MICROSTRUCTURE NOISE, REALIZED VARIANCE, AND OPTIMAL SAMPLING*

By Federico M. Bandi† and Jeffrey R. Russell‡

First draft: November 2003
This draft: July 2007

Abstract

A recent and extensive literature has pioneered the summing of squared observed intra-daily returns, "realized variance," to estimate the daily integrated variance of financial asset prices, a traditional object of economic interest. We show that, in the presence of market microstructure noise, realized variance does not identify the daily integrated variance of the frictionless equilibrium price. However, we demonstrate that the noise-induced bias at very high sampling frequencies can be appropriately traded off with the variance reduction obtained by high frequency sampling and derive an MSE optimal sampling theory for the purpose of integrated variance estimation. We show how our theory naturally leads to an identification procedure which allows us to recover the moments of the unobserved noise; this procedure may be useful in other applications. Finally, using the profits obtained by option traders on the basis of alternative variance forecasts as our economic metric, we find that explicit optimization of realized variance's finite sample MSE properties results in accurate forecasts and considerable economic gains.

Keywords: Realized variance, Quadratic variation, Market microstructure noise, Optimal sampling, Option pricing

JEL Classification: G12, C14, C22

---

*We are grateful to Tim Conley, Rob Engle, Neil Shephard, two anonymous referees, and Bernard Selanié, the Editor, for valuable comments and suggestions. We thank the participants at the CIRANO conference "Realized Volatility," Montreal, November 7-8, 2003, the conference "Econometric Forecasting and High-Frequency Data Analysis," Singapore, May 7-8, 2004, and seminar participants at various institutions for helpful discussions. We thank Jean Jacod for sending us a copy of his unpublished 1994 manuscript.

†Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637. E-mail: federico.bandi@gsb.uchicago.edu.

‡Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637. E-mail: jeffrey.russell@gsb.uchicago.edu.

1. INTRODUCTION

THE RECENT AVAILABILITY of quality high-frequency financial data has motivated a growing literature devoted to the model-free measurement of variance (see the review paper by Andersen et al. (2002) and the references therein). The main idea is to aggregate squared intra-daily returns to approximate the daily increments of the quadratic variation of the semimartingale that drives the underlying logarithmic price process. The consistency result justifying this procedure is the convergence in probability of the sum of squared returns to the daily increment of the quadratic variation of the logarithmic price process as returns are computed over intervals that are increasingly small asymptotically. This result is a cornerstone in semimartingale process theory (see, e.g., Chung and Williams (Theorem 4.1, page 76, 1990)). More recently, a nonparametric theory of inference for variance estimation has championed empirical implementation of these ideas and has termed these estimates "realized variance" (Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002), BN-S hereafter).

The theoretical validity of the procedure hinges on the observability of the true price process. However, the fine grain market dynamics (i.e., "the market microstructure") generate a divergence between the observed price process and the true or "frictionless equilibrium" price process, whose quadratic variation is the object of interest. This divergence could, for example, be induced by transaction price changes occurring as multiples of ticks (price discreteness) or by the existence of multiple prices for buyers and sellers (bid-ask spreads). It may be due to liquidity or information reasons (see, e.g., Stoll (2000)). In agreement with the extant literature, we denote the deviations of observed prices from frictionless equilibrium prices by "market microstructure noise" and the frictionless equilibrium price by "true price" or "equilibrium price."

The present paper provides a general treatment of market microstructure noise on realized variance estimates. Specifically, we consider both asymptotic and finite sample effects of noise.

Under our assumed price formation mechanism, we show that the realized variance estimates are asymptotically swamped by noise as the number of squared return data increases over a fixed time period. The theoretical manifestation of this effect is a realized variance estimator which fails to converge to the increment of the quadratic variation (or "integrated variance") of the underlying logarithmic true price process but, instead, increases without bound almost surely over any fixed period of time, however small. Hence, our results provide a theoretical justification for the diverging behavior at high frequencies of the realized variance estimates as shown by Andersen et al. (1999, 2000) using "volatility signature plots," namely plots of realized volatility versus alternative sampling frequencies. For illustration, Fig. 1 contains the (averaged across days) variance signature plot for the IBM data used in this paper. As implied by our asymptotic theory, the plot is upward sloping at high frequencies.

Figure 1 about here

Interestingly, despite the fact that realized variance is not consistent for the conventional object of interest (i.e., increments of quadratic variation or integrated variance), we show that a standardized version of the realized variance estimator can be employed to identify the variance of the unobservable *noise* component. More generally, we show that sample moments of the noisy return data can be put to work to identify moments of the underlying noise process. This contribution of the paper is important for the empirical analysis of the bias and variance of the realized variance estimator since, naturally, both quantities depend on the distributional features of the unobserved noise component in the recorded returns. Moreover, since the noise dictates deviations of observed prices from equilibrium prices, the noise moments may be of direct economic interest.

Our finite sample results begin with a characterization of the bias/variance trade-off induced by the presence of microstructure noise. When the true price process is observable, higher sampling frequencies over a fixed period of time result in more precise estimates of the integrated variance of the logarithmic price process. When the true price process is not observable, as is the case in the presence of microstructure frictions, frequency increases provide information about the underlying integrated variance but, necessarily, entail accumulation of noise affecting both the bias and the variance of the estimator. The sampling frequency can be chosen optimally to balance these two contrasting effects. Specifically, the bias/variance trade-off can serve as the basis for an optimal sampling theory for nonparametric variance estimates in the presence of microstructure noise. Under independence of the noise, we first derive an expression for the conditional (on the underlying volatility path) mean-squared-error (MSE) of the contaminated realized variance estimator as a function of the sampling frequency. A robust methodology is then provided to optimally choose the sampling frequency as the minimum of the conditional MSE. The method relies on the computation of sample moments of the contaminated high-frequency return data. As such, it is simple to implement. Subsequently, we study several extensions of our proposed methodology. First, we consider a bias-corrected realized variance estimator and discuss optimal sampling for the purpose of finite sample variance minimization. Second, we evaluate optimal sampling for estimating nonlinear functions of increments of quadratic variation. Finally, we relax the assumption of independent noise made initially and study optimal sampling under more general noise dependence properties.

We apply our methods to a sample of IBM quote data and find optimal sampling intervals that are shorter than the intervals typically implemented and/or conjectured in the literature. Our new intervals translate into considerable *statistical gains*. The last part of the paper employes the profits obtained from option trading on the basis of alternative variance forecasts as the economic metric used to evaluate the relative benefits of time-varying optimal (in an MSE sense) sampling intervals versus fixed, ad-hoc intervals. We show that our proposed time-varying optimal intervals result in improved variance forecasts and pervasive *economic gains*.

Despite general awareness of the necessity of accounting for price contaminations in high-frequency data for the purpose of integrated variance estimation (see, e.g., the discussions in Andersen et al. (2001) and BN-S (2002)), at the time of this paper's original writing most existing treatments of the effects of market microstructure noise on realized variance estimates were rather informal. There were important exceptions. These either imposed restrictions that we wished to avoid or failed to be operational in practise. In some cases, the focus was different.

Zhou (1996) proposes choosing the optimal sampling frequency of a bias-corrected realized variance measure by minimizing its variance expansion. Zhou's framework hinges on the i.i.d. nature of the noise components in the price process, on the Gaussianity of the noise, and on a constant spot price volatility over individual periods. Bai et al. (2004) derive an MSE expansion for variance estimates in the presence of dependent noise. However, they take a less structural approach than we do and are therefore unable to independently identify the two variance components of the observed returns, i.e., the equilibrium price variance and the variance of the microstructure noise. Consistent with Zhou (1996) and under similar assumptions, Aït-Sahalia et al. (2005) derive a closed-form expression for the unconditional MSE of a constant variance estimator. This paper's focus, however, is on important efficiency gains from maximum likelihood estimation of parametric diffusion models with market microstructure noise. Parametric likelihood-based identification in the presence of noise is also studied in work by Gloter and Jacod (2001a, 2001b). Oomen (2002) uses a structural model of price formation to provide simulated MSE plots for noisy quadratic variation estimates as a function of the sampling interval. He allows for a stochastic variance but resorts to simulations in the absence of a closed-form specification for the relation between the relevant MSE and the sampling frequency. There exists a related literature in probability theory which studies the identification of functionals of nearly observed continuous-time processes. Unobservability is generally either due to deterministic round-off errors or to Gaussian stochastic noise (see Delattre and Jacod (1997) and Picard (1993), for example). Contrary to the present paper and the papers cited previously, limiting results are obtained for contaminations that are assumed to be small asymptotically. While interesting from a theoretical perspective, these approaches are not appropriate for our observed price formation mechanism involving microstructure noise effects. Not surprisingly, our price formation mechanism leads us to a different approach to identification.

In contemporaneous and independent work, Zhang et al. (2005) have also derived a conditional MSE expansion for realized variance in the presence of i.i.d. noise, one of the theoretical contributions in the present paper. Under the same noise properties, they propose a methodology to obtain consistent estimates of integrated variance. Their suggested method relies on averaging (bias-corrected) realized variance estimates sampled at a specific frequency (i.e., it relies on "subsampling" in their terminology). While our focus is on a general study of the asymptotic and finite sample properties of the classical realized variance estimator in the presence of noise, their focus is on the asymptotic consistency properties of a novel approach to integrated variance estimation. A promising bias-corrected realized variance measure allowing for dependent noise and some form of dependence between noise and equilibrium price has been recently advocated by Hansen and Lunde (2006). As in Zhou (1996), their estimator is a kernel-based estimator. The usefulness of a rich class of kernel estimates of integrated variance has been further studied in recent work by Barndorff-Nielsen et al. (2006).

Since this paper's original draft, stimulating developments of optimal sampling methods have been proposed for a variety of issues in nonparametric variance estimation in the presence of noise-contaminated high-frequency data. We mention a few important contributions. Hansen and Lunde (2006) discuss MSE-based sampling of bias-corrected realized variance estimates. Oomen (2006) study MSE-based sampling for the purpose of evaluating the preferability of calendar time sampling versus business time sampling. Oomen (2005) study MSE-based optimal sampling in calendar time and business time for bias-corrected realized variance estimates. Bandi and Russell (2006d) propose MSE-based rules for selecting the optimal number of autocovariances (or subsamples) to be used in the definition of kernel estimates of integrated variance as in Zhang et al. (2005) and Barndorff-Nielsen et al. (2006). Mancino and Sanfelici (2006) derive MSE expansions for Fourier estimators of integrated variance. Andersen et al. (2006) and Ghysels and Sinko (2006) study optimal sampling for the purpose of realized variance forecasting. The research on high-frequency variance estimation and noise has been particularly vibrant in recent years. For extensive reviews, we refer the interested reader to Bandi and Russell (2006c), Barndorff-Nielsen and Shephard (2006), and McAleer and Medeiros (2006).

The paper proceeds as follows. In Section 2 we lay out the model. Section 3 is about the limiting properties of the realized variance estimator when microstructure noise affects asset prices. In Section 4 we present an expansion of the conditional MSE of the realized variance estimator when i.i.d. noise plays a role and discuss optimal sampling by virtue of MSE minimization. In Section 5 we study extensions of the method. Specifically, we discuss bias-correcting, optimal sampling for the purpose of estimating nonlinear functions of integrated variance, and optimal sampling in the presence of richer noise dependence properties. Section 6 contains simulations. In Section 7 we use quote-to-quote IBM price changes and apply our methods to the estimation of the integrated variance of the logarithmic price process and the second moment of the unobservable noise process. Section 8 discusses the economic benefits of variance forecasting using optimal sampling. We do so in the context of option pricing and option trading. Section 9 concludes. Proofs, technical details, and a glossary of notation are contained in the Appendixes.

2. THE MODEL

We introduce microstructure noise effects in the context of a model that is consistent with previous theoretical approaches to model-free variance estimation (see, e.g., BN-S (2002, 2003, 2004)). For convenience, we use similar notation as in BN-S (2002, 2003, 2004).

We consider a fixed time period h (a trading day, for instance) and write the observed price process at the end of the i-th period as

$$\widetilde{p}_{ih} = p_{ih}\vartheta_{ih} \qquad i = 1, 2, ..., n, \tag{2.1}$$

where p_{ih} is the frictionless equilibrium price, i.e., the price that would prevail in the absence of market microstructure frictions, and ϑ_{ih} denotes microstructure noise. A simple logarithmic transformation gives us

$$\underbrace{\ln\left(\widetilde{p}_{ih}\right) - \ln\left(\widetilde{p}_{(i-1)h}\right)}_{\widetilde{r}_i} = \underbrace{\ln(p_{ih}) - \ln(p_{(i-1)h})}_{r_i} + \underbrace{\eta_{ih} - \eta_{(i-1)h}}_{\varepsilon_i} \qquad i = 1, 2, ..., n, \tag{2.2}$$

where $\eta = \ln(\vartheta)$. We can now divide each period into M sub-periods and define the observed high-frequency continuously-compounded returns as

$$\widetilde{r}_{j,i} = \ln\left(\widetilde{p}_{(i-1)h+j\delta}\right) - \ln\left(\widetilde{p}_{(i-1)h+(j-1)\delta}\right) \qquad j = 1, 2, ..., M, \tag{2.3}$$

where $\delta = h/M$ is the time distance between adjacent logarithmic prices or, equivalently, the time horizon over which the continuously-compounded returns are computed. Hence, $\widetilde{r}_{j,i}$ is the j-th intra-period return over the i-th period. More precisely,

$$\widetilde{r}_{j,i} = r_{j,i} + \varepsilon_{j,i}, \tag{2.4}$$

where $r_{j,i}$ and $\varepsilon_{j,i}$ $(= \eta_{(i-1)h+j\delta} - \eta_{(i-1)h+(j-1)\delta})$ have straightforward interpretations in terms of equilibrium return and microstructure contamination in the return data, respectively. Both the equilibrium return $r_{j,i}$ and the microstructure noise contamination $\varepsilon_{j,i}$ are unobservable. The econometrician only observes the noisy return data $\widetilde{r}_{j,i}$. To simplify the notation, we suppress the subscript i and deal with the case $i = 1$. Thus, we write j in place of the double index j, i. Our discussion will then apply to each period h from 1 to n without loss of generality.

