Journal of Econometrics 147 (2008) 47–59

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error☆

Ilze Kalnina *, Oliver Linton

Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

ARTICLE INFO

Article history:
Available online 18 September 2008

JEL classification:
C12

Keywords:
Endogenous noise
Market microstructure
Realised volatility
Semimartingale

ABSTRACT

We propose an econometric model that captures the effects of market microstructure on a latent price process. In particular, we allow for correlation between the measurement error and the return process and we allow the measurement error process to have a diurnal heteroskedasticity. We propose a modification of the TSRV estimator of quadratic variation. We show that this estimator is consistent, with a rate of convergence that depends on the size of the measurement error, but is no worse than $n^{-1/6}$. We investigate in simulation experiments the finite sample performance of various proposed implementations.

1. Introduction

It has been widely recognized that using very high frequency data requires taking into account the effect of market microstructure (MS) noise. We are interested in the estimation of the quadratic variation of a latent price in the case where the observed log-price Y is a sum of the latent log-price X that evolves in continuous time and an error u that captures the effect of MS noise.

There is by now a large amount of literature that uses realized variance as a nonparametric measure of volatility. The justification is that, in the absence of market microstructure noise, it is a consistent estimator of the quadratic variation as the time between observations goes to zero. For a literature review, see Barndorff-Nielsen and Shephard (2007). In practice, ignoring microstructure noise seems to work well for frequencies below 10 min. For higher frequencies realized variance is not robust, as has been evidenced in the so-called 'volatility signature plots', see, e.g. Andersen et al. (2000).

The additive measurement error model where u is independent of X and i.i.d. over time was first introduced by Zhou (1996). The usual realized volatility estimator is inconsistent under this assumption. The first consistent estimator of quadratic variation of the latent price in the presence of MS noise was proposed by Zhang et al. (2005a) who introduced the Two Scales Realized Volatility (TSRV) estimator, and derived the appropriate central limit theory. TSRV estimates the quadratic variation using a combination of realized variances computed on two different time scales, performing an additive bias correction. It has a rate of convergence $n^{-1/6}$. Zhang (2006) introduced the more complicated Multiple Scales Realized Volatility (MSRV) estimator that combines multiple ($\sim n^{1/2}$) time scales, which has a convergence rate of $n^{-1/4}$. This is known to be the optimal rate for this problem. Both papers assumed that the MS noise was i.i.d. and independent of the latent price. This assumption, according to an empirical analysis of Hansen and Lunde (2006), "seems to be reasonable when intraday returns are sampled every 15 ticks or so". Further studies have tried to relax this assumption to allow modelling of even higher frequency returns. Aït-Sahalia et al. (2006a) modify TSRV and MSRV estimators and achieve consistency in the presence of serially correlated MS noise. Another class of consistent estimators of the quadratic variation was proposed by Barndorff-Nielsen et al. (forthcoming). They introduce realized kernels, a general class of estimators that extends the unbiased but inconsistent estimator of Zhou (1996), and is based on a general weighting of realized autocovariances as well as realized variances. They show that realized kernels can be designed to be consistent and derive the central limit theory. They show that for particular choices of weight functions they can be asymptotically equivalent to TSRV and MSRV estimators, or even more efficient. Apart from the benchmark setup where the noise is i.i.d. and independent from the latent price Barndorff-Nielsen et al. (forthcoming) have two additional sections, one allowing for AR(1) structure in the noise, another with an additional endogenous term, albeit one that is asymptotically degenerate. In discrete time framework, Robinson

☆ We would like to thank Jiti Gao, Dimitri Vayanos, Neil Shephard, Andrew Patton, Mathieu Rosenbaum, Yacine Aït-Sahalia, and two referees for helpful comments, as well as seminar participants at Perth, Alice Springs ESAM, Shanghai, and Yale. This research was supported by the ESRC and the Leverhulme foundation.

* Corresponding author.
E-mail addresses: i.kalnina@lse.ac.uk (I. Kalnina), o.linton@lse.ac.uk (O. Linton).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.09.016

(1986) shows how to deal with errors-in-variables problem when errors are possibly strongly serially correlated, contain seasonal effects and trends.

We generalize the standard additive noise model (where the noise is i.i.d. and independent from the latent price) in three directions. The first generalization is allowing for (asymptotically non-degenerate) correlation between MS noise and the latent returns. This is motivated by a paper of Hansen and Lunde (2006), where, for very high frequencies: "the key result is the overwhelming evidence against the independent noise assumption. This finding is quite robust to the choice of sampling method (calendar-time or tick-time) and the type of price data (transaction prices or quotation prices)".[1]

Another generalization concerns the magnitude of the MS noise. All of the papers above, like most of related literature, assume that the variance of the MS noise is constant and does not change depending on the time interval between trades. We call this a large noise assumption. We explicitly model the magnitude of the MS noise via a parameter α, where the $\alpha = 0$ case corresponds to the benchmark case of large noise. We allow also $\alpha > 0$ in which case the noise is "small" and specifically the variance of the noise shrinks to zero with the sample size n. The rate of convergence of our estimator depends on the magnitude of the noise, and can be from $n^{-1/6}$ to $n^{-1/3}$, where $n^{-1/6}$ is the rate of convergence corresponding to the "big" noise case when $\alpha = 0$.

