M}}\right)\right) O_p\left(\frac{1}{M^{1/4}}\right) \xrightarrow{p} 0. \tag{69}$$

Finally,

$$\frac{E}{M} \xrightarrow{p} \mathbf{E}\left(\varepsilon^4\right),$$

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

References

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

[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.* Amsterdam: North Holland. Forthcoming.

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

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

[5] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association,* 42-55.

[6] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003). 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 J.R. Russell (2003a). Volatility or noise? *Unpublished paper, GSB - University of Chicago.*

[10] Barndorff-Nielsen, O.E., and N. Shephard (2001a). 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.*

[11] Barndorff-Nielsen, O. E., and N. Shephard (2001b). How accurate is the asymptotic approximation to the distribution of realised variance? *Working paper.*

[12] 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.*

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

[14] 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.*

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

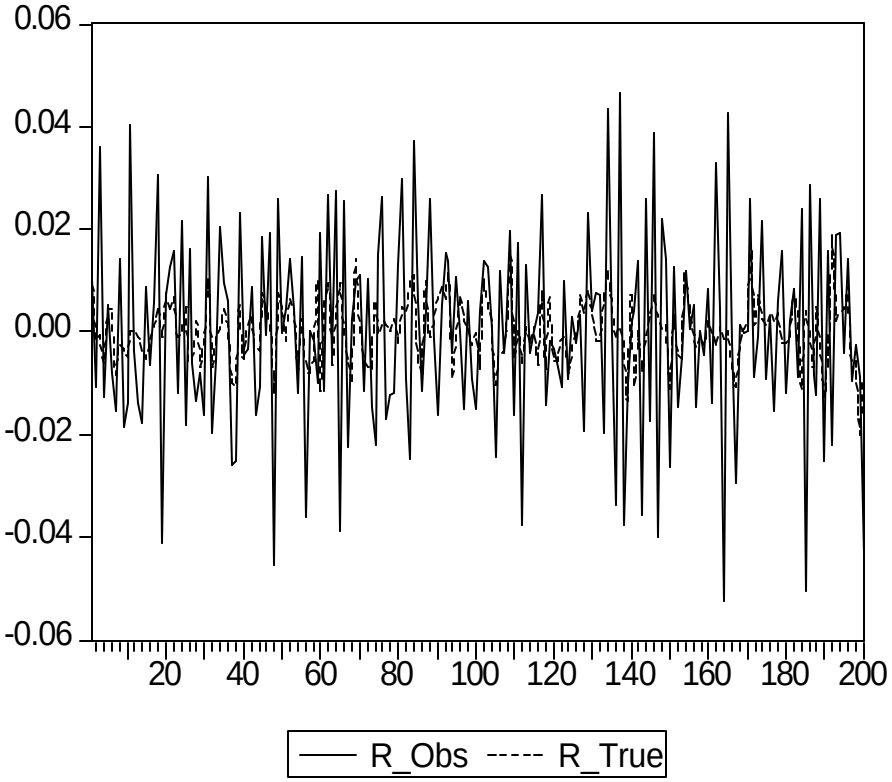

Figure 1. Plot of the true and observed return processes. We plot 200 second-by-second returns. The series are obtained using the simulations described in Section 5.