Our interest will be in characterizing the asymptotic and finite sample properties of the classical realized variance estimator $\widehat{V} = \sum_{j=1}^{M} \widetilde{r}_j^2$. As in Andersen et al. (2003) and BN-S (2002), among others, $\widehat{V}$ will be used to estimate the integrated variance of the logarithmic price process over each period. For clarity, we define this quantity formally after a discussion of the assumptions that we impose on the logarithmic price process $(\ln(p))$ and the microstructure contaminations in the price process (the η's), respectively.

Assumption 1: (The Price Process)
The logarithmic price process $\ln(p_t)$ *is a continuous stochastic volatility Brownian semimartingale. Specifically,*

$$\ln(p_t) = \ln(p_0) + \int_0^t \phi_s ds + \int_0^t \sigma_s dW_s, \tag{2.5}$$

where ϕ_t *is a continuous predictable drift process,* σ_t *is a càdlàg spot volatility process, and* W_t *is a standard Brownian motion.*

Assumption 2: (The Microstructure Noise)
The microstructure frictions in the price process $\eta_j's$ *are i.i.d. mean zero with a finite fourth moment.*

The object of econometric interest is the bounded integrated variance of the equilibrium price process over each period, i.e., $V = \int_0^h \sigma_s^2 ds$.

Assumption 1 implies that the equilibrium return process evolves in time as a stochastic volatility martingale difference plus an adapted process of finite variation.[1] The stochastic spot volatility is allowed to display jumps, diurnal effects, high-persistence (possibly of the long-memory type), and nonstationarities.[2] Dependence between σ and the Brownian motion W (i.e., leverage effects) is allowed.

Assumption 2 implies that the observed returns display an $MA(1)$ structure with a negative first-order autocorrelation. The $MA(1)$ market microstructure model in returns (or the i.i.d. market microstructure model in prices) is typically justified by bid-ask bounce effects (see, e.g., Roll (1984)). It is, in general, an approximation. However, it is a credible approximation in decentralized markets where traders arrive in a random fashion with idiosyncratic price setting behavior, the foreign exchange market being an important example (see, e.g., Bai et al. (2004)). It can also be a valid approximation in the case of equities when considering transaction prices and/or quotes posted on multiple exchanges (see Bandi and Russell (2006b)). Awartani et al. (2004) propose a formal test of the $MA(1)$ market microstructure model. In Section 5 we extend the model to a richer noise dependence structure. Section 7 shows the empirical validity of the $MA(1)$ model for the sample of IBM high-frequency price data used in the present study.

Importantly, while the equilibrium return process is $O_p\left(\sqrt{\delta}\right)$ over any intra-period time horizon of size $\delta = \frac{h}{M}$, the contaminations in the return process are $O_p(1)$. This result, which is a consequence of Assumptions 1 and 2, implies that longer period returns are less contaminated by noise than shorter period returns. In other words, the size of the contaminations does not decrease in probability with the distance between subsequent time stamps. The rounding of recorded prices to a grid, alone, makes this feature of the model compelling.[3]

One final observation is in order before turning to our results. A large and successful literature in macroeconomics has focused on methods to study nonstationary time series (like our observed price process $\ln \widetilde{p}$) expressed as the sum of a nonstationary component (like our unobserved equilibrium price process $\ln p$ with $\phi \equiv 0$) and a residual stationary component (like our noise effect

[1] The study of jumps in the price process is beyond the scope of the present paper and is left for future research. The presence of jumps would require re-thinking about the object of interest since emphasis could be placed on the diffusive component of integrated variance over the period, on the total variance inclusive of squared jumps, or simply on the sum of the squared jumps. Importantly, the main limiting results in this paper, namely the diverging behavior of realized variance at high frequencies (Theorem 1, Point (*ii*)) and the consistency of the noise moment estimates (Theorem 2) would hold even in the presence of infrequent Poisson jumps, for example.

[2] For jumps in volatility, see Eraker et al. (2003) and the references therein. For diurnal effects, see Andersen and Bollerslev (1998) and the references therein. For persistence in volatility, see Andersen et al. (2003) and the references therein. For potential nonstationarities in volatility, see Bandi and Perron (2006) and the references therein.

[3] We are of course assuming that sampling does not occur between high-frequency price updates.

η) (see, e.g., Stock and Watson (1988), for a review). While these methods have also proved very useful in empirical work on market microstructure (see, e.g. Hasbrouck (1993)), our identification procedures are novel. As we will show below, we rely on the different stochastic orders of the unobserved components of the observed returns as described in the previous paragraph. In other words, we rely on the identification potential of high-frequency return data sampled at different frequencies.

3. ASYMPTOTIC THEORY FOR REALIZED VARIANCE

Under the set-up in Eq. (2.4), we can rewrite the realized variance estimator as the sum of three components, namely

$$\widehat{V} = \underbrace{\sum_{j=1}^{M} r_j^2}_{A} + \underbrace{\sum_{j=1}^{M} \varepsilon_j^2}_{B} + \underbrace{2\sum_{j=1}^{M} r_j \varepsilon_j}_{C}. \tag{3.1}$$

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

Theorem 1 below characterizes our asymptotic findings regarding realized variance. The term Q in the theorem denotes the bounded integrated quarticity discussed, e.g., by BN-S (2002), i.e., $Q = \int_0^h \sigma_s^4 ds$. We use this notation throughout the paper.

THEOREM 1:

(i) Assume absence of market microstructure noise, i.e., $\eta_j \equiv 0\, \forall j$. If Assumption 1 is satisfied, then

$$\sqrt{\frac{M}{h}}\left(\widehat{V} - V\right) \Rightarrow \mathbf{MN}\left(0, 2Q\right) \qquad as\ M \to \infty. \tag{3.2}$$

(ii) Assume presence of market microstructure noise. If Assumption 1 and 2 are satisfied, then

$$\widehat{V} \overset{a.s.}{\to} \infty \qquad as\ M \to \infty. \tag{3.3}$$

PROOF: *See Appendix A.*

Remark 1: (Absence of Microstructure Noise) In the absence of market microstructure contaminations, the estimation error between the realized variance estimator and integrated variance over the period converges weakly to a mean-zero mixed Gaussian distribution at speed $\sqrt{M}$. Result (i) replicates a finding obtained by Jacod (1994, Proposition 9.1) and Jacod and Protter (1998). Under the additional restriction of independence between σ and W (absence of leverage effects), the result was also obtained by BN-S (2002). Here, we obtain it in the context of a functional central limit theory. Our derivation (in Appendix A) is of independent econometric interest. It is novel and simpler than the original derivation in Jacod (1994) and Jacod and Protter (1998).

Remark 2: (Presence of Microstructure Noise) When microstructure effects play a role, the realized variance estimator does not consistently estimate the integrated variance over a period. Intuitively, the summing of an increasing number of contaminated return data entails infinite accumulation of noise as the frequency increases asymptotically. Specifically, while the term A in Eq. (3.1) above converges to the integrated variance over the period (from Eq. (3.2) in (i)), the term B diverges to infinity almost surely.[4] The term C is stochastically dominated by B. The limiting result in (ii) is an asymptotic approximation suggesting that for large M, as is the case for high-frequency data, the researcher must be wary of microstructure contaminations as the effect of the noise can be substantial. Hence, any statement about the informational content of the conventional realized variance estimator as a measurement of the integrated variance of the underlying logarithmic price process ought to be a finite sample statement.

Interestingly, however, sample moments of the *observed* return series can be used to learn about population moments of the *unobserved* noise returns at high frequencies. This result is formalized in the following theorem.

Theorem 2: *If Assumptions 1 and 2 are satisfied and* $\mathbf{E}\left(\eta^8\right) < \infty$, *then*

$$\frac{1}{M}\sum_{j=1}^{M} \widetilde{r}_j^{\,q} \xrightarrow{p} \mathbf{E}\left(\varepsilon^q\right) \qquad q = 2, 3, 4 \tag{3.4}$$

as $M \to \infty$.

Proof: *See Appendix A.*

For large M, Theorem 2 implies that (over any fixed period h) one can consistently estimate

[4] This result does not hinge on the dependence properties of the noise. It simply relies on the different stochastic orders of the noise returns ($O_p(1)$) and equilibrium returns ($O_p(\sqrt{\delta})$). In other words, the result will also hold under the more general noise structure in Subsection 5.3 below.

moments of the unobserved noise process at *all* frequencies by using data sampled at the highest frequency.

We now move from asymptotic arguments to a characterization of the finite sample bias/variance trade-off induced by the accumulation of noise.

4. FINITE SAMPLE THEORY FOR REALIZED VARIANCE

This section derives the MSE of the realized variance estimator conditional on the volatility path. Our strategy is to learn about the underlying integrated variance of the logarithmic price process by sampling at a rate which minimizes the conditional expected squared loss of the realized variance estimator. We show that the minimum MSE is achieved for a finite number of return observations M^*. We also show that M^* depends on moments of the microstructure noise distribution as well as on the integrated variance and the integrated quarticity of the underlying logarithmic price process.

For empirical tractability, the results in this section are stated under three additional conditions which we jointly collect in Assumption 3. We emphasize that these conditions are not required to establish the asymptotic results in the previous section.

Assumption 3:

(1) $\phi \equiv 0$.

(2) $\sigma \perp W$.

(3) *The frictions η_j's are independent of the price process.*

Assumption 3(1) implies unpredictability of the equilibrium returns. While the presence of time-varying risk premia would invalidate this assumption, the drift component is known to be negligible at the sampling frequencies considered in the realized variance literature.[5] Coherently, classical market microstructure theory predicts that the unobservable equilibrium price should evolve as a martingale at high frequencies (see, e.g., O'Hara (1995)).

[5] Consider a stock like IBM, for example. Assume a realistic annual constant drift of 0.08. The magnitude of the drift over a minute interval would then be $0.08/(365*24*60) = 1.52 \times 10^{-7}$. Computing the standard deviation of IBM return data over the same horizon by using data discussed in Section 7 yields a value equal to 9.5×10^{-4}. Hence, at the one minute interval the drift component is 1.6×10^{-4}, or nearly $1/10,000$, the magnitude of the return standard deviation. Similarly, the daily value of the drift based on a 6-hour trading period would be 5.47×10^{-5}. The average daily standard deviation is 1.5×10^{-2}. Therefore, the daily drift is 3.6×10^{-3} the magnitude of the daily standard deviation.

Assumption 3(2) implies absence of leverage effects. While assuming absence of leverage effects is known to be empirically reasonable in the case of exchange rate data, the same condition appears restrictive when examining high-frequency stock returns. Nonetheless, recent work uses tractable parametric models to show that the effect of leverage on the *unconditional* MSE of the realized variance estimator in the absence of market microstructure noise is asymptotically negligible (Meddahi (2002) and Andersen et al. (2005)). This work provides a justification for the standard assumption of no-leverage in the literature (see the review paper by Andersen et al. (2002) for further discussions).

Assumption 3(3) implies independence between the equilibrium returns and the noise components at all frequencies. At low frequencies this assumption provides a reasonable approximation. One way of assessing the empirical plausibility of the assumption at high frequencies is by using "volatility signature plots" (see Andersen et al. (1999, 2000)), as advocated by Hansen and Lunde (2006). Hansen and Lunde (2006) show that the presence of a negative correlation between noise and equilibrium returns might determine realized variance estimates that do not diverge as the sampling frequency increases. Such estimates might even underestimate the true integrated variance over the period. Bandi and Russell (2006b) point out that the diverging behavior of realized variance at high frequencies is rather pervasive across markets, sampling methods (trade time sampling versus calendar time sampling), and observed prices (mid-quotes versus transaction prices). In agreement with our limiting results (Theorem 1, Point (*ii*)), Fig. 1 in the Introduction shows the diverging behavior of the *variance* signature plot of the stock analyzed in Section 7 (IBM). In our data we do not find obvious evidence of a significant, negative correlation.

Theorem 3 lays out the conditional (on the volatility path) MSE of realized variance. In what follows, the symbol $\mathbf{E}_\sigma$ will denote expectation conditional on the path of volatility over the period. We continue to use the symbol $\mathbf{E}$ when the conditioning statement is moot. The symbol ε_{-j} will denote the j-th lag of the variable ε.