How could the size of the noise "depend" on the sample size? We give a fuller discussion of this issue below, but we note here two arguments. First, there is a negative relationship between the bid-ask spread (an important component of the MS noise for transaction data) and a number of (other) liquidity measures, including number of transactions during the day. This negative relationship is a stylized fact from the market microstructure literature. See, for example, Copeland and Galai (1983) and McInish and Wood (1992). Also, Awartani et al. (2004) write that "an alternative model of economic interest [to the standard additive noise model] would be one in which the microstructure noise variance is positively correlated with the time interval". This is in principle a testable hypothesis. Using Dow Jones Industrial Average data, the authors test for and reject the hypothesis of constant variance of the MS noise across frequencies.

The third feature of our model is that we allow the MS noise to exhibit diurnal heteroscedasticity. This is motivated by the stylized fact in market microstructure literature that intradaily spreads and intradaily stock price volatility are described typically by a U-shape (or reverse J-shape). See Andersen and Bollerslev (1997), Gerety and Mulherin (1994), Harris (1986), Kleidon and Werner (1996), Lockwood and Linn (1990), and McInish and Wood (1992). Allowing for diurnal heteroscedasticity in our model has the effect that the original TSRV estimator may not be consistent because of end effects. In some cases, instead of estimating the quadratic variation, it would be estimating some function of the noise. We propose a modification of the TSRV estimator that is consistent, without introducing new parameters to be chosen. Our model is not meant to be definitive and can be generalized in a number of ways.

The structure of the paper is as follows. Section 2 introduces the model. Section 3 describes the estimator. Section 4 gives the main result and the intuition behind it. Section 5 investigates the numerical properties of the estimator in a set of simulation experiments. Section 6 illustrates the ideas with an empirical study of IBM transaction prices. Section 7 concludes. We use $\Longrightarrow$ to denote convergence in distribution.

2. The model

Suppose that the latent (log) price process $\{X_t, t \in [0, T]\}$ is a Brownian semimartingale solving the stochastic differential equation

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t, \tag{1}$$

where W_t is standard Brownian motion, μ_t is a locally bounded predictable drift function, and σ_t a càdlàg volatility function; both are independent of the process $\{W_t, t \in [0, T]\}$. The (no leverage) assumption of $\{\sigma_t, \mu_t, t \in [0, T]\}$ being independent of $\{W_t, t \in [0, T]\}$, though reasonable for exchange rate data, is unrealistic for stock price data. However, it is frequently used and makes the theoretical analysis more tractable. The simulation results suggest that this assumption does not change the result. Furthermore, in many other contexts the presence of leverage does not affect the limiting distributions, see Barndorff-Nielsen and Shephard (2002).

The additive noise model says that the noisy price Y is observed at times $t_1, \ldots, t_n$ on some fixed domain $[0, T]$

$$Y_{t_i} = X_{t_i} + u_{t_i}, \tag{2}$$

where u_{t_i} is a random variable representing measurement error. Without loss of much generality we are going to restrict attention to the case of equidistant observations with $T = 1$. This type of model was first introduced by Zhou (1996) who assumed that u_{t_i} is i.i.d. over i and independent of $\{X_t, t \in [0, 1]\}$. In this case the signal to noise ratio for returns decreases with sample size, i.e., $\mathrm{var}(\Delta X_{t_i})/\mathrm{var}(\Delta u_{t_i}) \to 0$ as $n \to \infty$, and at a specific rate such that $\lim_{n\to\infty} n\mathrm{var}(\Delta X_{t_i})/\mathrm{var}(\Delta u_{t_i}) < \infty$, which implies inconsistency of realized volatility. We are going to modify the properties of the process $\{u_{t_i}\}$ and its relation to $\{X_t, t \in [0, 1]\}$.