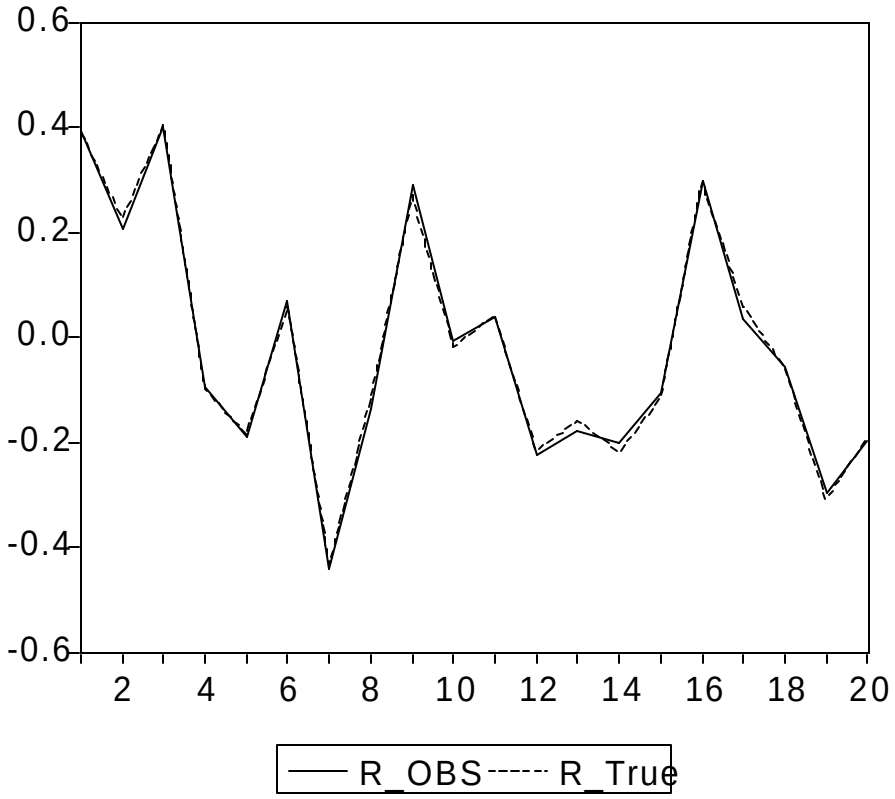

Figure 2. Plot of the true and observed return processes. We plot 20 15-minute returns. The series are obtained using the simulations described in Section 5.

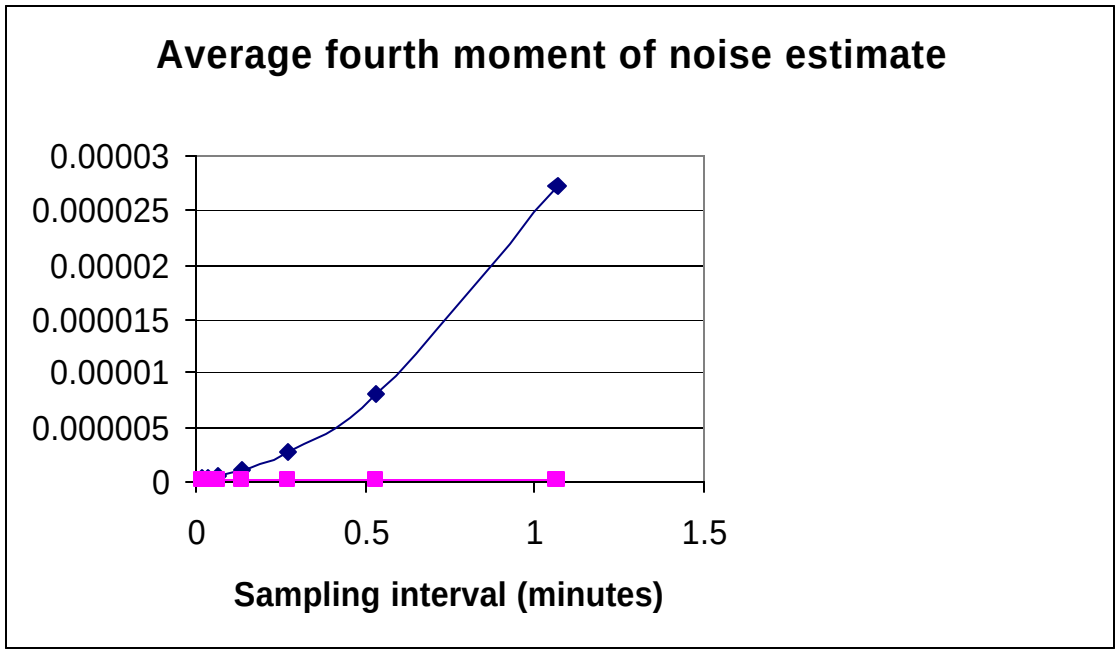

Figure 3. We plot the average of the fourth moment estimates across the 1,000 simulations described in Section 5. The horizontal line denotes the known fourth moment of the noise process.