THEOREM 3: *If Assumptions 1, 2, and 3 are satisfied, then*

$$\mathbf{E}_\sigma \left(\widehat{V} - V\right)^2 = 2\frac{h}{M}\left(Q + o(1)\right) + Mb + M^2 a + c, \tag{4.1}$$

where the parameters $a, b,$ *and* c *are defined as*:

$$a = \left(\mathbf{E}(\varepsilon^2)\right)^2, \tag{4.2}$$

$$b = \mathbf{E}\left(\varepsilon^4\right) + 2\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) - 3\left(\mathbf{E}(\varepsilon^2)\right)^2, \tag{4.3}$$

and

$$c = 4\mathbf{E}(\varepsilon^2)V - 2\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) + 2\left(\mathbf{E}(\varepsilon^2)\right)^2. \tag{4.4}$$

PROOF: *See Appendix A.*

Should the observed return series not be affected by microstructure noise, then the MSE would decrease to zero as the number of observations M diverges to infinity. In fact, it would simply reduce to the conditional variance of the sum of squared equilibrium returns, i.e., $2\frac{h}{M}(Q + o(1))$ (see BN-S (2002) and Appendix A for a derivation). When microstructure noise is present, the 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 a bias term, i.e.,

$$\mathbf{E}_\sigma\left(\widehat{V} - V\right) = M\mathbf{E}\left(\varepsilon^2\right), \tag{4.5}$$

and an additional variance term (jointly captured by Λ_M). Thus, the term Λ_M induces a trade-off which can be optimized by choosing the number of observations M^* so that

$$M^* \equiv \arg\min\left(2\frac{h}{M}(Q + o(1)) + \Lambda_M\right). \tag{4.6}$$

Notice that $\mathbf{E}(\varepsilon^2\varepsilon_{-1}^2) = \frac{1}{2}\mathbf{E}\left(\varepsilon^4\right)$ under Assumption 2. Hence, the minimum of the conditional MSE expansion of the realized variance estimator only depends on the second and fourth moment of the noise component in the return data as well as on the integrated quarticity Q and the integrated variance V. The second and fourth moment of the noise can be estimated consistently by virtue of Theorem 2.[6] BN-S (2002) have provided an estimator of the integrated quarticity which is consistent in the absence of market microstructure noise, i.e., $\widehat{Q} = \frac{M}{3h}\sum_{j=1}^{M}\widetilde{r}_j^4$. Inevitably, $\widehat{Q}$, like $\widehat{V}$, loses its consistency properties when microstructure effects play a role. In light of this observation, to evaluate the MSE empirically, we suggest using relatively low frequencies (i.e., 15- or 20-minute frequencies) to compute preliminary (roughly unbiased) estimates of $\widehat{V}$ and $\widehat{Q}$.[7] The simulations in Section 6 provide evidence for the empirical usefulness of this straightforward estimation procedure.[8]

[6] The estimators' finite sample biases and bias-corrections are discussed in Subsection 5.3.

[7] The design of efficient quarticity estimators is an important topic for future research.

[8] We refer the reader to the simulations in Bandi and Russell (2006a) and the unpublished Appendix to this paper (posted on the journal's web site) for further evidence.

Next, we offer a simple approximation to choose the optimal frequency in the presence of $MA(1)$ noise.

REMARK 3: *For a high optimal number of observations M^*, we can write*

$$M^* \approx \left(\frac{hQ}{\left(\mathbf{E}(\varepsilon^2)\right)^2} \right)^{1/3}. \qquad (4.7)$$

When the quadratic term in Eq. (4.1) dominates the linear term (i.e., for values of M^* sufficiently large), the approximation in Eq. (4.7) provides a good representation of the optimal number of observations M^*. In Section 7 we show that this property is valid for a liquid stock like IBM. Bandi and Russell (2006a) confirm the validity of this approximation for a large number of S&P100 stocks.

Remark 3 is important for two reasons. First, it provides a straightforward rule-of-thumb to choose M^* without having to go through the otherwise simple minimization of Eq. (4.1) above. Second, it clearly illustrates what the main determinants of the optimal frequency are, namely the integrated quarticity of the logarithmic price process and the squared variance of the microstructure noise component in the return process. Naturally, M^* can be regarded as a signal-to-noise ratio: the stronger the signal, the higher the optimal frequency.

5. EXTENSIONS

We consider three extensions of the previous framework: (*1*) bias-corrected realized variance, (*2*) non-linear functions of integrated variance, and (*3*) dependent noise.

5.1. *Bias-corrected realized variance*

Eq. (4.5) shows that the bias of $\widehat{V}$ is equal to $M\mathbf{E}(\varepsilon^2)$. The squared bias term can be a substantial component of the MSE and bias-correcting the realized variance estimator might be very beneficial in practise (see Zhou (1996), Hansen and Lunde (2006), and Oomen (2005) for important, alternative approaches to bias-correcting in the context of integrated variance estimation). Given an estimate of $\mathbf{E}(\varepsilon^2)$ (obtained as in Theorem 2), a bias-corrected realized variance estimator can be derived. The following theorem presents the conditional MSE in this case. The MSE provides a means of selecting an optimal number of observations M^*_{bc} to be used to *optimally* bias-correct (bc) $\widehat{V}$.

THEOREM 4: *Define $\widehat{V}^{bc} = \widehat{V} - M\mathbf{E}(\varepsilon^2)$. If Assumptions 1, 2, and 3 are satisfied, then*

$$\mathbf{E}_\sigma\left(\widehat{V}^{bc}-V\right)^2 = 2\frac{h}{M}\left(Q+o(1)\right)+Mb+c, \tag{5.1}$$

where the parameters b and c are defined in Eq. (4.3) and Eq. (4.4).

Remark 4: *In this case, the minimum of the MSE (variance) is conveniently defined in closed-form as*

$$M_{bc}^* = \left(\frac{2hQ}{2\mathbf{E}\left(\varepsilon^4\right)-3\left(\mathbf{E}(\varepsilon^2)\right)^2}\right)^{1/2}, \tag{5.2}$$

where

$$2\mathbf{E}\left(\varepsilon^4\right)-3\left(\mathbf{E}(\varepsilon^2)\right)^2 = 4\mathbf{E}\left(\eta^4\right) > 0. \tag{5.3}$$

The optimal number of observations M_{bc}^* of the bias-corrected estimator is in general larger than M^*. This is intuitive in that bias reduction permits sampling at higher frequency. Section 7 provides an empirical application.

It should be noted that we assume absence of estimation uncertainty in the identification of the bias term $M\mathbf{E}\left(\varepsilon^2\right)$. This assumption is justified by the fact that the second moment of the noise has the potential to be estimated accurately in the presence of data sampled at high frequency. We expect the empirical relevance (for the purpose of evaluating M_{bc}^*) of incorporating estimation uncertainty regarding $M\mathbf{E}\left(\varepsilon^2\right)$ to be small. Work on this subject is beyond the scope of the present paper and is left for future research.

5.2. *Nonlinear functions of integrated variance*

Interest could be placed in nonlinear functions of integrated variance. Typical examples in finance are volatility ($\sqrt{V}$) and Sharpe ratios. Neglecting scale factors (i.e., expected excess returns), the latter require computation of $\frac{1}{\sqrt{V}}$. As is the case for Remark 3 and Remark 4 above, Remark 5 provides a simple MSE-based rule to choose the optimal frequency in this case.

Remark 5: *Consider a twice continuously differentiable function $f(.)$. For a high optimal number of observations M_f^*, and a second-order Taylor expansion of $f(.)$, we can write*

$$M_f^* \approx \left(\frac{2h(f'(V))^2 Q}{(f''(V))^2\left(\mathbf{E}(\varepsilon^2)\right)^4}\right)^{1/5}, \tag{5.4}$$

provided $f'(.)\neq 0$ and $f''(.)\neq 0$.

Proof: *See Appendix A.*

As earlier, M_f^* depends on a signal-to-noise ratio. Importantly, as expected, "the signal" (i.e., the variance term) in the numerator depends on the first derivative of the function, whereas "the noise" term in the denominator depends on the function's curvature.[9] For the types of nonlinearities routinely encountered in finance, the non-zero condition on the function's derivatives is of course easily satisfied in general.

In Section 7 we employ Eq. (5.4) to compute the optimal frequency of the realized *volatility* estimator for IBM and a typical day in our sample.

5.3. *Noise dependence*

This subsection generalizes the persistence properties of the noise. To this extent, we revise Assumption 2.

Assumption 2b : (The Microstructure Noise)

(1) *The microstructure frictions in the price process η_j's have mean zero and are strictly stationary with joint density $f_M(.)$.*

(2) *The variance of $\varepsilon_j = \eta_j - \eta_{j-1}$ is $K_M + o(1)$ with $K_M > 0$.*

Assumption 2b permits general dependence features for the microstructure noise components in the recorded prices. The correlation structure of the microstructure noise contaminations can capture, as earlier, first-order negative autocorrelation in the recorded high-frequency returns, as determined by bid-ask bounce effects, as well as higher-order dependence as induced by clustering in order flows. In this case, however, the characteristics of the noise returns ε's may, in general, depend on the sampling frequency $\delta = \frac{h}{M}$. The joint density of the η's has a subscript M to make this dependence explicit. Consistently, the symbol $\mathbf{E}_M$ will be used in what follows to denote expectation given M (and the measure $f_M(.)$ in the case of the noise returns). As before, the symbol $\mathbf{E}_{\sigma,M}$ will denote expectation conditional on the volatility path.

Assumption 2b(2) implies that the variance of the noise returns does not converge to zero as the observation frequency increases. This is effectively an identification condition justified by the economics of high-frequency price formation provided sampling does not occur within price updates. This assumption yields, as earlier, the asymptotic diverging behavior of realized variance

[9] Naturally, the accuracy of the approximation depends on the function $f(.)$ as well as on the moments of the unobserved price components. Given a Taylor expansion of the relevant function, one could, in principle, minimize the full-blown MSE. This said, even in the second-order expansion case, the form of the MSE is very involved and of little practical applicability. This MSE can be provided by the authors upon request.

over every fixed time period (as in Theorem 1 above). If the variance of the noise returns converged to zero at the same speed as the variance of the equilibrium returns (as implied by a diffusion model for the price contaminations, for example), then realized variance would converge to the sum of (increments of) equilibrium price and noise price quadratic variation. However, given the discreetness in the observed prices, it is unlikely that this modelling alternative would provide a meaningful approximation.

Theorem 4 presents the conditional MSE of the realized variance estimator under general noise dependence.

THEOREM 5: *If Assumptions 1, 2b, and 3 are satisfied, then*

$$\mathbf{E}_{\sigma,M}\left(\widehat{V}-V\right)^2 = 2\frac{h}{M}\left(Q+o(1)\right)+\Lambda_M, \tag{5.5}$$

where

$$\Lambda_M = M\mathbf{E}_M\left(\varepsilon^4\right)+2\sum_{j=1}^{M}(M-j)\mathbf{E}_M\left(\varepsilon^2\varepsilon_{-j}^2\right)+4\mathbf{E}_M(\varepsilon^2)V. \tag{5.6}$$

Empirical evaluation of the MSE (and choice of M^*) is harder in this case than in the $MA(1)$ case. On the one hand, the MSE is not a closed-form function of M since the noise moments change with the sampling frequency. On the other hand, sample moments of the observed returns (over each period) cannot deliver consistent estimates of the noise moments for each M. However, for each M, we can derive (roughly) unbiased estimates of the noise moments. These can be used, in conjunction with low-frequency estimates of $\widehat{V}$ and $\widehat{Q}$, as earlier, to evaluate the MSE for every frequency before finding its minimum. To this extent, we notice that

$$\begin{aligned}\mathbf{E}_M(\varepsilon^2) &= \mathbf{E}_{\sigma,M}\left(\frac{1}{M}\sum_{j=1}^{M}\widetilde{r}_j^2\right)-\frac{V}{M}, &(5.7)\\ \mathbf{E}_M(\varepsilon^4) &= \mathbf{E}_{\sigma,M}\left(\frac{1}{M}\sum_{j=1}^{M}\widetilde{r}_j^4\right)-\frac{6\mathbf{E}_M(\varepsilon^2)V}{M}+O\left(\frac{1}{M^2}\right), &(5.8)\end{aligned}$$

and

$$\mathbf{E}_M(\varepsilon^2\varepsilon_{-s}^2) = \mathbf{E}_{\sigma,M}\left(\frac{1}{M-s}\sum_{j=s+1}^{M}\widetilde{r}_j^2\widetilde{r}_{j-s}^2\right)-\frac{2\mathbf{E}_M(\varepsilon^2)V}{M-s}+O\left(\frac{1}{M(M-s)}\right) \tag{5.9}$$

for all fixed $s < M$, given the volatility path. These expressions can be used to bias-correct the sample moments of the observed returns for each frequency $\frac{h}{M}$ within each period of interest.

When ignoring the order terms in Eqs. (5.8) and (5.9), the procedure results in a first-order bias-correction. However, if one is worried about accuracy, the full bias can be characterized. Specifically, the order term in Eq. (5.8) is equal to $\approx -\frac{3hQ}{M^2}$ whereas the order term in Eq. (5.9) is equal to

$$-\frac{1}{M-s}\sum_{j=s+1}^{M}\left(\int_{(j-1)\delta}^{j\delta}\sigma_u^2 du\right)\left(\int_{(j-s-1)\delta}^{(j-s)\delta}\sigma_u^2 du\right)$$
$$+\frac{\mathbf{E}_M(\varepsilon^2)}{M-s}\left(\sum_{j=M-s+1}^{M}\left(\int_{(j-1)\delta}^{j\delta}\sigma_u^2 du\right)+\sum_{j=1}^{s}\left(\int_{(j-1)\delta}^{j\delta}\sigma_u^2 du\right)\right). \quad (5.10)$$

Both the first-order bias-correction and the full bias-correction can be applied to straight sample moments of the observed return data. As in Theorem 2, these moments can of course be computed over a fixed time span h. Hence, we do not need to use information across multiple periods to evaluate the single-period MSEs. However, under the assumption that the properties of the noise do not change from period to period, we can average the sample moments of the observed returns across periods and employ conventional limiting arguments as the number of periods n diverges to infinity for a *fixed* M. Given the volatility path, this procedure would result in consistent estimation of the relevant moments with the quantities that depend on volatility being replaced by their averages across periods. The single-period object V, for example, would have to be replaced by its average across periods, i.e., $\sum_{i=1}^{n} V_i/n$. Section 7 applies these methods.

