

Appendix to “Microstructure noise, realized variance, and optimal sampling”

Federico M. Bandi and Jeffrey R. Russell

*Graduate School of Business
University of Chicago*

1 The biases in the realized variance estimates: simulations based on IBM

This section of the Appendix simulates discrete data from a continuous-time stochastic volatility local martingale with $MA(1)$ contaminations using realistic parameter values based on IBM. By virtue of the simulated data the realized variance estimates can be compared to the value of the true integrated variance. Consistently with the results in Theorem 1, we show that realized variance grows without bounds as the sampling interval goes to zero. We also show that, when appropriately standardized, realized variance converges to the variance of the noise process. Finally, for the parameter values used, we show that a small bias is present in the realized variance estimates at sampling intervals around 15 minutes. Nonetheless, the bias can be considerable at higher frequencies. For instance, it is non-negligible at the commonly-employed 5-minute interval.

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

$$d \ln(p_t) = \sigma_t dW_t, \tag{1}$$

with

$$d\sigma_t^2 = \kappa (\bar{v} - \sigma_t^2) dt + \omega \sqrt{\sigma_t^2} dB_t, \quad (2)$$

where $\{W_t, B_t : t \geq 0\}$ is a standard Brownian motion on the plane. Assuming that the η 's are i.i.d. Gaussian completes the specification of the observed return process as given in Eq. (2.4) in the paper.¹

We now turn to parameter selection 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 continuous-time models for stock returns. We normalize the mean variance to unity so that \bar{v} is one. The parameter ω controls the magnitude of the variance of variance and is set equal to .05. Finally, notice that the sample estimates of σ_η and \sqrt{V} (using a 15-minute interval) are equal to .0197% and 1.652%, respectively.² These values imply that the ε 's have a variance equal to .0284% of the average daily variance since $\frac{2(.0197\%)^2}{(1.652\%)^2} = .0284\%$. Hence, to reflect the features of IBM in our sample, we set the standard deviation of the Gaussian noise η equal to $\frac{\sqrt{.0284\%}}{\sqrt{2}}$ (since the average true price variance is normalized to one).

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

Fig. 1 shows the mean realized variance 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., .9957). In agreement with the predictions of Theorem 1, the sharp spike at zero shows that realized variance explodes as the sampling interval goes to zero. Hence, the classical realized variance estimator cannot be a consistent estimate of the integrated variance of the underlying logarithmic price process in the presence of microstructure noise.

¹Without loss of generality, we assume Gaussianity only for simplicity in the simulations. While it is known that alternative, possibly discrete, distributions (such as the multinomial distribution) describe microstructure noise better than the Gaussian distribution, the theoretical results in the paper are fully robust to alternative distributional assumptions.

²The numbers reflect a slightly different sample from the one employed in the paper. We only use NYSE quotes here. The sample period is the same as in the paper.

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 .9957 whereas the value of the average realized variance 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 “variance signature plots,” using the terminology in Andersen et al. (1999, 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 the paper.

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 which is 4 times as large as that in Fig. 2 (i.e., .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 value 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.

Theorem 2 in the paper indicates that a re-scaled version of the quadratic variation estimator should converge to the variance of the noise process. Fig. 5 represents the re-scaled realized variance, namely $\frac{\hat{V}}{M}$, for different sampling frequencies and a value of the ratio between variance of the noise process and underlying quadratic variation equal to .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 re-scaled realized variance estimates converge to the second moment of the ε 's as the

sampling interval goes to zero. 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). We confirm again the validity of the predictions contained in Theorem 2.

2 The biases in the quarticity estimates and their impact on the conditional MSE of the realized variance estimator: simulations based on IBM

In this section of the Appendix we show that alternative, credible sampling frequencies used to compute the underlying quarticity have little impact on the sampling distribution of the optimal number of observations needed to calculate realized variance. In addition, we show that, when the quarticity is estimated relatively accurately, the rule-of-thumb in Remark 3 in the paper delivers a distribution of the estimated optimal sampling frequencies which 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 which are more biased, and considerably more volatile, than 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 just be 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 variance 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. (1999, 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 (2006a)). Naturally, as shown in Fig. 7, such choice is 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. (4.1) for values of the quarticity estimates obtained by sampling at the correct 2-minute interval. 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 which has been widely used in 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 variance estimator. Finally, for the range of values in Fig. 8, the incremental impact on the MSE of the realized variance estimator of lowering the sample frequency is rather small. The value of the MSE at the optimal 1.7-minute frequency is .014. It is .0143 at the 2-minute interval and .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 which 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 variance 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 sampling frequencies higher than 15 minutes are 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 (2006b) confirm this finding). Furthermore, it is an easy interval to use. Therefore, it deserves attention in applied work. Consistently, the 15-minute frequency is the frequency that we utilize in the empirical application in the

paper.