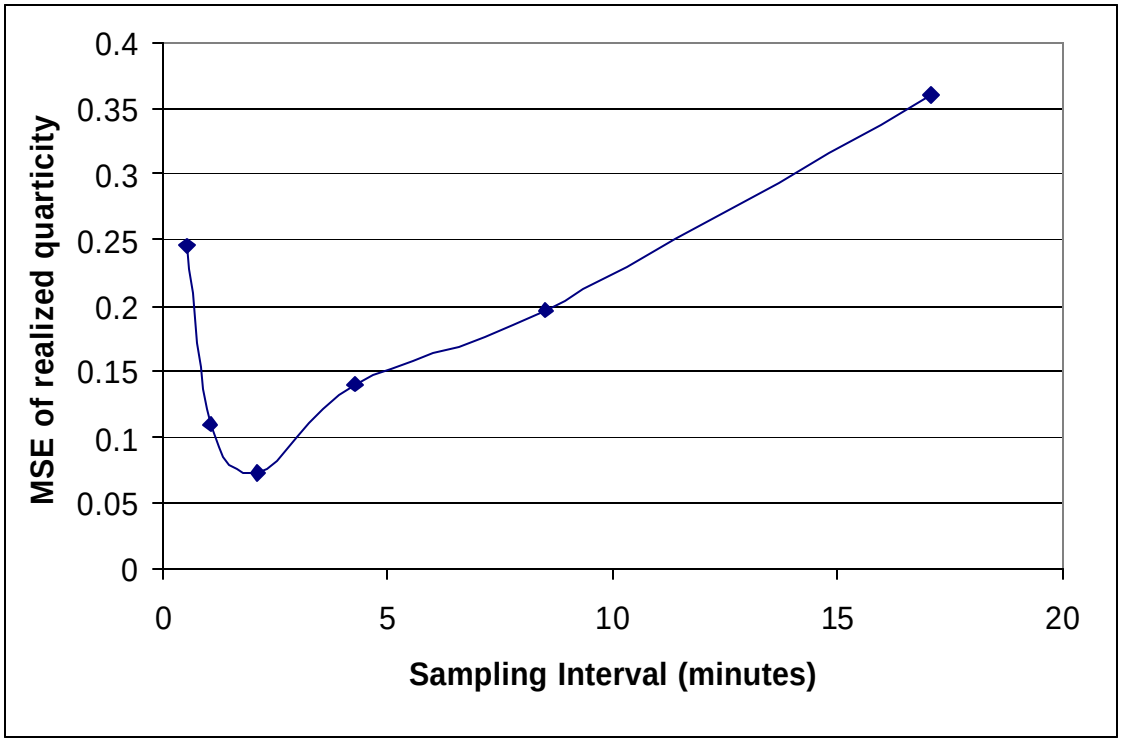

Figure 4. We plot the simulated conditional mean-squared error of the realized quarticity. We use 1,000 simulation as described in Section 5.

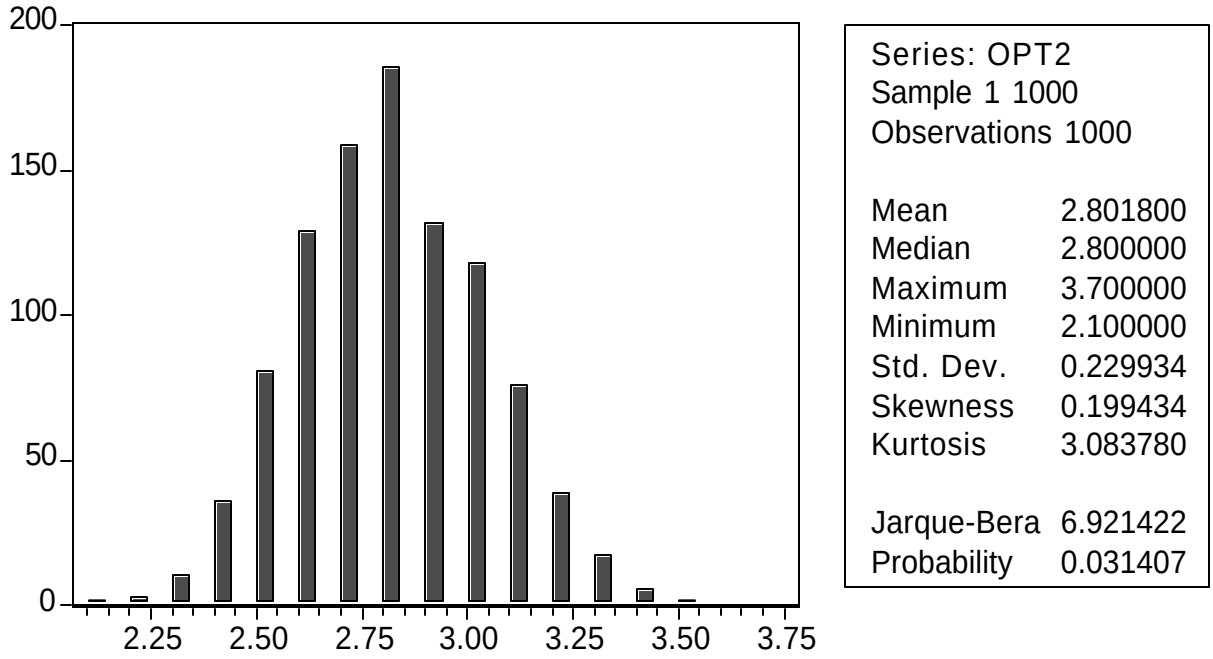

Figure 5. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 2 minute sampling interval. The table contains the corresponding descriptive statistics.