One can of course bias-correct realized variance even in the presence of noise dependence and then optimize the estimator's MSE (i.e., variance). In this case, the MSE of the bias-corrected estimator is the same as that in Theorem 5 without a term equal to $M^2\left(\mathbf{E}_M(\varepsilon^2)\right)^2$. While, as earlier, the MSE can be evaluated for any fixed time period h, meaningful bias-correcting requires computation of $\mathbf{E}_M(\varepsilon^2)$ by using information across periods:

$$\widehat{V}^{bc} = \widehat{V} - M_{bc}^{*}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{M_{bc}^{*}}\sum_{j=1}^{M_{bc}^{*}}\widetilde{r}_{j,i}^{2} - \frac{V_i}{M_{bc}^{*}}\right)\right]. \quad (5.11)$$

Of course, the averaging in the bias-correction can be local (and rely on a moving window of periods, rather than on the full sample n) if one wishes to allow for properties of the noise that change over time.

6. SIMULATIONS

Our theoretical results deliver an optimal sampling frequency which is determined by moments of the market microstructure noise and moments of the equilibrium price process. In practice these moments must be estimated from the data. In this section we provide Monte Carlo evidence about the empirical performance of our optimal sampling theory when the moments are estimated from data.

In our Monte Carlo setting we can evaluate the true MSE (given moments of the data generating process) for any choice of sampling interval. We can therefore compare the relative performance (in an MSE sense) of realized variance estimators constructed using *any* sampling interval of interest. Specifically, we compare the MSEs of three estimators. The first estimator is the true minimum MSE estimator which chooses an interval which minimizes the true MSE (i.e., the estimator which uses M^* as defined in Eq. (4.6)). This estimator is intended to represent the best realized variance estimator one can obtain in an MSE sense. Obviously, this estimator is *infeasible* in practice since it requires knowledge of the true noise moments, integrated quarticity, and integrated variance. The second estimator relies on fixed 15-minute intervals. The MSE for this estimator represents a benchmark reflecting practical applications of realized variance estimates (see, e.g., Andersen et al. (1999, 2000)). Finally, we consider the estimator which uses sample estimates of the noise moments, integrated quarticity, and integrated variance. This estimator represents empirical application of our proposed optimal sampling theory. As such, we refer to it as the *feasible* optimal estimator.

Since the observed price is the equilibrium price plus market microstructure noise, our Monte Carlo experiment requires a model for the noise and for the equilibrium price process. Recent research suggests the importance of a two-factor volatility model for the dynamics of the equilibrium price. We follow Chernov et al. (2003) and write:

$$d\ln(p_t) = \alpha_{10}dt + s\text{-}\exp\left(\beta_0 + \beta_1 v_{1t} + \beta_2 v_{2t}\right) dW_{pt}, \tag{6.1}$$

$$dv_{1t} = \alpha_1 v_{1t}dt + dW_{v_1 t}, \tag{6.2}$$

$$dv_{2t} = \alpha_2 v_{2t}dt + \left(1 + \beta_3 v_{2t}\right) dW_{v_2 t}, \tag{6.3}$$

where $dW_{v_1 t}$ and $dW_{v_2 t}$ are independent Brownian motions and s-exp denotes the exponential function splined with appropriate growth conditions ensuring the existence of a unique, stationary solution (see Appendix A of Chernov et al. (2003)). The leverage correlations are

$corr(dW_p, dW_{v_1}) = \rho_1$ and $corr(dW_p, dW_{v_2}) = \rho_2$. We use the same parameter values as those used in Huang and Tauchen (2006).

We simulate 5,000 days, each of 6.5 hours, using 1-second discretized increments. The initial value of the volatility process is set equal to its unconditional mean of 14% annualized (obtained from a preliminary set of simulations). The variance of the simulated volatilities is extremely high. The smallest and largest daily volatilities in the sample correspond to annualized return volatilities of .08% and over 100%, respectively. This is reflected in a very large value of the return kurtosis of 230. While it is not clear that any single stock would have such a broad volatility range, this process is useful to empirically establish the performance of our realized variance estimates for a broad range of states of the underlying volatility process.

We specify the market microstructure noise as a mean zero, i.i.d. Gaussian process. We set the variance of the market microstructure noise equal to $.0012\sigma^2$, where σ^2 is the unconditional variance of the daily true return process. The ratio between noise variance and integrated variance of the underlying true price is fairly typical (see Bandi and Russell, 2006a). The daily unconditional true return variance is .76. Our noise variance is therefore equal to .000912. The observed return data is generated by summing the equilibrium returns and the noise returns over 1-second intervals.

For each day we estimate the moments of the noise using 10-second sampling intervals. Preliminary estimates of Q and V are obtained using 15-minute return data. For each day, we construct the true MSE function using the true noise moments, the true daily integrated variance, and the true daily integrated quarticity. We then evaluate the MSE at three points, the infeasible minimum MSE sampling interval, the 15-minute sampling interval, and the feasible minimum MSE sampling interval.

We first discuss the performance of our feasible minimum MSE estimator relative to the infeasible minimum MSE estimator. Fig. 2 presents the empirical cumulative distribution function (CDF) for the ratio of the root MSE (RMSE) of the infeasible optimal estimator to the RMSE of the feasible optimal estimator.

Figure 2 about here

This plot summarizes the efficiency loss which results from using estimated moments to construct the optimal sampling intervals rather than true moments. The feasible estimator performs very well. For half of the days in the simulated sample the RMSE of the infeasible estimator is at least 95% that of the feasible estimator. For three quarters of the days the RMSE of the infeasible estimator is at least 83% of the RMSE of the feasible estimator. The largest differences result in a RMSE of the infeasible estimator which is about 35% of the RMSE of the feasible estimator. Thus, the imprecision induced by moment estimation error appears small from an MSE perspective.

We now turn to the comparison between the feasible optimal sampling interval and the fixed 15-minute interval. Fig. 3 presents the empirical CDF for the ratio of the RMSE of the feasible optimal estimator to the RMSE of the 15-minute estimator.

Figure 3 about here

This plot compares two feasible estimators which might be used in practice. The feasible optimal sampling interval outperforms the 15-minute estimator 88% of the time. Additionally, the improvement in RMSE can be substantial - as much as 5 times more accurate. Not surprisingly, the days in which the 15-minute interval outperforms the optimal interval always correspond to optimal intervals near 15 minutes. Even on the rare occasions when the 15-minute interval outperforms the optimal interval, the loss in efficiency is generally extremely small with the RMSE for the feasible optimal interval exceeding the 15-minute RMSE by 10% or more on fewer than 2.5% of the days. Because the relevant comparison is between the feasible optimal estimator and fixed-interval estimators used and/or suggested in the extant literature (such as the 15-minute estimator), these results are important. The results suggest that, while estimation error does not play an important role in practise, the upside to optimal sampling can be large and the downside is very small.

7. A SPECIFIC STOCK: IBM

7.1. *The data*

We use quote-to-quote mid-point prices. The quotes were obtained from the TAQ data set for the month of February 2002. We restrict our attention to NYSE and MIDWEST quote updates.[10] Quotes prior to 9:30am are removed to ensure our sample contains no opening quotes. We have a total of $140,614$ quote arrivals over the month. On average a new quote arrives every 2.92

[10] Ideally, one would want to use all available quotes from the consolidated market to construct the mid-quote return series. However, quotes from the satellite markets tended to either be far noisier than those generated by the NYSE specialist or have a suspicious correlation structure. This observation is particularly true in the case of the NASDAQ quotes. A notable exception was the MIDWEST exchange. We therefore constructed mid-quote return series for IBM by using quotes obtained both from the NYSE and the MIDWEST exchange.

seconds. We construct 10-second continuously-compounded returns. The smallest return in our sample is -0.92% and the largest is 1%. The first autocorrelation is significantly negative and equal to -0.12. The second autocorrelation is 0.045 and the third is 0.02. Thus, the $MA(1)$ approximation appears to capture the main economic effects in our data.[11] We confirm this finding in the next subsection.

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

If the second moment of the noise return does not vary across frequencies, as in the $MA(1)$ model, then Theorem 2 implies that a re-scaled version of the realized variance estimator computed at high frequencies consistently estimates the variance of the unobservable noise process over each day. Here we also average the single period estimates across days. If the second moment of the noise component changes from day to day, as typically the case in practise, then the resulting value can be interpreted as an average daily estimate. If the second moment does not change from day to day, then averaging the single period estimates across days would result in efficiency gains in the estimation of the constant noise second moment. We find that the square root of the re-scaled realized variance (computed over a 6.5-hour trading period) is .038%. Notice that this estimate is essentially the sample standard deviation imposing a mean return of zero. 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, under the $MA(1)$ model, the variance of the noise term η in Eq. (2.2) is one half the variance of the return contamination ε. Thus, the standard deviation of the logarithmic noise price is given by $\sigma_\eta = .038\%/\sqrt{2} = .026\%$. 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 2.6 cents. For added perspective, the average spread for IBM in our sample is 10.6 cents. Hence, the standard deviation is small relative to the spread with a $+/-$ 2 standard deviation interval just about equal to the average spread. Since most trades take place inside the bid/ask spread, the estimated magnitude of the noise variance appears very plausible.

7.3. *Computing the optimal frequency for IBM*

In Fig. 4 we plot the estimated conditional MSE of the realized variance estimator under the $MA(1)$ assumption as well as the (nonparametric) MSE for the case of dependent noise as described in Subsection 5.3.

[11] The corresponding numbers for the first three autocorrelations of the 5 second returns are -0.15, 0.003, and 0.022.

Figure 4 about here

The MSEs are estimated using the methods discussed in Section 4 and in Subsection 5.3. Specifically, since the noise moments do not vary across frequencies under the $MA(1)$ model, we use sample moments constructed using quote-to-quote return data to consistently estimate the relevant population moments of the noise components as in the previous subsection. In the case of the nonparametric MSE, the moments are evaluated on a grid of frequencies and recomputed for each frequency as required by the dependent-noise case. They are also fully bias-corrected to avoid contaminations induced by the underlying equilibrium price process. Thus, unlike the $MA(1)$ case, the nonparametric MSE allows for a general correlation structure of the η's within frequencies. Also, differently from the $MA(1)$ case, we allow for dependence in the squared noise returns through lag 4. Hence, we impose that the terms $\mathbf{E}_M\left(\varepsilon^2\varepsilon^2_{-s}\right)$ are equal to $\left(\mathbf{E}_M\left(\varepsilon^2\right)\right)^2$ only for $s \geq 5$. In the $MA(1)$ model this property is satisfied for $s \geq 2$. In both cases, we obtain preliminary estimates of Q and V by constructing $\widehat{Q}$ and $\widehat{V}$ using 15-minute returns.

Interestingly, we find that the two MSEs largely overlap. Again, the $MA(1)$ approximation seems well suited for our data. The minimum of the $MA(1)$ MSE is 2.7 minutes. The minimum of the nonparametric MSE is 2 minutes. The corresponding MSE values are virtually identical. Both intervals are shorter than the 5-minute interval which is used in some empirical work on the subject (see, e.g., Andersen et al. (2001)). Consequently, they are shorter than recent conjectures on optimal sampling based on the 15/20-minute interval (see, e.g., Andersen et al. (1999, 2000)). The reported findings of course hinge on the features of IBM. Bandi and Russell (2006a) find that stocks with a higher noise-to-signal ratio $\frac{\mathbf{E}(\varepsilon^2)}{V}$ than IBM require lower sampling frequencies leading to optimal intervals which can be in the vicinity of the 5-minute interval. The estimated average (over the month) daily IBM realized variance is about .0003.

In virtue of the empirical relevance of the $MA(1)$ model in our data, it is worth examining the impact of bias-correcting in the context of the $MA(1)$ model. As we have shown, the optimal frequency of the bias-corrected estimator can be conveniently determined in closed-form (c.f., Remark 4). Fig. 5 reports the $MA(1)$ MSE in Fig. 4 and compares it to the corresponding MSE of the bias-corrected estimator from Subsection 5.1.

Figure 5 about here

The value of the MSE of the realized variance estimator at its minimum is 2.01×10^{-9}. The minimum value of the bias-corrected estimator's MSE is 1.82×10^{-9}. We also find that the rule-of-thumb in Remark 3 delivers a very reasonable approximation to the true minimum. The value of the MSE of the estimator at the approximate minimum is 2.22×10^{-9}. For an application of these procedures to a wider set of stocks, namely the S&P100 stocks, we refer the interested reader to Bandi and Russell (2006a).