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 distribution on the optimal sampling frequencies for the realized variance estimator obtained by employing the rule-of-thumb in Remark 3 in the paper. 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 sampling frequency for the quarticity. In effect, the approximation provided by Remark 3 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 is still informative about the magnitude of the optimal realized variance's sampling frequency.

In Figs. 14 and 15 we plot the conditional MSE, as implied by Eq. (4.1) in the paper, and corresponding 95% bands based on 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. In agreement 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. Naturally, stocks whose price updates occur more frequently should lead to more accurate evaluations of our proposed MSE expansions.

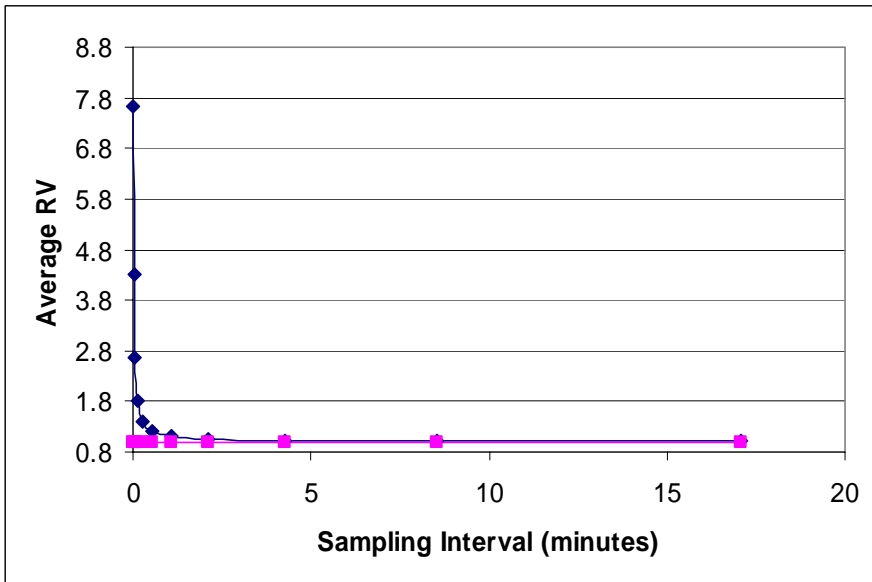


Figure 1. We plot the average realized variance across the 1,000 simulations described in Section 1. The horizontal line denotes the known quadratic variation for the day (.9957).

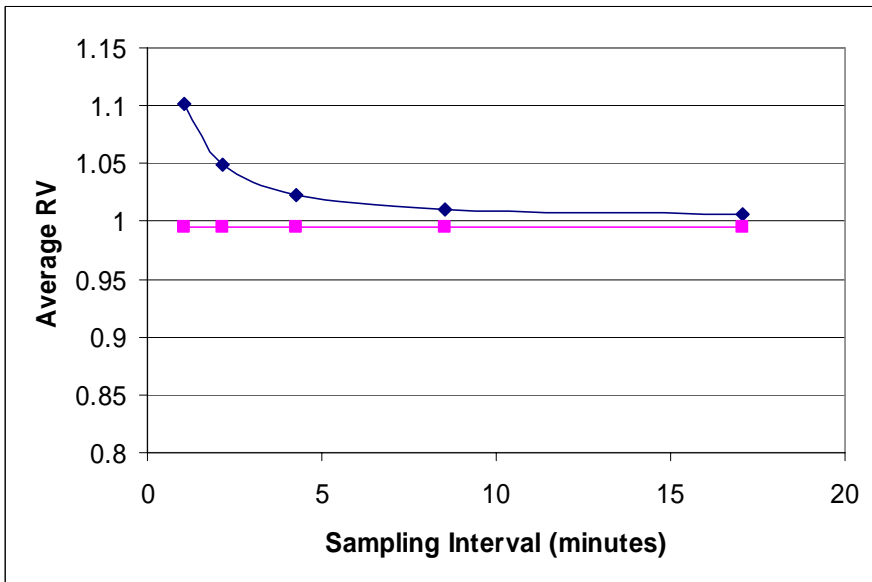


Figure 2. We plot the average realized variance across the 1,000 simulations described in Section 1. The horizontal line denotes the known quadratic variation for the day (.9957). This plot is the same as the plot in Figure 1 but on a different scale.