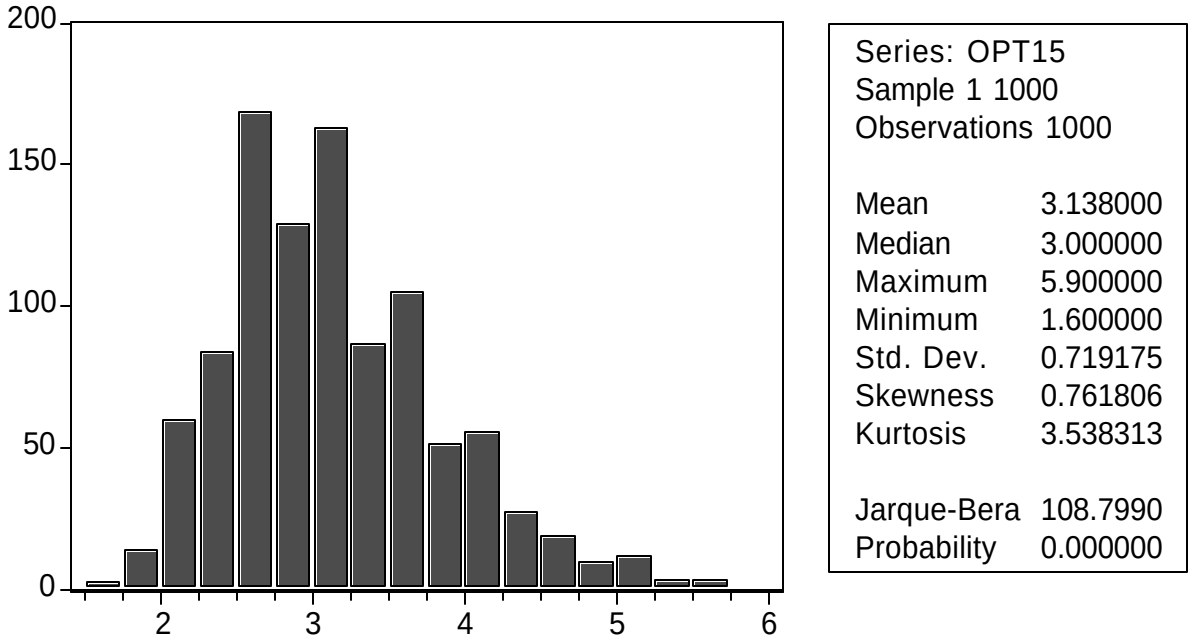

Figure 6. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The table contains the corresponding descriptive statistics.

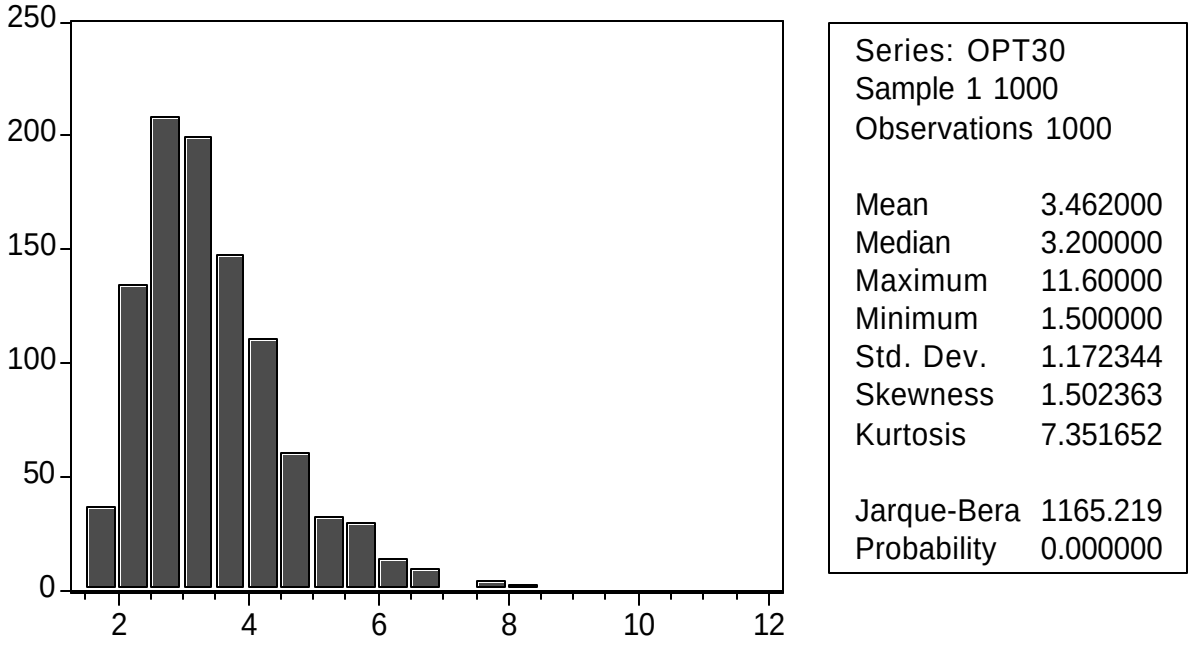

Figure 7. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 30 minute sampling interval. The table contains the corresponding descriptive statistics.