Finally, we consider optimal sampling for the purpose of computing $\sqrt{V}$, i.e., volatility, as in Subsection 5.2. The expression in Eq. (5.4) suggests that $M_f^* = (M^*)^{3/5}\left(\frac{2(f^{'}(V))^2}{(f^{''}(V))^2(\mathbf{E}(\varepsilon^2))^2}\right)^{1/5}$, where M^* is the optimal number of observations from Remark 3. Using $V = .0003$ and $\mathbf{E}(\varepsilon^2) = 1.5 \times 10^{-7}$, which are the sample averaged (over the month) realized variance and noise return variance respectively, we find that $M^*_{\sqrt{\cdot}} = (\frac{6.5*60}{2.7})^{3/5}\left(\frac{8(.0003)^2}{(1.5\times 10^{-7})^2}\right)^{1/5} \approx 622$ for our IBM data. Specifically, for our data the approximate optimal frequency for volatility estimation is about 36 seconds. Interestingly, it is easy to show that $M^*_{\sqrt{\cdot}} > M^*$ provided $\mathbf{E}(\varepsilon^2) < 8^{\frac{3}{2}}\frac{V^3}{Q}$ and that this condition is generally satisfied in practise. The orders of magnitude of V^2 and Q are empirically similar and $\mathbf{E}(\varepsilon^2)$ is typically three orders of magnitude smaller than V (see Bandi and Russell, 2006a, for evidence based on an extended sample of S&P100 stocks). Hence, the optimal sampling frequency of the realized volatility estimator is typically higher than the optimal sampling frequency of the realized variance estimator.

8. OPTION PRICING and OPTION TRADING

The previous section considered the *statistical* gains associated with optimal sampling in an MSE sense. As in Engle et al. (1990) we construct an hypothetical option market. In this market, we now evaluate the *economic* benefit of optimal sampling. Since variance estimates and variance forecasts play a central role in option pricing, this is a natural setting to evaluate the economic gains from optimal sampling. For economic gains associated with optimal sampling in dynamic portfolio allocation problems, we refer the reader to Bandi and Russell (2006a) and Bandi et al. (2006).

Our hypothetical option market has 4 different agents/traders. Each agent uses a different Method to estimate daily realized variance - fixed 5-minute sampling, 15-minute sampling, 30-minute sampling, and optimal time-varying (from day-to-day) MSE-based sampling as described in Section 4, Remark 3. The agents then employ their resulting daily variance estimates to construct out-of-sample 1-day ahead variance forecasts. Given their forecasts, the agents price 1-day at-the-money options on a $100 stock using the Black-Scholes formula. The risk-free rate is taken to be zero.

When the strike price is equal to the spot price and the risk-free rate is zero, the Black-Scholes call price evaluated at the one-day ahead variance *forecast* $\widetilde{V}$ reduces to $P_i = 2\Phi\left(\frac{1}{2}\widetilde{V}_i^{\frac{1}{2}}\right) - 1$, where $\Phi(.)$ is the normal CDF. By put/call parity, P_i is also the corresponding put price. Although variance is time-varying within day, short-term, at-the-money option prices are expected to be well approximated by Black-Scholes prices evaluated at the expected variance over the life of the option (see, e.g., Poteshman (2002) for discussions).

We are now specific about the various stages of the pricing and trading process.

1. Given a Method and an associated variance forecast, each agent computes his/her fair option price for a 1-day at-the-money option on a \$100 stock.

2. The pair-wise trades take place at the mid-point of the Black-Scholes prices for the optimal sampling agent and each of the fixed interval agents. Agents with variance forecasts leading to prices higher than the mid-point will perceive the options to be underpriced. Hence, they will buy a straddle (one put and one call) from their counterpart. Thus, agents with high variance forecasts effectively speculate on variance in that, on the one hand, they perceive the straddle to be underpriced while, on the other hand, they count on either option to end up considerably in the money due to the expected high variance. The daily profit to a trader who buys the straddle is then given by $|R_i| - 2P_i$. The daily profit to a seller is of course $2P_i - |R_i|$.

3. The average profits for the optimal agent are computed at the end of the trading. These average profits constitute our economic metric.

We consider daily options on SBC, Merrill Lynch (MER) and Exxon Mobile Corporation (XOM). These are the same stocks used in Bandi and Russell (2006a) and Bandi et al. (2006).[12] Our sample period extends from January 2, 1997 through December 31, 2003. Since the realized variances are for a 6.5-hour day and we are pricing daily options, we must make an adjustment to account for the overnight period. Following Bandi and Russell (2006a), we convert the 6.5-hour realized variances $\widehat{V}_i$ for a given stock to daily (or 24-hour) variances by multiplying each realized variance by $\zeta = \sum_{i=1}^{n} R_i^2 / \sum_{i=1}^{n} \widehat{V}_i$, where R_i^2 is the i-th squared daily stock return. Hence, the average of the converted variance series is the same as the variance of the daily returns over the same time period. Subsequently, 1-day ahead forecasts are constructed for each of the 4 agents and for each

[12] They correspond to the first, median, and ninth decile of the distribution of the noise-to-signal ratios $\frac{\mathbf{E}(\varepsilon^2)}{V}$ for the S&P100 stocks examined in Bandi and Russell (2006a)

stock (for a total of 12 series) using an ARFIMA(2,d,2) model. The forecasts are out-of-sample in that each subsequent 1-day ahead ARFIMA forecast uses all available data prior to the forecast period. The first forecast uses the first 200 observations for estimation.[13]

If neither agent of a given pair of traders has a superior variance forecast, profits should be random and zero on average. Similarly, if one agent has a superior variance forecast, then that agent should, on average, earn profits at the other agent's expense. Hence, average profits provide a natural economic benchmark to assess the relative quality of alternative variance forecasts. Table 1 presents the average profits per trade for the trader using optimal sampling. Profits are in cents and reported for each of the three stocks and each pair of Methods (optimal sampling vs. fixed intervals).

Table 1 about here

Interestingly, for the 9 sets of profits, we see that all are positive for the optimal sampling agent except for one instance when the profits are nearly zero (this occurs when trading with the 15-minute agent in the XOM case). This finding indicates that the mid-point of the optimal and fixed-interval option prices miss-prices the straddle relative to the price predicted by the optimal sampling agent. Not only are the miss-pricings pervasive for these stocks, but they are also economically substantial. Consider SBC and the 5-minute agent. The optimal agent earns average profits of 2.76 cents per trade. The typical call/put price is about 92 cents so the straddle averages $1.84. Since profits can be made by the optimal agent, the mid-point miss-prices the straddle by about 1.5%, on average.

A natural summary of the overall profitability of optimal sampling is given by the profits of the portfolio of all 9 straddle positions. Average profits for the portfolio are 20.9 cents. The average total price of the 9 calls/puts is $8.37. Hence, the miss-pricing of the portfolio is about 1.25% of the average straddle price. We test the statistical significance of the profits on the portfolio. Since the portfolio profits are highly heteroskedastic, we account for time-varying variance by using a GARCH(1,1) model. The null of zero average profits is easily rejected with a p-value of 1.7%. Hence, the profits are pervasive across stocks and choice of fixed sampling interval, economically meaningful, and statistically significant.

[13] As in Bandi and Russell (2006a) and Bandi et al. (2006) we use the entire sample to estimate the long-memory parameter d. This parameter is then fixed and only the ARMA terms are re-estimated for each forecast. Consistent with most empirical studies, we find long-memory parameter estimates near .45. Despite this not being a completely out-of-sample forecasting exercise, the results are not sensitive to the precise value of d. In other words, any reasonable value of the long-memory parameter delivers very similar results.

The goal of this section is to provide a way to convert the statistical gains yielded by our optimal sampling methods into a dollar metric. To do so, we follow the standard practice in the industry and academic literature to derive option prices based on alternative variance forecasts. However, while we empirically show that the MSE gains translate into important economic gains in our option pricing framework, the profit-maximizing optimal forecasts are of course not, in general, the same as the forecasts obtained from MSE-optimal variance estimates. On the one hand, there is important scope for selecting an optimal frequency for the purpose of optimizing the forecasts directly. This selection could be based on statistical loss functions such as forecast MSEs. On the other hand, the selection might be tailored to relevant economic criteria such as option pricing errors. Initial work on the former issue is considered in Andersen et al. (2006), Bandi et al. (2007), and Ghysels and Sinko (2006).

9. CONCLUSIONS

Recorded high-frequency asset prices are known to diverge from their equilibrium values due to the presence of market microstructure frictions. In light of this observation, we provide an asymptotic and finite sample treatment of the effects of general microstructure noise on realized variance estimates.

Asymptotically, we find that the presence of market microstructure noise renders the consistent estimation of the quadratic variation of the underlying logarithmic price process through the conventional realized variance estimator unachievable. The summing of increasingly-frequent contaminated return data entails substantial accumulation of noise. Importantly, we show that, under assumptions, a standardized version of the realized variance estimator can be employed to identify the second moment of the unobservable noise process. More generally, we find that sample moments of the contaminated *observed* returns can be put to work to evaluate the moments of the underlying *unobservable* noise process.

Moving from asymptotic arguments to finite sample results, we argue that the microstructure noise creates a dichotomy in the use of realized variance. 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 squared returns entails considerable noise accumulation. The final effect is the determination of a finite sample bias/variance trade-off. We quantify this trade-off and provide a sampling theory for realized variance. Our theory includes bias-corrected realized variance estimates, nonlinear functions of integrated variance, and general noise-dependence of the stationary type. Using a sample of representative stocks, we illustrate the *statistical* and *economic* benefits of our finite sample methods.

In our view, this paper makes three main contributions. It proposes a novel identification procedure for noise and equilibrium price moments relying on the information content of high-frequency data sampled at different frequencies. It highlights the relevance of finite sample methods for nonparametric variance estimates relying on noise-contaminated high-frequency data. It emphasizes the importance of evaluating the performance of these methods in the context of well-posed economic metrics. These ideas can been applied more generally and to less classical estimators than realized variance as shown in recent work conducted by the authors and others. The recent survey papers by Bandi and Russell (2006c), Barndorff-Nielsen and Shephard (2006), and McAleer and Medeiros (2006) contain discussions.

10. APPENDIX A: PROOFS

Proof of Theorem 1: Recall that $\frac{h}{\delta} = M$ and write term A in Eq. (3.1) as

$$A = \underbrace{\sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta}\phi_s ds\right)^2}_{\alpha} + \underbrace{\sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta}\sigma_s dW_s\right)^2}_{\beta} \tag{10.1}$$

$$\underbrace{+2\sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta}\phi_s ds\right)\left(\int_{(j-1)\delta}^{j\delta}\sigma_s dW_s\right)}_{\gamma}. \tag{10.2}$$

We study β first. Define

$$L_j = \int_0^{j\delta}\sigma_s dW_s. \tag{10.3}$$

Then,

$$\beta = \sum_{j=1}^{h/\delta}(L_j - L_{j-1})^2 = \sum_{j=1}^{h/\delta}\left(L_j^2 + L_{j-1}^2 - 2L_jL_{j-1}\right) \tag{10.4}$$

$$= \sum_{j=1}^{h/\delta}\left(L_j^2 - L_{j-1}^2 - 2L_{j-1}\left(L_j - L_{j-1}\right)\right). \tag{10.5}$$

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

$$\beta = \sum_{j=1}^{h/\delta}\left(2\int_{(j-1)\delta}^{j\delta} L_s dL_s + [L]_{j-1,j} - 2L_{j-1}\left(L_j - L_{j-1}\right)\right), \tag{10.6}$$

where

$$[L]_{j-1,j} = [L]_{j\delta} - [L]_{(j-1)\delta} = \int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds. \tag{10.7}$$

Hence,

$$\beta - V = 2\sum_{j=1}^{h/\delta}\overline{L}_{j-1,j}, \tag{10.8}$$

where

$$\overline{L}_{j-1,t} = \int_{(j-1)\delta}^{t\delta}\left(L_s - L_{j-1}\right)dL_s \qquad j-1 \le t \le j \tag{10.9}$$

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

$$\Phi_t^M := 2\sqrt{\frac{M}{h}}\left(\sum_{j=1}^{h/\delta-1}\overline{L}_{j-1,j} + \int_{(\frac{h}{\delta}-1)\delta}^{t\delta}\left(L_s - L_{\frac{h}{\delta}-1}\right)dL_s\right), \tag{10.10}$$

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

$$
\begin{aligned}
& 4\frac{M}{h}\sum_{j=1}^{h/\delta-1}\int_{(j-1)\delta}^{j\delta}(L_s-L_{j-1})^2\,d\,[L]_s + 4\frac{M}{h}\int_{\left(\frac{h}{\delta}-1\right)\delta}^{t\delta}\left(L_s-L_{\frac{h}{\delta}-1}\right)^2 d\,[L]_s \\
= \; & 4\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\int_{(j-1)\delta}^{j\delta}[L-L_{j-1}]_s\,d\,[L]_s+\int_{\left(\frac{h}{\delta}-1\right)\delta}^{t\delta}\left[L-L_{\frac{h}{\delta}-1}\right]_s d\,[L]_s\right)+O_p\left(\sqrt{\frac{h}{M}}\right) && (10.11)\\
= \; & 4\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\int_{(j-1)\delta}^{j\delta}[L-L_{j-1}]_s\,d\,[L-L_{j-1}]_s+\int_{\left(\frac{h}{\delta}-1\right)\delta}^{t\delta}\left[L-L_{\frac{h}{\delta}-1}\right]_s d\left[L-L_{\frac{h}{\delta}-1}\right]_s\right) && (10.12)\\
= \; & 4\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\frac{[L-L_{j-1}]^2}{2}\Big|_{j-1}^{j}+\frac{\left[L-L_{\frac{h}{\delta}-1}\right]^2}{2}\Big|_{\frac{h}{\delta}-1}^{t}\right) && (10.13)\\
= \; & 2\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}[L_j-L_{j-1}]^2+\left[L_t-L_{\frac{h}{\delta}-1}\right]^2\right) && (10.14)\\
= \; & 2\frac{M}{h}\left(\sum_{j=1}^{h/\delta-1}\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds\right)^2+\left(\int_{\left(\frac{h}{\delta}-1\right)\delta}^{t\delta}\sigma_s^2 ds\right)^2\right). && (10.15)
\end{aligned}
$$