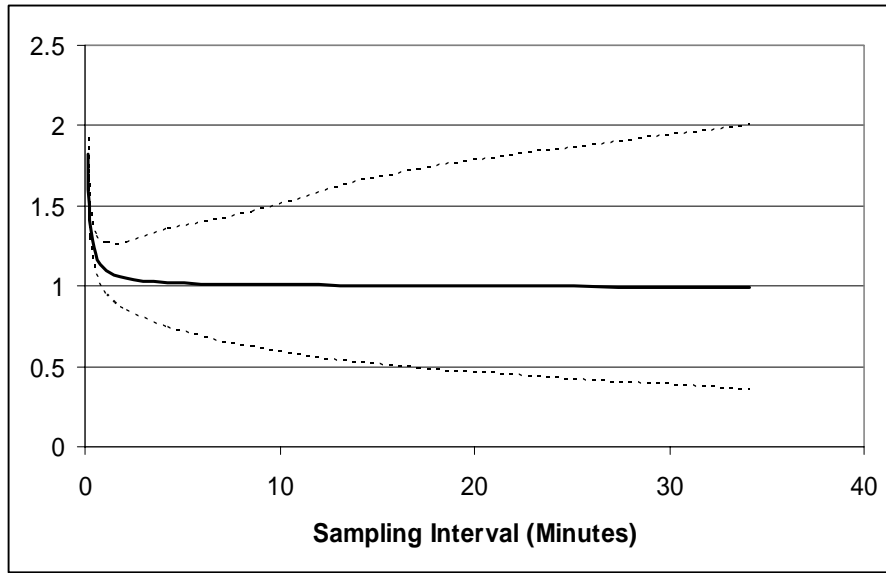


Figure 3. We plot the average realized variance and 95% empirical bands computed across the 1,000 simulations described in Section 1.

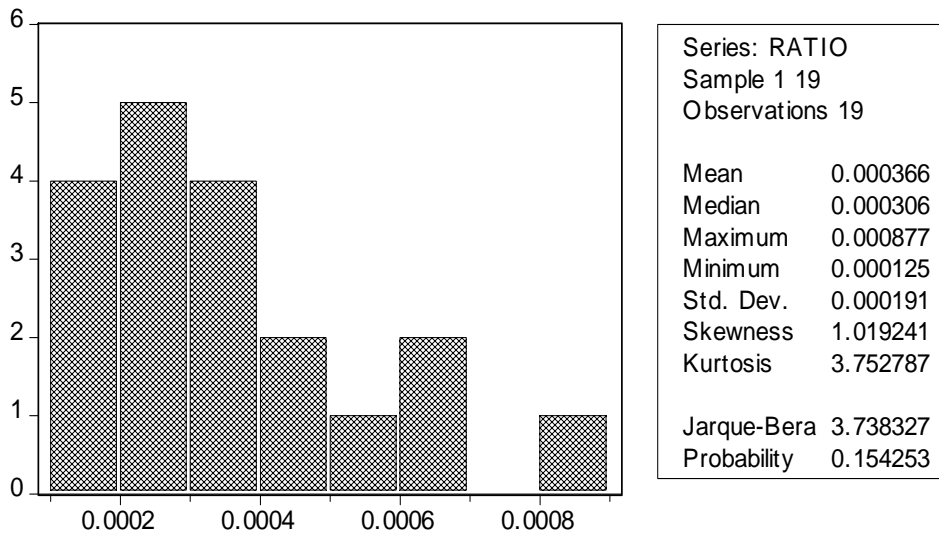


Figure 4. We plot the histogram of the ratio between the estimated second moment of the noise process and the daily realized variances (computed over a 6.5 hour period). The table contains the corresponding descriptive statistics.

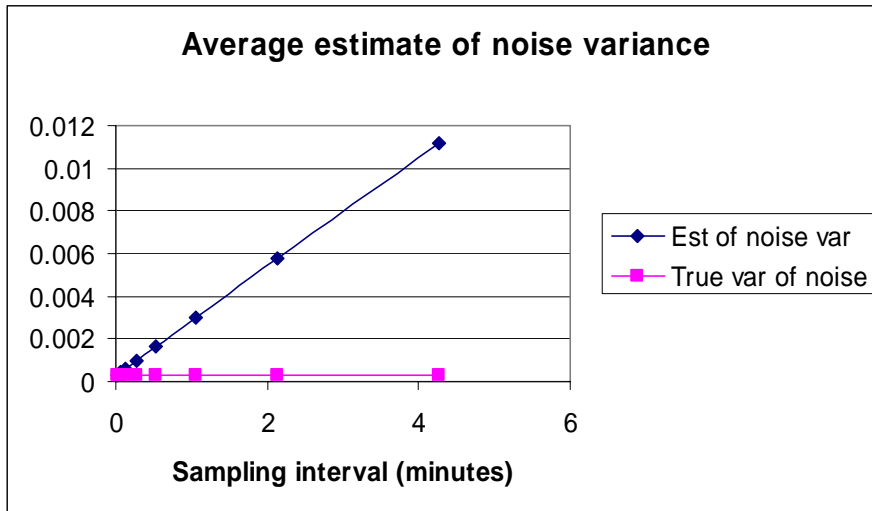


Figure 5. We plot the average of the standardized (by the number of observations) realized variance estimates across the 1,000 simulations described in Section 1. The horizontal line denotes the known second moment of the noise process.