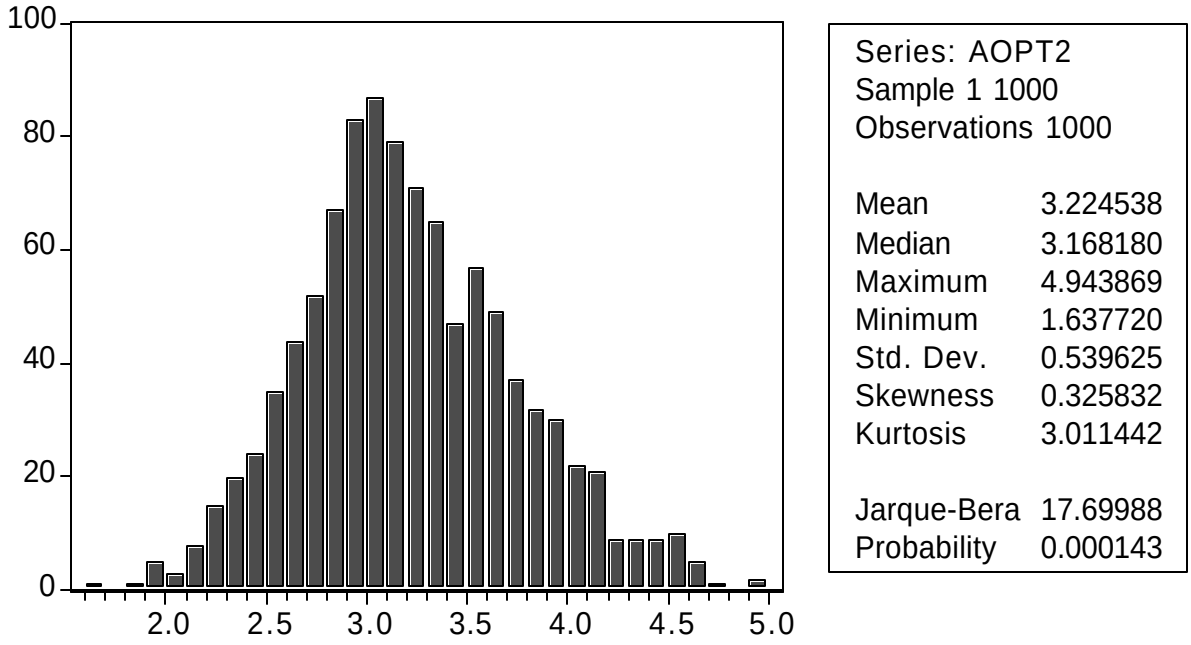

Figure 8. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Section 4 across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 2 minute sampling interval. The table contains the corresponding descriptive statistics.

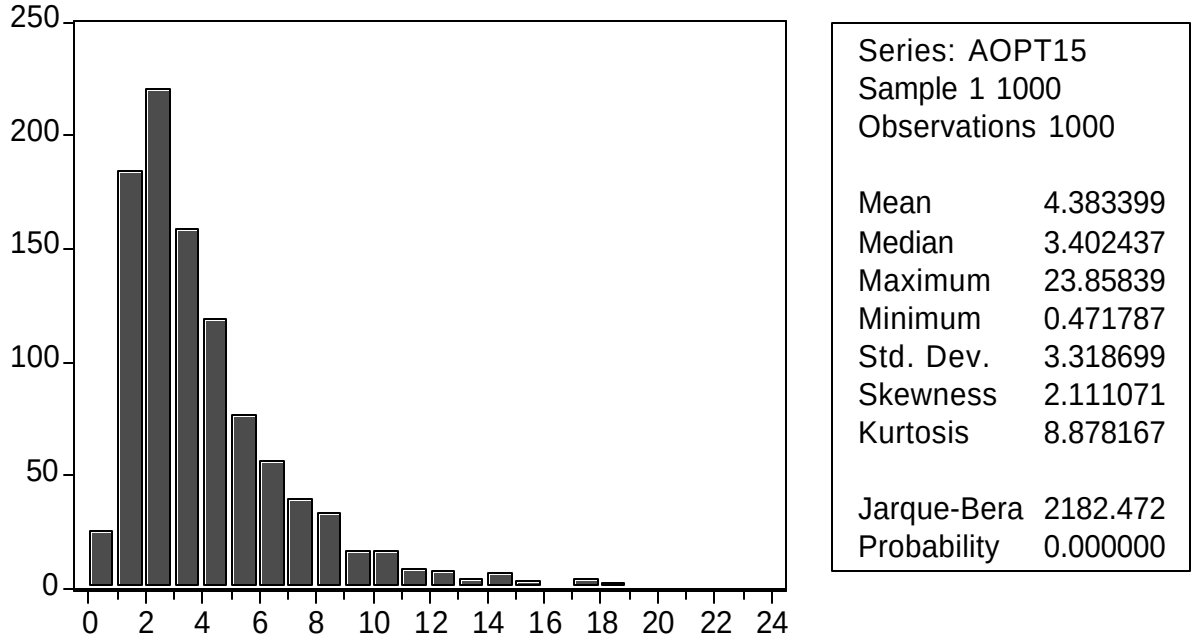

Figure 9. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Section 4 across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The table contains the corresponding descriptive statistics.