The asymptotic order term in Eq. (10.11) derives from the following argument. Consider

$$4\frac{M}{h}\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}(L_s-L_{j-1})^2 d[L]_s. \qquad (10.16)$$

By Ito's lemma,

$$
\begin{aligned}
(L_s-L_{j-1})^2 &= \int_{(j-1)\delta}^{s}2L_u dL_u+\int_{(j-1)\delta}^{s}d[L]_u-\int_{(j-1)\delta}^{s}2L_{j-1}dL_u && (10.17)\\
&= \int_{(j-1)\delta}^{s}2\left(L_u-L_{j-1}\right)dL_u+\int_{(j-1)\delta}^{s}\sigma_u^2 du && (10.18)\\
&= \int_{(j-1)\delta}^{s}2\left(L_u-L_{j-1}\right)dL_u+[L-L_{j-1}]_s && (10.19)
\end{aligned}
$$

since $L_s-L_{j-1}=\int_{(j-1)\delta}^{s}\sigma_u dW_u$ and $[L-L_{j-1}]_s=\int_{(j-1)\delta}^{s}\sigma_u^2 du$. Hence, we need to show that

$$\Psi^M = 4\frac{M}{h}\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}\left[\int_{(j-1)\delta}^{s}2\left(L_u-L_{j-1}\right)dL_u\right]d[L]_s \xrightarrow{p} 0. \qquad (10.20)$$

Integrate Ψ^M by parts to obtain

$$4\frac{M}{h}\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}\left(-2\left(L_s-L_{j-1}\right)[L]_s\,dL_s+2\left(L_s-L_{j-1}\right)[L]_j\,dL_s\right)$$

$$= 4\frac{M}{h}\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}2\left(L_s-L_{j-1}\right)\left([L]_j-[L]_s\right)dL_s \tag{10.21}$$

$$= 4\frac{M}{h}\sum_{j=1}^{M-1}\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2ds\right)\int_{(j-1)\delta}^{j\delta}2\left(L_s-L_{j-1}\right)dL_s$$

$$-4\frac{M}{h}\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}2\left(L_s-L_{j-1}\right)\left([L]_s-[L]_{j-1}\right)dL_s \tag{10.22}$$

$$= \Psi_1^M+\Psi_2^M. \tag{10.23}$$

As earlier in the case of the expression in Eq. (10.10), the quantity Ψ_2^M can be embedded in a local martingale. Its quadratic variation is given by

$$16\left(\frac{M}{h}\right)^2\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}4\left(L_s-L_{j-1}\right)^2\left([L]_s-[L]_{j-1}\right)^2\sigma_s^2ds$$

$$\leq 64\left(\frac{M}{h}\right)^2\max_{1\leq j\leq M-1}\sup_{j-1\leq s\leq j}\left(L_s-L_{j-1}\right)^2\max_{1\leq j\leq M-1}\sup_{j-1\leq s\leq j}\left([L]_s-[L]_{j-1}\right)^2\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}\sigma_s^2ds \tag{10.24}$$

$$= O_p\left(\frac{h}{M}\right), \tag{10.25}$$

implying that $\Psi_2^M=O_p\left(\sqrt{\frac{h}{M}}\right)$. As for Ψ_2^M, using the same argument leading to Eq. (10.25), it is immediate to show that $\sqrt{\frac{M}{h}}\sum_{j=1}^{M-1}\int_{(j-1)\delta}^{j\delta}2\left(L-L_{j-1}\right)dL=O_p(1)$. Since $\sqrt{\frac{M}{h}}\,[L]_{j-1,j}=O_p\left(\sqrt{\frac{h}{M}}\right)\ \forall j$, then $\Psi_1^M=O_p\left(\sqrt{\frac{h}{M}}\right)$. We now define the stopping times $\tau=\inf\left\{s:\left[\Phi^M\right]_s>t\right\}$. Hence, $B_t=\Phi_\tau^M$ is a $\Im_\tau$-Brownian motion and $\Phi_t^M=B_{[\Phi^M]_t}$ is the DDS (Dambis, Dubins-Schwartz) Brownian motion of Φ_t (see Revuz and Yor (Theorem 1.6, page 173, 1994)). Finally,

$$\sqrt{\frac{M}{h}}\left[\beta-V\right] = \Phi_{h/\delta}^M \underset{M\to\infty}{\Rightarrow} B\left(2\frac{M}{h}\sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2ds\right)^2\right) \tag{10.26}$$

$$\overset{d}{=} B\left(2\left(\int_0^h\sigma_s^4ds\right)\right), \tag{10.27}$$

where $Q=\int_0^h\sigma_s^4ds$ is the integrated quarticity over h as defined in BN-S (2002), by the asymptotic Knight's theorem (see Revuz and Yor (Theorem 2.3, page 496, 1994)). Let us now consider the terms α and γ. Since $\frac{M}{h}\sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta}\phi_sds\right)^2=\int_0^h\phi_s^2ds+o_{a.s.}(1)$ by Lemma 1 in BN-S (2002), for example, then $\sqrt{\frac{M}{h}}\alpha=O_p\left(\sqrt{\frac{h}{M}}\right)$. Now observe that $\sum_{j=1}^{h/\delta}L_j=O_p(1)$ which implies $\frac{M}{h}\gamma=O_p(1)$. Then, $\sqrt{\frac{M}{h}}\gamma=O_p\left(\sqrt{\frac{h}{M}}\right)$. Finally,

$$\sqrt{\frac{M}{h}}\left[A-V\right] = \sqrt{\frac{M}{h}}\left[\beta-V\right]+O_p\left(\sqrt{\frac{h}{M}}\right) \tag{10.28}$$

$$\underset{M\to\infty}{\Rightarrow} B\left(2\left(\int_0^h \sigma_s^4 ds\right)\right). \tag{10.29}$$

In addition, the covariation process between

$$\sum_{j=1}^{h/\delta-1}\left(\sigma_{j\delta}-\sigma_{(j-1)\delta}\right)+\int_{\left(\frac{h}{\delta}-1\right)\delta}^{t\delta}\left(\sigma_s-\sigma_{\frac{h}{\delta}-1}\right)+\sigma_0, \tag{10.30}$$

with $\frac{h}{\delta}-1<t\leq\frac{h}{\delta}$ and Φ_t^M converges to zero in probability as $M\to\infty$ (see, e.g., Bandi and Phillips (2007), Theorem 5). Hence,

$$B\left(2\left(\int_0^h \sigma_s^4 ds\right)\right)\overset{d}{=}\mathbf{MN}\left(0,2\left(\int_0^h \sigma_s^4 ds\right)\right). \tag{10.31}$$

This proves the stated result in Theorem 1, Point (i). As for the result in Point (ii), notice that $B=\sum_{j=1}^{M}\varepsilon_j^2=O_{a.s.}(M)$ under Assumption 2 and $C=O_p(1)$ since $\sum_{j=1}^{h/\delta}L_j=O_p(1)$ and $\varepsilon_j=O_p(1)\ \forall j$. Thus,

$$A+B+C=\left(V+O_p\left(\sqrt{\frac{h}{M}}\right)\right)+O_{a.s.}(M)+O_p(1) \tag{10.32}$$

which implies divergence to infinity with probability one as M increases without bound. ■

Proof of Theorem 2: We show the stated result for the case $q=4$. The proofs in the remaining cases are simplified versions of the following proof. Recall that $\frac{h}{\delta}=M$ and write

$$\sum_{j=1}^{h/\delta}\tilde{r}_j^4=\sum_{j=1}^{h/\delta}\left(r_j+\varepsilon_j\right)^4=\underbrace{\sum_{j=1}^{h/\delta}r_j^4}_{A}+4\underbrace{\sum_{j=1}^{h/\delta}r_j^3\varepsilon_j}_{B}+6\underbrace{\sum_{j=1}^{h/\delta}r_j^2\varepsilon_j^2}_{C}+4\underbrace{\sum_{j=1}^{h/\delta}r_j\varepsilon_j^3}_{D}+\underbrace{\sum_{j=1}^{h/\delta}\varepsilon_j^4}_{E}. \tag{10.33}$$

Note that, for any $\delta>0$, by the Chebyshev's inequality,

$$\begin{aligned}
&\mathbf{P}\left(\left|\frac{E}{M}-\mathbf{E}_M\left(\varepsilon^4\right)\right|>\delta\right)\\
\leq\ & \frac{\frac{1}{M^2}\mathbf{E}_M\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_j^4-\mathbf{E}_M\left(\varepsilon^4\right)\right)\right)^2}{\delta^2} && (10.34)\\
=\ & \frac{1}{\delta^2 M}\left(\Gamma_0^4+2\left(\frac{M-1}{M}\right)\Gamma_1^4+2\left(\frac{M-2}{M}\right)\Gamma_2^4+...+2\left(\frac{1}{M}\right)\Gamma_{M-1}^4\right) && (10.35)\\
\leq\ & 2\frac{1}{\delta^2 M}\left(\left|\Gamma_0^4\right|+\left|\Gamma_1^4\right|+...+\left|\Gamma_{M-1}^4\right|\right)+.... && (10.36)
\end{aligned}$$

Under Assumption 2, the bound vanishes if $\mathbf{E}\left(\eta^8\right)<\infty$. Hence,

$$\frac{E}{M}-\mathbf{E}_M\left(\varepsilon^4\right)\underset{M\to\infty}{\overset{p}{\to}}0 \tag{10.37}$$

or

$$\frac{E}{M}\underset{M\to\infty}{\overset{p}{\to}}\mathbf{E}\left(\varepsilon^4\right), \tag{10.38}$$

since, given Assumption 2, the expectation does not depend on M. Next, write

$$\frac{A}{M} \leq \frac{\left(\left(\sum_{j=1}^{h/\delta} r_j^2\right)\left(\sum_{j=1}^{h/\delta} r_j^2\right)\right)}{M} \leq \frac{1}{M}\left(V^2 + O_p\left(\sqrt{\frac{h}{M}}\right)\right) \underset{M\to\infty}{\overset{p}{\to}} 0, \tag{10.39}$$

by Theorem 1, Point (i), and the delta method. Now notice that

$$\frac{C}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} r_j^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} \varepsilon_j^4\right)^{1/2}}{M} \leq \sqrt{\frac{1}{M}} O_p(1)\left(\left(\mathbf{E}_M\left(\varepsilon^4\right)\right)^{1/2} + o_p(1)\right) \underset{M\to\infty}{\overset{p}{\to}} 0, \tag{10.40}$$

where the first bound follows from the Cauchy-Schwarz (CS, hereafter) inequality and the second bound derives from Eq. (10.39) and Eq. (10.37) by virtue of the delta method. Now consider term B/M and notice that

$$\frac{B}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} r_j^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} r_j^2\varepsilon_j^2\right)^{1/2}}{M} \leq \sqrt{\frac{1}{M}}\left(\frac{1}{M}\right)^{1/4} O_p(1) \underset{M\to\infty}{\overset{p}{\to}} 0, \tag{10.41}$$

where the first bound follows again from the CS inequality and the second bound derives from Eqs. (10.39) and (10.40). Finally, we turn to term D/M and write

$$\frac{D}{M} \leq \frac{\left(\sum_{j=1}^{h/\delta} \varepsilon_j^4\right)^{1/2}\left(\sum_{j=1}^{h/\delta} r_j^2\varepsilon_j^2\right)^{1/2}}{M} \leq \left(\left(\mathbf{E}_M\left(\varepsilon^4\right)\right)^{1/2} + o_p(1)\right)\left(\frac{1}{M}\right)^{1/4} O_p(1) \underset{M\to\infty}{\overset{p}{\to}} 0 \tag{10.42}$$

where, again, the first bound derives from the CS inequality and the second bound derives from the delta method (given Eq. (10.37)) and Eq. (10.40). This proves the stated results.■

Proof of Theorem 3: We show the result in the general dependent-noise case. The result readily specializes to the i.i.d. noise case in Theorem 3. Recall that $\frac{h}{\delta} = M$. We expand the conditional MSE of $\widehat{V}$ and obtain

$$\begin{aligned}
& \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}(r_j+\varepsilon_j)^2 - \int_0^h \sigma_s^2 ds\right)^2 \\
= \; & \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\left(r_j^2+\varepsilon_j^2+2r_j\varepsilon_j\right) - \int_0^h \sigma_s^2 ds\right)^2 \qquad (10.43) \\
= \; & \underbrace{\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta} r_j^2 - \int_0^h \sigma_s^2 ds\right)^2}_{A} + \underbrace{\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_j^2 + 2r_j\varepsilon_j\right)\right)^2}_{B} \\
& \underbrace{+2\mathbf{E}_{\sigma,M}\left(\left(\sum_{j=1}^{h/\delta} r_j^2 - \int_0^h \sigma_s^2 ds\right)\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_j^2+2r_j\varepsilon_j\right)\right)\right)}_{C}. \qquad (10.44)
\end{aligned}$$