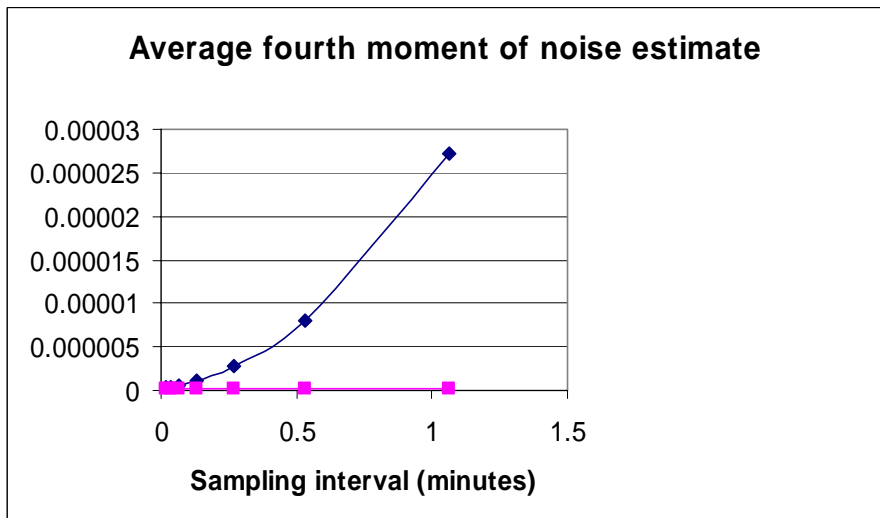


Figure 6. We plot the average of the fourth moment estimates across the 1,000 simulations described in Section 1. The horizontal line denotes the known fourth moment of the noise process.

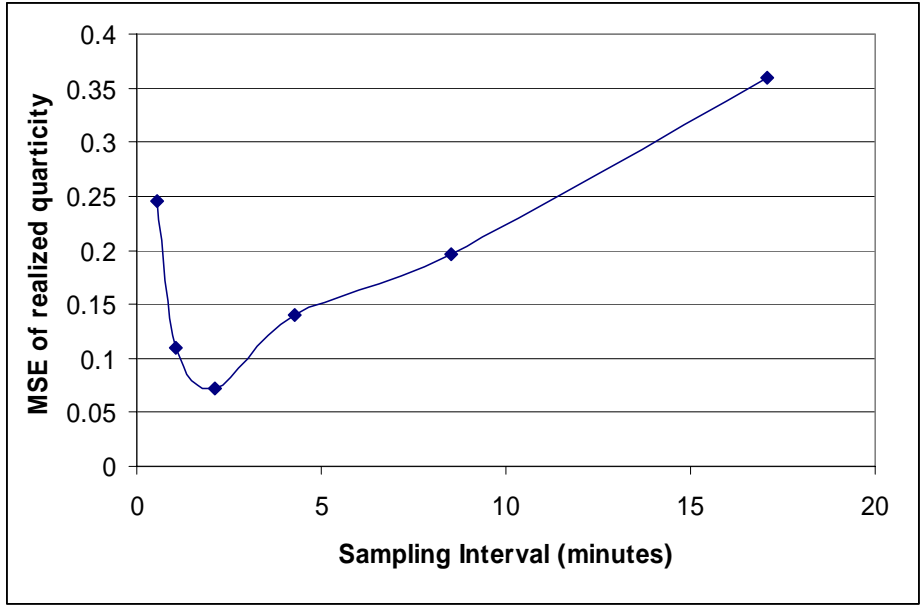


Figure 7. We plot the simulated conditional MSE of the realized quarticity. We use 1,000 simulations as described in Section 1.