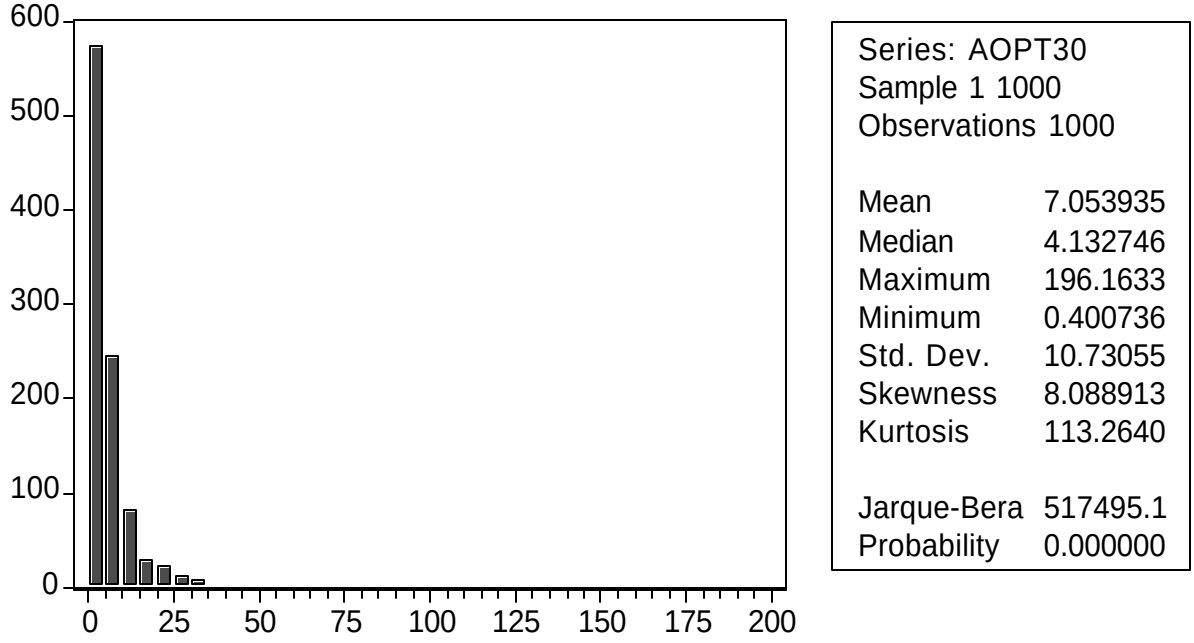

Figure 10. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Section 4 across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 30 minute sampling interval. The table contains the corresponding descriptive statistics.