We start with B.

$$B = \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}\left(\varepsilon_j^2 + 2r_j\varepsilon_j\right)\left(\varepsilon_g^2 + 2r_g\varepsilon_g\right)\right) \tag{10.45}$$

$$= \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}\varepsilon_j^2\varepsilon_g^2 + 2\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}\varepsilon_j^2 r_g\varepsilon_g + 2\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}r_j\varepsilon_j\varepsilon_g^2 + 4\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}r_j\varepsilon_j r_g\varepsilon_g\right) \tag{10.46}$$

$$= \mathbf{E}_{M}\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}\varepsilon_j^2\varepsilon_g^2\right) + 4\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}r_j\varepsilon_j r_g\varepsilon_g\right) \tag{10.47}$$

$$= \sum_{j=1}^{h/\delta}\mathbf{E}_M\left(\varepsilon_j^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_M\left(\varepsilon_j^2\varepsilon_g^2\right) + 4\left(\sum_{j=1}^{h/\delta}\mathbf{E}_{\sigma,M}\left(r_j^2\varepsilon_j^2\right)\right) + 8\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_{\sigma,M}(r_j\varepsilon_j r_g\varepsilon_g) \tag{10.48}$$

$$= \frac{h}{\delta}\mathbf{E}_M(\varepsilon^4) + 2\sum_{g=1}^{h/\delta}\left(\frac{h}{\delta} - j\right)\mathbf{E}_M\left(\varepsilon^2\varepsilon_{-j}^2\right) + 4\mathbf{E}_M(\varepsilon^2)V. \tag{10.49}$$

We recall that

$$\sum_{j=1}^{h/\delta}\mathbf{E}_{\sigma,M}(r_j^2) = \sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds\right) = V. \tag{10.50}$$

This result is used in Eq. (10.49) above as well as in Eq. (10.53) below. Now consider C.

$$C = 2\left(\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}r_j^2 - \int_0^h\sigma_s^2 ds\right)\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_j^2 + 2r_j\varepsilon_j\right)\right)\right) \tag{10.51}$$

$$= 2\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}r_j^2\right)\mathbf{E}_M\left(\sum_{g=1}^{h/\delta}\varepsilon_g^2\right) + 4\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta}r_j^2 r_g\varepsilon_g\right) - 2V\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\left(\varepsilon_j^2 + 2r_j\varepsilon_j\right)\right) \tag{10.52}$$

$$= 2\frac{h}{\delta}V\mathbf{E}_M(\varepsilon^2) - 2\frac{h}{\delta}V\mathbf{E}_M(\varepsilon^2) = 0. \tag{10.53}$$

We now turn to A. Write

$$A = \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta} r_j^2\right)^2 - 2\mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta} r_j^2\right)\left(\int_0^h \sigma_s^2 ds\right) + \left(\int_0^h \sigma_s^2 ds\right)^2 \quad (10.54)$$

$$= \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta}\sum_{g=1}^{h/\delta} r_j^2 r_g^2\right) - \left(\int_0^h \sigma_s^2 ds\right)^2 \quad (10.55)$$

$$= \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta} r_j^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_{\sigma,M}\left(r_j^2 r_g^2\right) - \left(\int_0^h \sigma_s^2 ds\right)^2 \quad (10.56)$$

$$= \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta} r_j^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\mathbf{E}_{\sigma,M}\left(r_j^2\right)\mathbf{E}_{\sigma,M}\left(r_g^2\right) - \left(\int_0^h \sigma_s^2 ds\right)^2 \quad (10.57)$$

$$= \mathbf{E}_{\sigma,M}\left(\sum_{j=1}^{h/\delta} r_j^4\right) + 2\sum_{g=1}^{h/\delta}\sum_{j<g}\left(\int_{(j-1)\delta}^{j\delta} \sigma_s^2 ds\right)\left(\int_{(g-1)\delta}^{g\delta} \sigma_s^2 ds\right) - \left(\int_0^h \sigma_s^2 ds\right)^2 \quad (10.58)$$

$$= \sum_{j=1}^{h/\delta}\mathbf{E}_{\sigma,M}\left(r_j^4\right) - \sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta} \sigma_s^2 ds\right)^2 \quad (10.59)$$

$$= \sum_{j=1}^{h/\delta}\mathbf{V}_{\sigma,M}\left(r_j^2\right) + \sum_{j=1}^{h/\delta}\left(\mathbf{E}_{\sigma,M}\left(r_j^2\right)\right)^2 - \sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta} \sigma_s^2 ds\right)^2 \quad (10.60)$$

$$= \sum_{j=1}^{h/\delta}\mathbf{V}_{\sigma,M}\left(r_j^2\right) \quad (10.61)$$

It is noted that, conditionally on the volatility path $\{\sigma_s\}_{s\in(0,h)}$, the quantity r_j is Gaussian. Hence, $\frac{r_j^2}{\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds\right)}$ is conditionally Chi-squared with one degree of freedom. Then,

$$\sum_{j=1}^{h/\delta}\mathbf{V}_{\sigma,M}\left(r_j^2\right) = \sum_{j=1}^{h/\delta}\mathbf{V}_{\sigma,M}\left(\frac{r_j^2}{\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds\right)}\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds\right)\right) \quad (10.62)$$

$$= \sum_{j=1}^{h/\delta} 2\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds\right)^2 \quad (10.63)$$

$$= 2\frac{h}{M}\left(\frac{M}{h}\sum_{j=1}^{h/\delta}\left(\int_{(j-1)\delta}^{j\delta}\sigma_s^2 ds\right)^2\right) \quad (10.64)$$

$$= 2\frac{h}{M}\left(Q + o(1)\right), \quad (10.65)$$

where convergence to the integrated quarticity follows from Lemma 1 in BN-S (2002). This proves the result as stated earlier.■

Proof of Remark 5:

Neglecting the higher order terms, write

$$\mathbf{E}_\sigma(f(\widehat{V}) - f(V))^2 = \left(f^{'}(V)\right)^2\mathbf{E}_\sigma(\widehat{V} - V)^2 + \frac{1}{4}\left(f^{''}(V)\right)^2\mathbf{E}_\sigma(\widehat{V} - V)^4 + f^{'}(V)f^{''}(V)\mathbf{E}_\sigma(\widehat{V} - V)^3. \quad (10.66)$$

We now consider the form of the terms only involving the equilibrium returns. We start with:

$$
\begin{aligned}
& \mathbf{E}_\sigma \left(\sum_{j=1}^{h/\delta} r_j^2 - V \right)^4 \\
= \; & \mu_4 \left(\sum_{j=1}^{h/\delta} r_j^2 \right) + 3\mu_2^2 \left(\sum_{j=1}^{h/\delta} r_j^2 \right) \quad \textit{[fourth cumulant + 3 × squared second cumulant]} & (10.67) \\
= \; & \sum_{j=1}^{h/\delta} \mu_4(r_j^2) + 3\left(2\frac{h}{M}Q + o(1) \right)^2 \quad \textit{[by the additivity of the cumulants for independent random variables]} & (10.68) \\
= \; & \sum_{j=1}^{h/\delta} \mu_4 \left(\frac{r_j^2}{\int_{(j-1)\delta}^{j\delta} \sigma_s^2 ds} \int_{(j-1)\delta}^{j\delta} \sigma_s^2 ds \right) + 3\left(2\frac{h}{M}Q + o(1) \right)^2 & (10.69) \\
= \; & \sum_{j=1}^{h/\delta} \left(\int_{(j-1)\delta}^{j\delta} \sigma_s^2 ds \right)^4 \mu_4 \left(Chi_1 \right) \; + 3\left(2\frac{h}{M}Q + o(1) \right)^2 \quad \textit{[by the homogeneity of order k of the k-th cumulant]} & (10.70) \\
= \; & 48 \sum_{j=1}^{h/\delta} \left(\int_{(j-1)\delta}^{j\delta} \sigma_s^2 ds \right)^4 \; + 3\left(2\frac{h}{M}Q + o(1) \right)^2 \quad \textit{[the fourth cumulant of the } Chi_1 \textit{ is 48]} & (10.71) \\
= \; & 48 \left(\frac{h}{M} \right)^3 \int_0^h \sigma_s^8 ds + 12 \left(\frac{h}{M} \right)^2 Q^2 \; + o(1) \quad \textit{[by Lemma 1 in BN-S (2002)].} & (10.72)
\end{aligned}
$$

Similarly, it is straightforward to show that

$$
\mathbf{E}_\sigma \left(\sum_{j=1}^{M} r_j^2 - V \right)^3 = 8 \left(\frac{h}{M} \right)^2 \int_0^h \sigma_s^6 ds + o(1). \tag{10.73}
$$

Hence,

$$
\begin{aligned}
\mathbf{E}_\sigma (f(\widehat{V}) - f(V))^2 \;=\; & \left(f^{'}(V) \right)^2 \left[2\frac{h}{M}Q + M^2 \left(\mathbf{E}(\varepsilon^2) \right)^2 + ... \right] + \\
& \frac{1}{4} \left(f^{''}(V) \right)^2 [48 \left(\frac{h}{M} \right)^3 \int_0^h \sigma_s^8 ds + 12 \left(\frac{h}{M} \right)^2 Q^2 + M^4 \left(\mathbf{E}(\varepsilon^2) \right)^4 + ...] \\
& + f^{'}(V) f^{''}(V) [8 \left(\frac{h}{M} \right)^2 \int_0^h \sigma_s^6 ds + M^3 \left(\mathbf{E}(\varepsilon^2) \right)^3 + ...]. & (10.74)
\end{aligned}
$$

The result derives from the dominating (for a large M) bias and variance components. ■

11. APPENDIX B: NOTATION

$\xrightarrow{p}$	convergence in probability
$\xrightarrow{a.s.}$	almost sure convergence
$\Rightarrow$	weak convergence
$\equiv$	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
$\overset{d}{=}$	distributional equivalence
$\approx$	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., P. Mykland, and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18, 351-416.

[2] Andersen, T.G. and T. Bollerslev (1998). Deutsche market volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 53, 219-265.

[3] 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.*

[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 (1999). (Understanding, optimizing, using and forecasting) Realized volatility and correlation. *Working paper.*

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

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

[8] Andersen, T.G., T. Bollerslev, and N. Meddahi (2005). Correcting the errors: a note on volatility forecast evaluation based on high-frequency data and realized volatilities. *Econometrica* 73, 279-296.

[9] Andersen, T.G., T. Bollerslev, and N. Meddahi (2006). Realized volatility forecasting and market microstructure noise. *Working paper.*

[10] Awartani, B., V. Corradi, and W. Distaso (2004). Testing and modelling market microstructure effects with an application to the Dow Jones Industrial Average. *Working paper.*

[11] Bai, X., J.R. Russell, and G. Tiao (2004). Effects on non-normality and dependence on the precision of variance estimates using high-frequency data. *Working paper.*

[12] Bandi, F.M., and B. Perron (2006). Long memory and the relation between realized and implied volatility. *Journal of Financial Econometrics* 4, 636-670.

[13] Bandi, F.M. and P.C.B. Phillips (2007). A simple approach to the parametric estimation of potentially-nonstationary diffusion processes. *Journal of Econometrics* 137, 354-395.

[14] Bandi, F.M., and J.R. Russell (2006a). Separating microstructure noise from volatility. *Journal of Financial Economics* 79, 655-692.

[15] Bandi, F.M., and J.R. Russell (2006b). Comment on Hansen and Lunde. *Journal of Business and Economic Statistics* 24, 167-173.

[16] Bandi, F.M., and J.R. Russell (2006c). Volatility. In J.R. Birge and V. Linetski (Eds.) *Handbook of Financial Engineering.* Elsevier North-Holland. *Forthcoming.*

[17] Bandi, F.M., and J.R. Russell (2006d). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Working paper.*

[18] Bandi, F.M., J.R. Russell, and C. Yang (2007). Realized volatility forecasting in the presence of time-varying noise. *Working paper.*

[19] Bandi, F.M., J.R. Russell, and Y. Zhu (2006). Using high-frequency data in dynamic portfolio choice. *Econometric Reviews, forthcoming.*

[20] 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.

[21] 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.

[22] Barndorff-Nielsen, O.E., and N. Shephard (2004). Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72, 885-925.

[23] Barndorff-Nielsen, O.E., and N. Shephard (2006). Variation, jumps, market frictions and high-frequency data in financial econometrics. In R. Blundell, P. Torsten and W. K. Newey

(Eds.) *Advances in Economics and Econometrics. Theory and Applications*. Ninth World Congress. Econometric Society Monographs, Cambridge University Press.

[24] Barndorff-Nielsen, O.E., P. Hansen, A. Lunde and N. Shephard (2006). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Working paper.*

[25] Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics* 116, 225-257.

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

[27] Delattre, S., and J. Jacod (1997). A central limit theorem for normalized functions of increments of a diffusion process, in the presence of round-off errors. *Bernoulli* 3, 1-28.

[28] Engle, R.F., C.H. Hong, and A. Kane (1990). Valuation of variance forecasts with simulated option markets. NBER Working paper No: 3350.

[29] Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps and volatility in returns. *Journal of Finance* 58, 1269-1300.