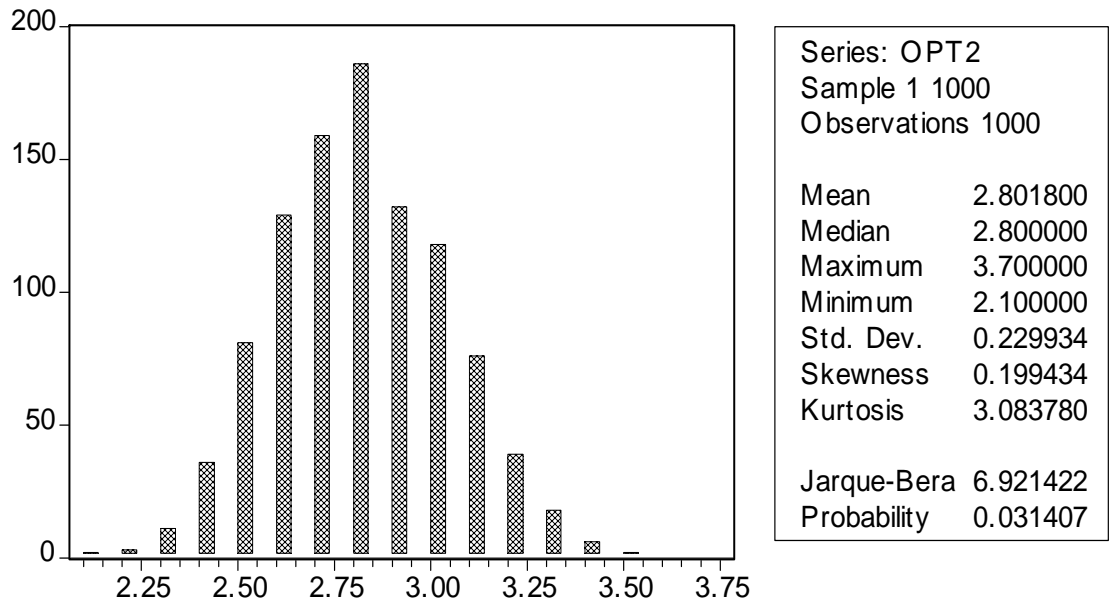


Figure 8. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 1. The realized quarticity is computed using a 2-minute sampling interval. The table contains the corresponding descriptive statistics.

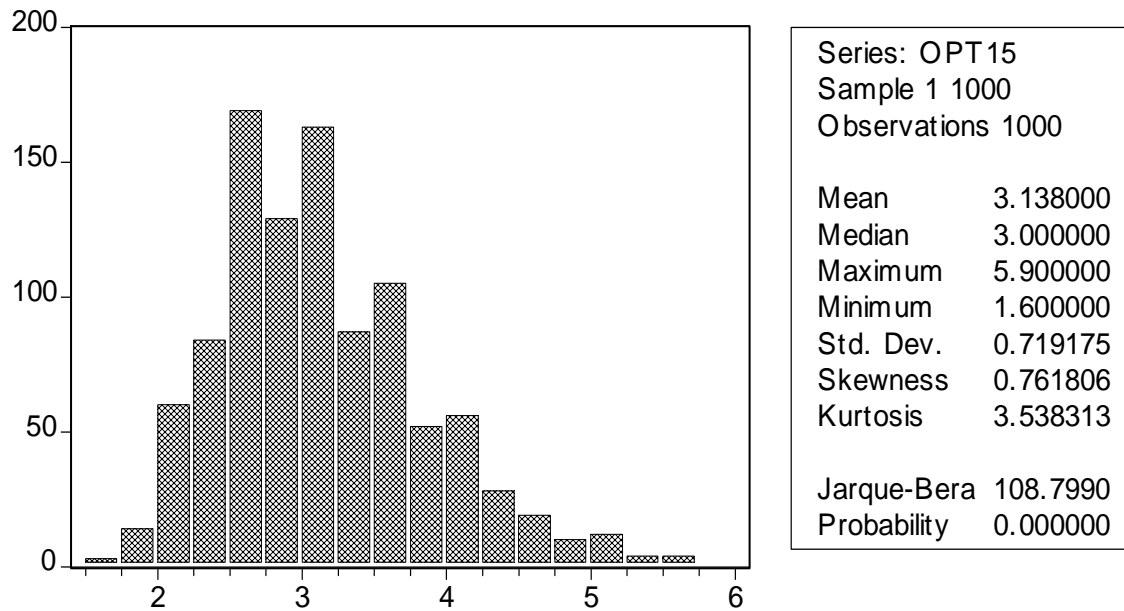


Figure 9. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 1. The realized quarticity is computed using a 15-minute sampling interval. The table contains the corresponding descriptive statistics.