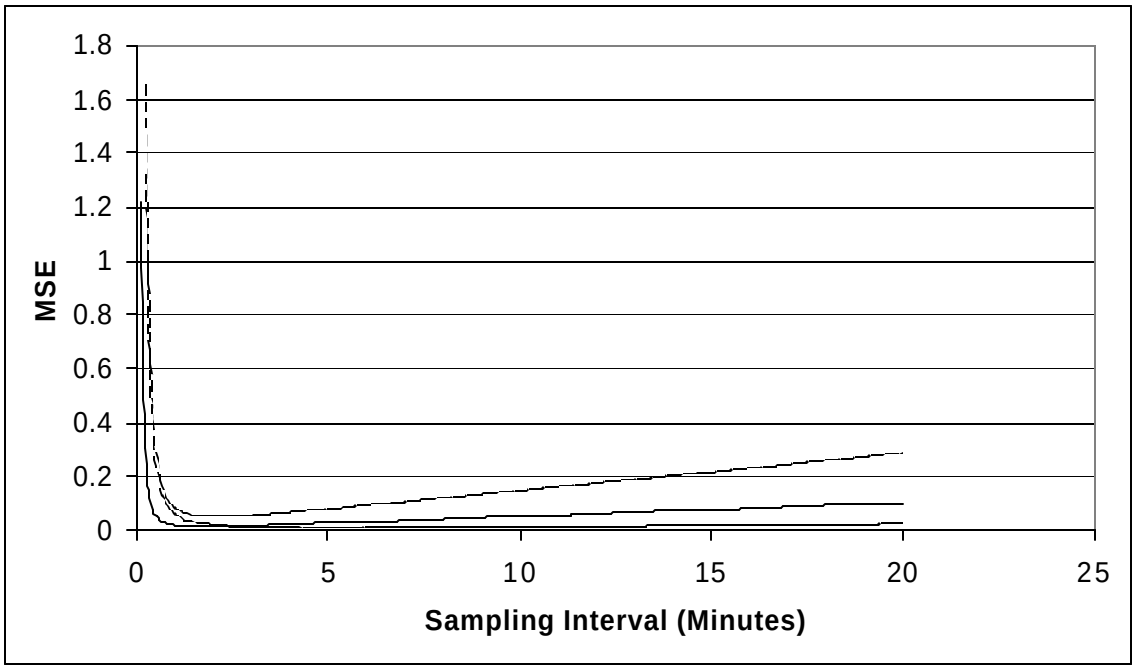

Figure 11. We plot the true MSE expansion in Section 4 and the corresponding 95% bands obtained by implementing the 1,000 simulations in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The moments of the noise process are computed using a 10 second sampling interval.

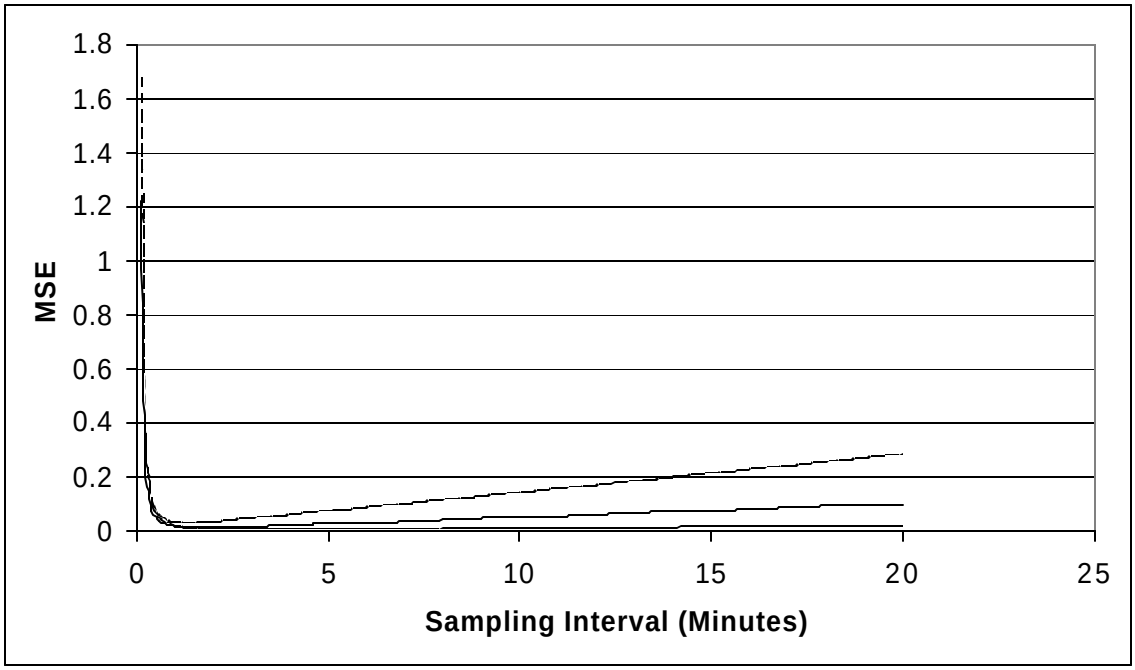

Figure 12. We plot the true MSE expansion in Section 4 and the corresponding 95% bands obtained by implementing the 1,000 simulations in Section 5. The realized quarticity is computed using a 15 minute sampling interval. The moments of the noise process are computed using a 1 second sampling interval.

Symbol	N	Mean Duration	Std. Duration	Min. Duration	Max. Duration
IBM	41841	10.59	11.21	1	334
AA	20356	21.85	25.87	1	931
ARG	5920	75.09	123.28	1	1591

Table 1. Descriptive statistics for IBM, Alcoa (AA), and Airgas (ARG)