[30] Ghysels, E. and A. Sinko (2006b). Volatility forecasting and microstructure noise. *Working paper.*

[31] Gloter, A. and J. Jacod (2001a). Diffusions with measurement errors: I - Local asymptotic normality. *ESAIM: Probability and Statistics* 5, 225-242.

[32] Gloter, A. and J. Jacod (2001b). Diffusions with measurement errors: II - Optimal estimators. *ESAIM: Probability and Statistics* 5, 243-260.

[33] Hansen, P.R., and A. Lunde (2006). Realized variance and market microstructure noise (with discussions). *Journal of Business and Economic Statistics* 24, 127-161.

[34] Hasbrouck, J. (1993). Assessing the quality of a security market: a new approach to transaction cost measurement. *Review of Financial Studies* 6, 191-212.

[35] Huang, X., and G. Tauchen (2005). The relative contribution of jumps to total price variation. *Journal of Financial Econometrics* 3, 456-499.

[36] Jacod, J. (1994). Limit of random measures associated with the increments of a Brownian semimartingale. *Working paper.*

[37] Jacod, J. and P. Protter (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability* 26, 267-307.

[38] Mancino, M. E., and S. Sanfelici (2006). Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Working paper.*

[39] McAleer, M., and M. Medeiros (2006). Realized volatility: A review. *Econometric Reviews, forthcoming.*

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

[41] O'Hara, M. (1995). *Market Microstructure Theory.* Blackwell Publishers Ltd, Oxford.

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

[43] Oomen, R.C.A. (2005). Properties of bias-corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics* 3*,* 555–577.

[44] Oomen, R.C.A. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business and Economic Statistics* 24*,* 219-237.

[45] Picard, J. (1993). Estimation of the quadratic variation of nearly-observed semimartingales with application to filtering. *SIAM Journal of Control and Optimization* 31, 494-517.

[46] Poteshman, A. (2000). Forecasting future volatility from option prices. *Working paper.*

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

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

[49] Stock, J. and M. Watson (1988). Variable trends in economic time series. *Journal of Economic Perspectives* 2, 147-174.

[50] Stoll, H.R. (2000). Friction. *Journal of Finance* 55, 1479-1514.

[51] Zhang, L., P. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394-1411.

[52] Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics* 14, 45-52.

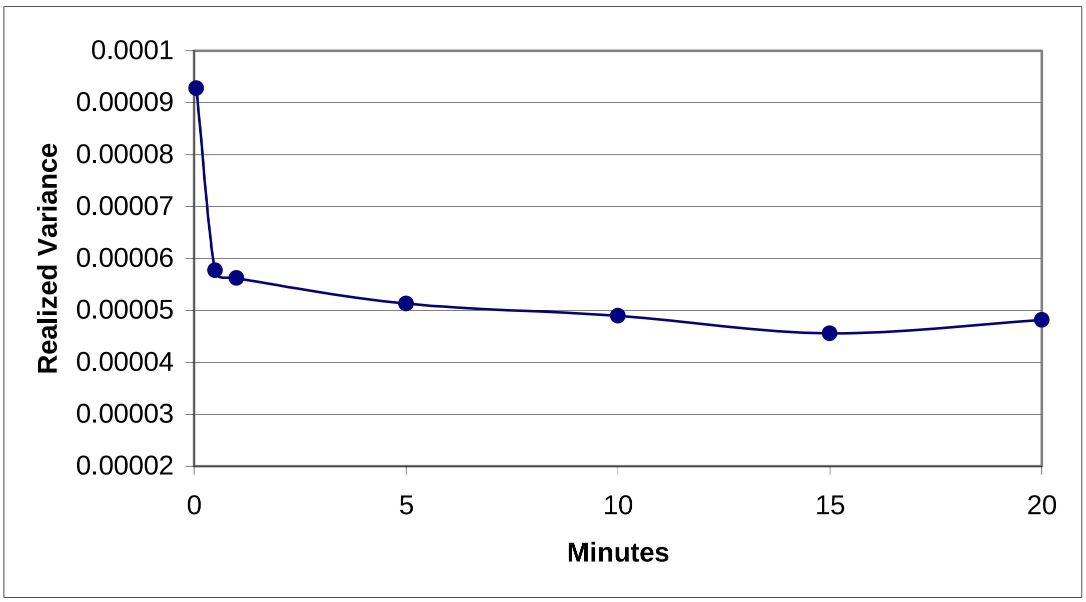

Figure 1. For IBM, we plot the average signature plot across all the days in our sample. The shortest sampling interval corresponds to quote-to-quote returns. All others intervals are in calendar time.

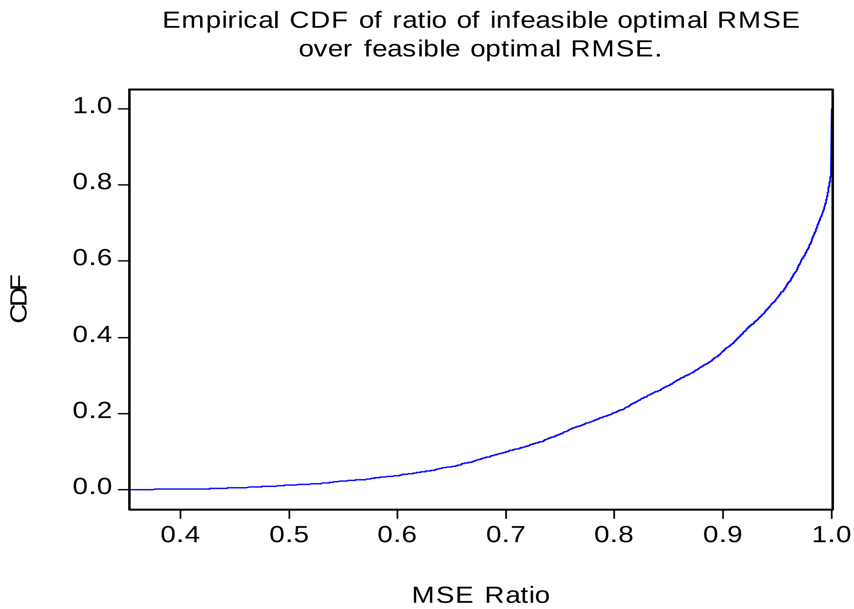

Figure 2. Cumulative distribution function (across simulations) of the ratio between the rootMSE (RMSE) of the infeasible optimally-sampled realized variance estimator and the RMSE of the feasible optimally-sampled realized variance estimator.

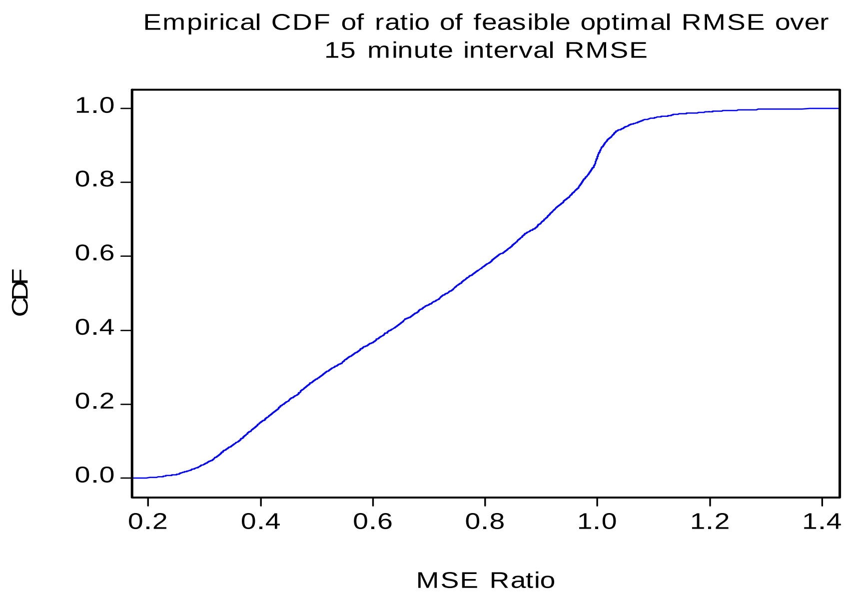

Figure 3. Cumulative distribution function (across simulations) of the ratio between the rootMSE (RMSE) of the feasible optimally-sampled realized variance estimator and the rootMSE (RMSE) of the 15-minute realized variance estimator.

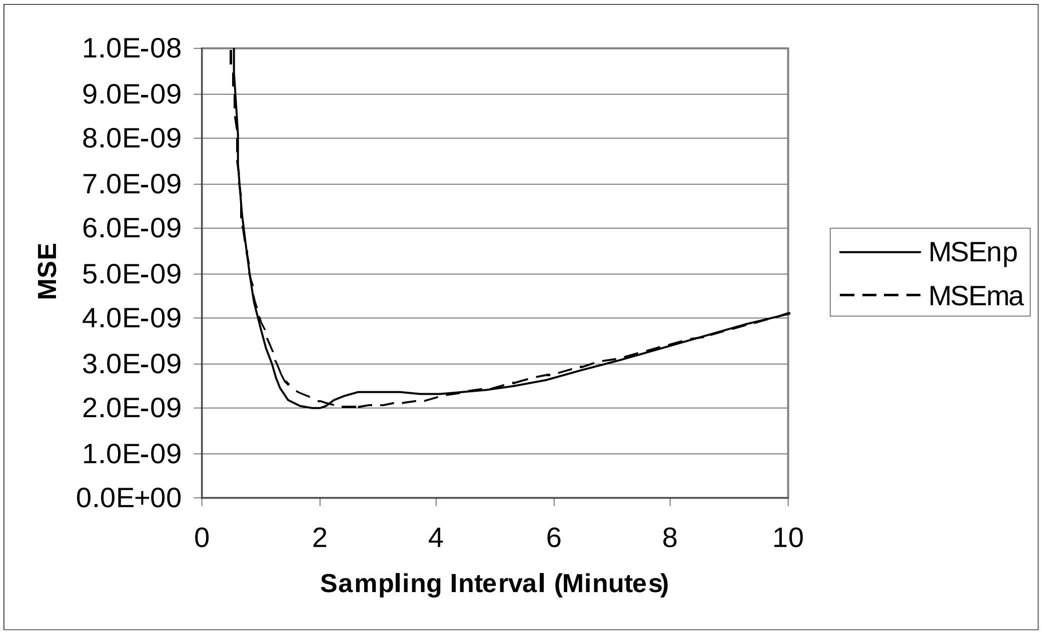

Figure 4. For IBM, we plot the nonparametric MSE for the case of dependent noise (in Theorem 5) and the MA(1) MSE (in Theorem 3).

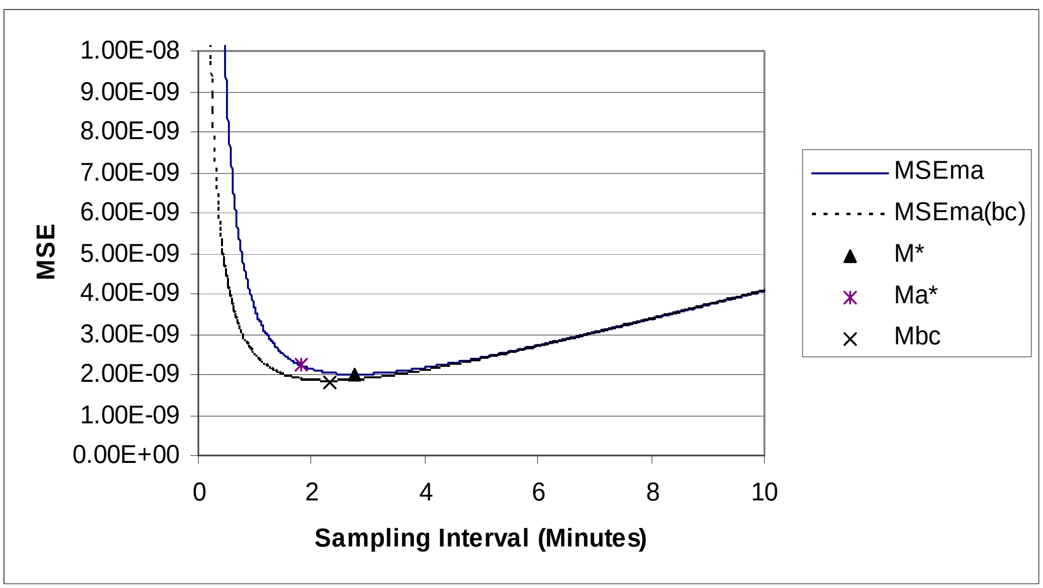

Figure 5. For IBM, we plot the MA(1) MSE (in Theorem 3) and the bias-corrected MA(1) MSE (in Theorem 4). We also report the true optimal frequency, the approximate optimal frequency in Remark 3 and the optimal frequency of the bias-corrected estimator in Remark 4.

Table 1. Profits from option pricing and option trading.

	5 Minutes			15 Minutes			30 Minutes		
	SBC	XOM	MER	SBC	XOM	MER	SBC	XOM	MER
Avg. Profits	2.76	2.78	.24	3.07	-.06	.36	4.22	.49	7.10
Avg. Put/Call Prices	92.09	67.25	118.56	92.46	67.81	119.17	92.79	67.68	119.92

We report the average profits of the optimal-sampling agent obtained by trading with 5, 15, and 30-minute agents for the three stocks SBC, XOM, and MER. Agents price at-the-money options on a $100 stock. Average Put/Call prices are given below the profits. The units are in cents.