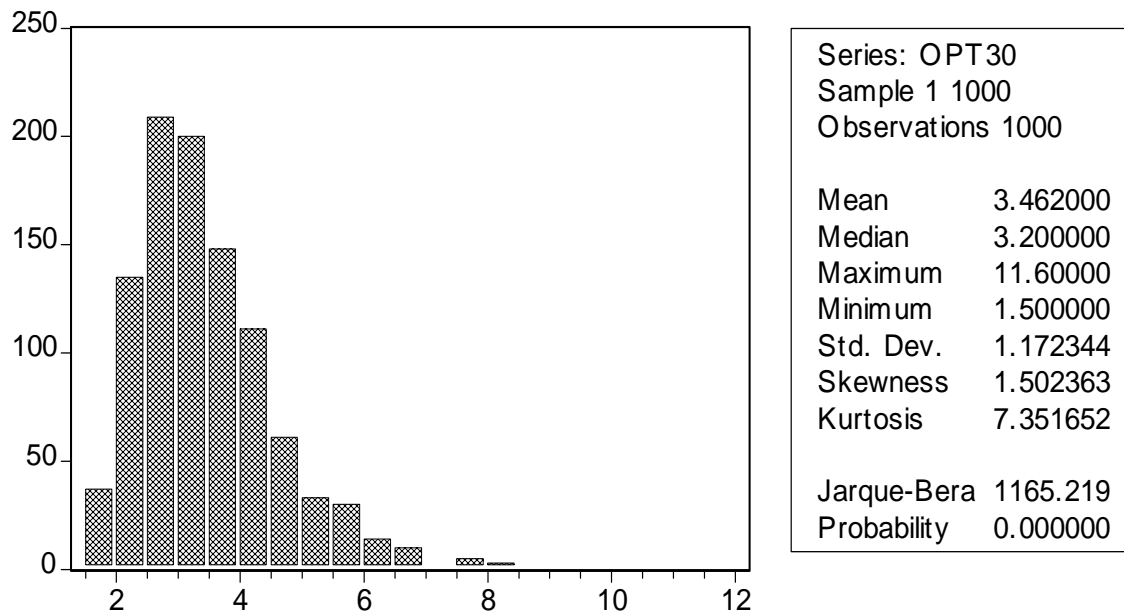


Figure 10. We plot the distribution of the optimal sampling frequencies across the 1,000 simulations described in Section 1. The realized quarticity is computed using a 30-minute sampling interval. The table contains the corresponding descriptive statistics.

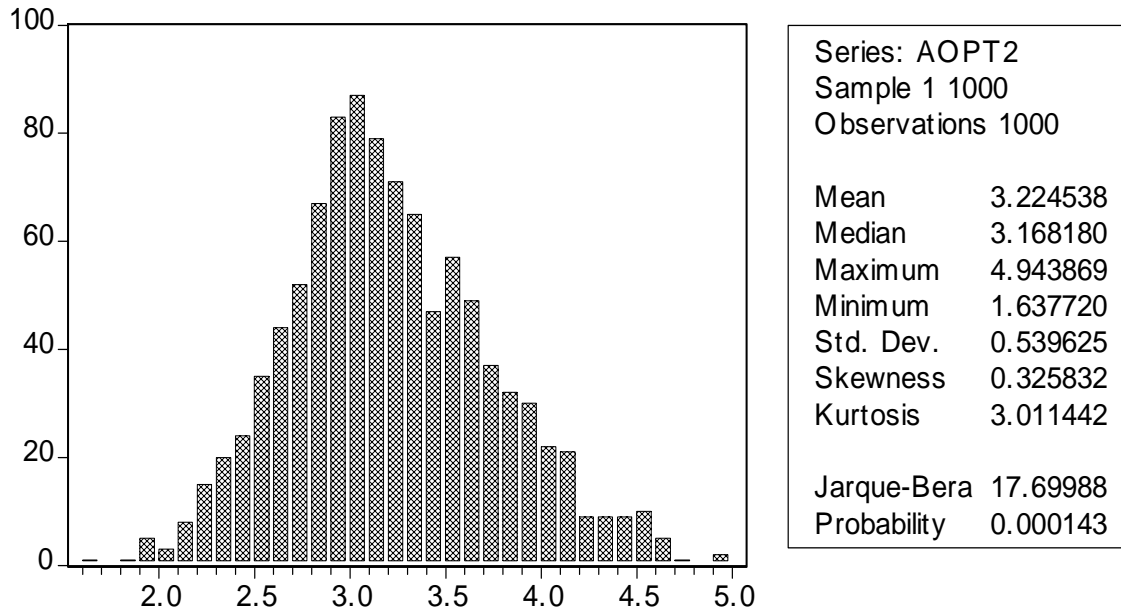


Figure 11. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Remark 3 in the paper across the 1,000 simulations described in Section 1. The realized quarticity is computed using a 2-minute sampling interval. The table contains the corresponding descriptive statistics.