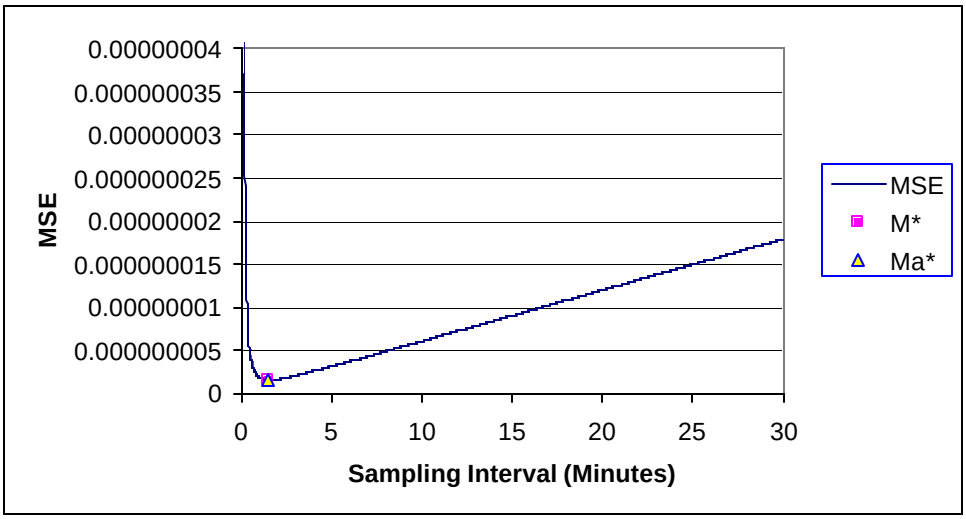

Figure 13. We plot the estimated conditional MSE expansion in Section 4 for the stock IBM. The realized quarticity is computed using a 15 minute sampling interval. M* and Ma* stand for the values corresponding to the true and approximate minima, respectively.

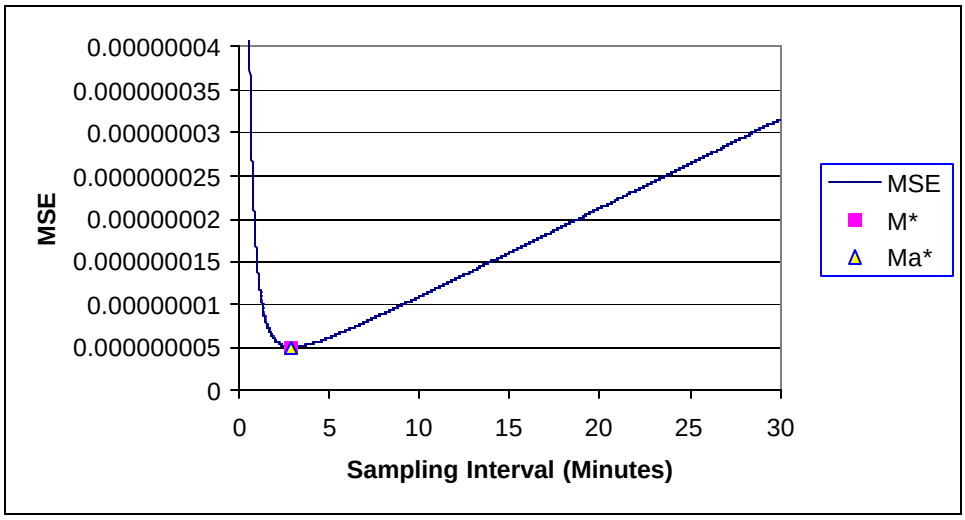

Figure 14. We plot the estimated conditional MSE expansion in Section 4 for the stock Alcoa. The realized quarticity is computed using a 15 minute sampling interval. M* and Ma* stand for the values corresponding to the true and approximate minima, respectively.

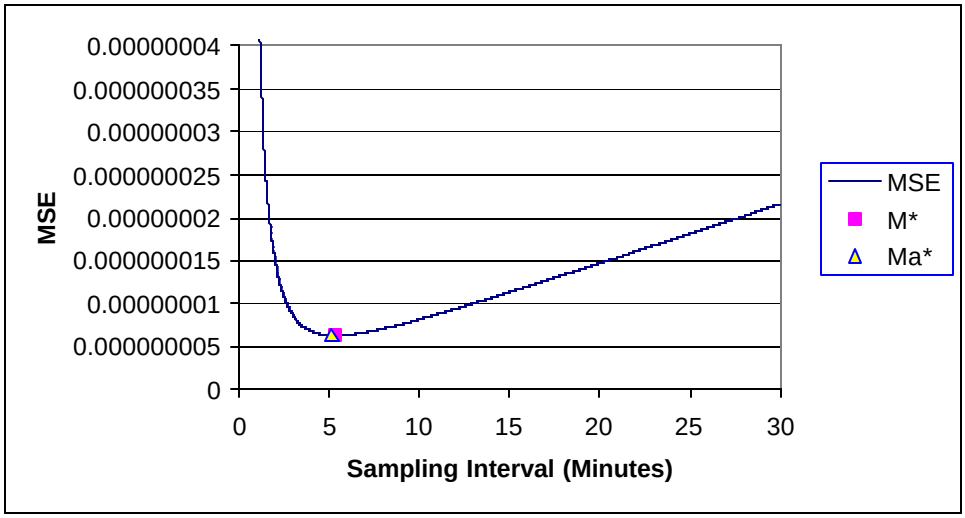

Figure 15. We plot the estimated conditional MSE expansion in Section 4 for the stock Airgas. The realized quarticity is computed using a 15 minute sampling interval. M* and Ma* stand for the values corresponding to the true and approximate minima, respectively.