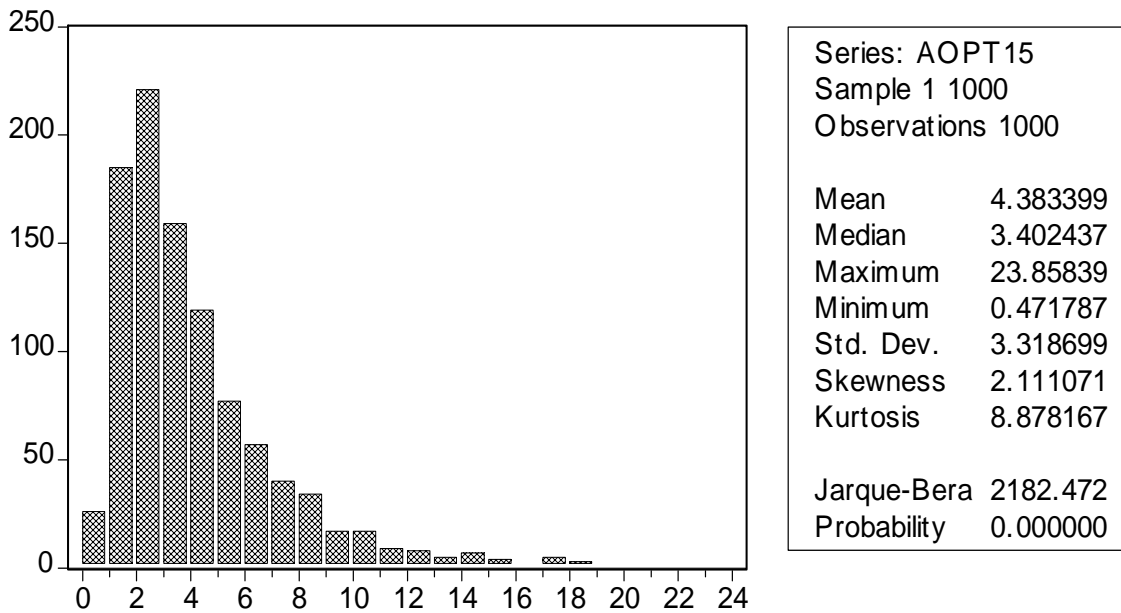


Figure 12. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Remark 3 in the paper across the 1,000 simulations described in Section 1. The realized quarticity is computed using a 15-minute sampling interval. The table contains the corresponding descriptive statistics.

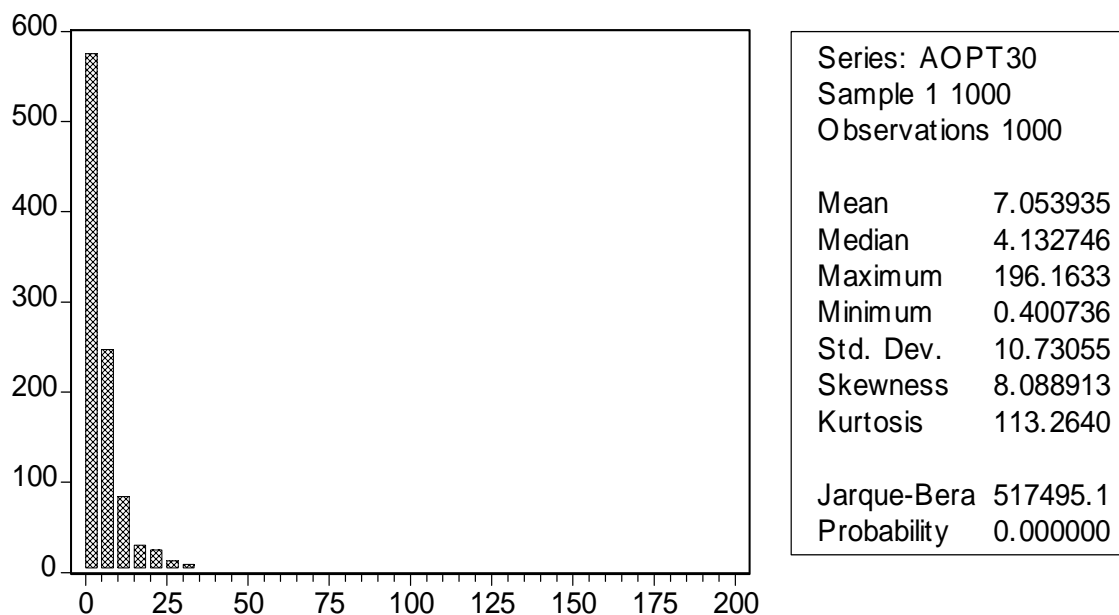


Figure 13. We plot the distribution of the optimal sampling frequencies obtained by using the rule-of-thumb in Remark 3 in the paper across the 1,000 simulations described in Section 5. The realized quarticity is computed using a 30-minute sampling interval. The table contains the corresponding descriptive statistics.

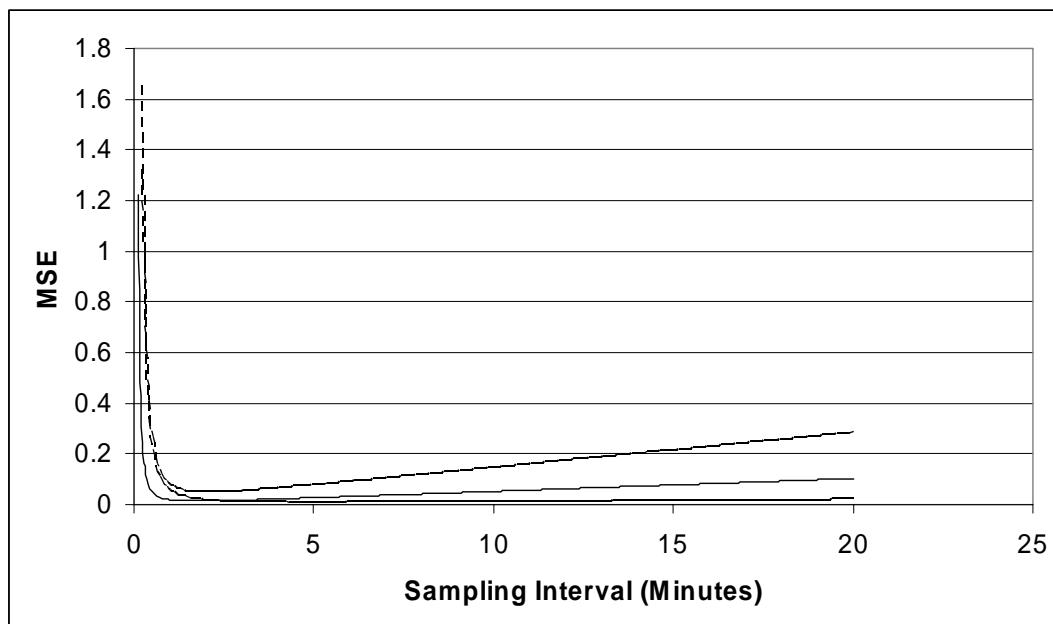


Figure 14. We plot the estimated MSE expansion in Theorem 3 in the paper and the corresponding 95% bands obtained by implementing the 1,000 simulations in Section 1. The realized quarticity is computed using a 15-minute sampling interval. The moments of the noise process are computed using a 10-second sampling interval.

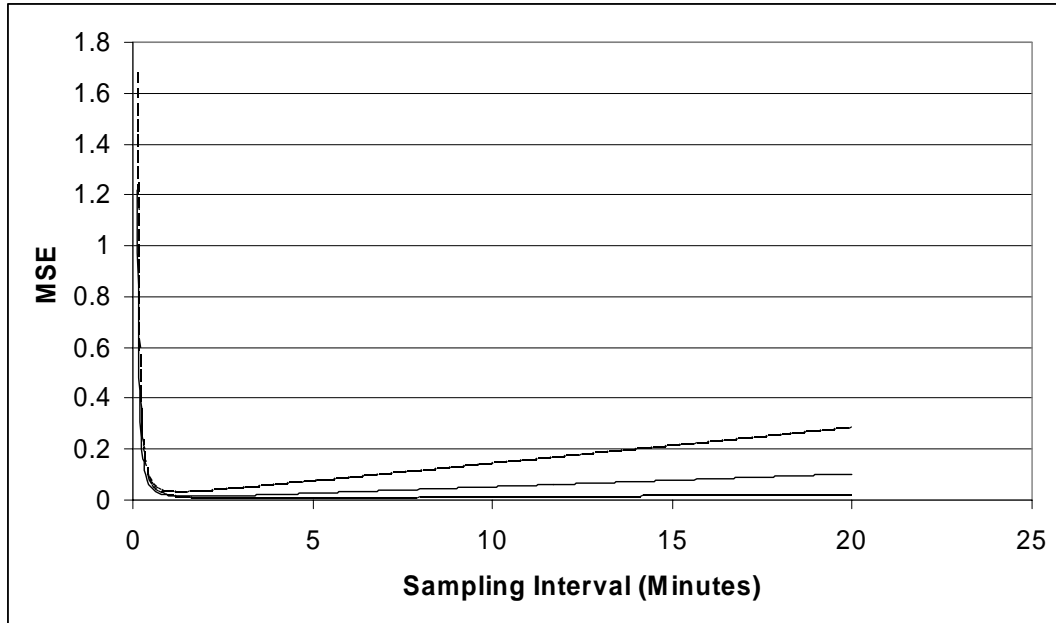


Figure 15. We plot the estimated MSE expansion in Theorem 3 in the paper and the corresponding 95% bands obtained by implementing the 1,000 simulations in Section 1. The realized quarticity is computed using a 15-minute sampling interval. The moments of the noise process are computed using a 1-second sampling interval.