We would like to capture the idea that the measurement error can be small. This can be addressed by adopting a model $u_{t_i} = \sigma_\epsilon \epsilon_{t_i}$, where ϵ_{t_i} is an i.i.d. sequence with mean zero and variance one, and σ_ϵ is a parameter such that $\sigma_\epsilon \to 0$. Many authors have found small σ_ϵ in practice. As usual one wants to make inferences about data drawn from the true probability measure of the data where both n is finite and $\sigma_\epsilon > 0$ by working with a limiting case that is more tractable. In this case there are a variety of limits that one could take. Bandi and Russell (2006a) for example calculate the exact MSE of the statistic of interest, and then in Eq. (24) implicitly take $\sigma_\epsilon \to 0$ followed by $n \to \infty$. We instead take the sound and well established practice in econometrics of taking pathwise limits, that is we let $\sigma_\epsilon = \sigma_\epsilon(n)$ and then let $n \to \infty$. Such a limit with "small" noise has been used before to derive Edgeworth approximations (Zhang et al., 2005b), to calculate optimal sampling frequency of inconsistent estimator for QV_x (Zhang et al., 2005a, Eq. (53)), to estimate QV_x consistently when X follows a pure jump process and Y is observed fully and continuously Large (2007), and to estimate QV_x consistently in a pure rounding model (Li and Mykland, 2006; Rosenbaum, 2007). An example from MS modelling literature in microeconomics is Back and Baruch (2004) who show the link between the two key papers in asymmetric information modelling, Glosten and Milgrom (1985) and Kyle (1985) using a limit with small noise. In particular, they consider a limit of Glosten and Milgrom (1985) as the arrival rate of trades explodes (so the number of trades in any interval goes to infinity) and order size (and hence incremental information per trade) goes to zero, thus reaching the Kyle (1985) model as a limit. We are also mindful not to preclude the case where $\sigma_\epsilon(n)$ is "large" i.e., (in our framework)

[1] By "independent noise" Hansen and Lunde (2006) mean the combination of the i.i.d. assumption and the assumption that the noise is independent from the latent price. Our paper proposes to relax the second assumption. As to the first assumption, we do not allow for serial correlation in the noise. At the same time, we only impose approximate stationarity compared to Hansen and Lunde (2006) since we allow for intraday heteroscedasticity of the noise.

does not vanish with n, and our parameterization below allows us to do that.

We next present our model. We assume that

$$u_{t_i} = v_{t_i} + \varepsilon_{t_i} \tag{3}$$
$$v_{t_i} = \delta\gamma_n \left(W_{t_i} - W_{t_{i-1}}\right)$$
$$\varepsilon_{t_i} = m\left(t_i\right) + n^{-\alpha/2}\omega\left(t_i\right)\epsilon_{t_i}, \quad \alpha \in [0, 1/2)$$

with ϵ_{t_i} i.i.d. mean zero and variance one and independent of the Gaussian process $\{W_t, t \in [0, 1]\}$ with $E|\epsilon_{t_i}|^{4+\eta} < \infty$ for some $\eta > 0$. The functions m and ω are differentiable, nonstochastic functions of time. They are unknown as are the constants δ and α. The usual benchmark measurement error model with noise being i.i.d. and independent from the latent price has $\alpha = 0$, $\gamma_n = 0$ and $\omega(.)$ and $m(.)$ constant (see, e.g., Barndorff-Nielsen and Shephard (2002), Zhang et al. (2005a) and Bandi and Russell (2006b)).

The process for the latent log-price is motivated by the fundamental theory of asset prices, which states that, in a frictionless market, log-prices must obey a semi-martingale; we are specializing to the Brownian semimartingale case (1). We want to model log-prices at very high frequency where frictions are important and observed prices do not follow a semimartingale. One way of partly reconciling the evidence in volatility signature plots of the price behavior in very high and moderate frequencies, is to assume that observed prices can be decomposed as in (2). The first component X is a semi-martingale with finite quadratic variation, while the second component u is not a semi-martingale and has infinite quadratic variation. In particular, the increments in u are of larger magnitude than that of X, and this difference is the key in identifying the quadratic variation of X. We split the noise component u into an independent term ε that has been considered in the literature, and a 1-dependent endogenous part v, which is correlated with X due to being driven by the same Brownian motion. At the same time, v preserves the features of not being a semi-martingale and having infinite quadratic variation, the main motivation of the way ε is modelled.

There are three key parts to our model: the correlation between u and X, the relative magnitudes of u and X, and the heterogeneity of u. We have $E[u_{t_i}] = m(t_i)$ and $\text{var}[u_{t_i}] = \delta^2\gamma_n^2(t_i - t_{i-1}) + 2n^{-\alpha}\sigma_\epsilon^2(i/n)$. To have the variance of both terms in u equal, we set $\gamma_n^2 = n^{1-\alpha}$. This seems like a reasonable restriction if both components are generated by the same mechanism. In this case, both of the measurement error terms are $O_p(n^{-\alpha})$. In our model the signal to noise ratio of returns varies with sample size in a way depending on α so that only $\lim_{n\to\infty} n^{1-\alpha}\text{var}(\Delta X_{t_i})/\text{var}(\Delta u_{t_i}) < \infty$. We exploit the fact that for consistency of the TSRV estimator, it is enough to assume that noise increments are of larger order of magnitude than the latent returns, and the usual stronger assumption $\lim_{n\to\infty} n\text{var}(\Delta X_{t_i})/\text{var}(\Delta u_{t_i}) < \infty$ is not necessary.

The process ε_{t_i} is a special case of the more general class of locally stationary processes of Dahlhaus (1997). The generalization of allowing time varying mean and variance in the measurement error, allows one to capture diurnal variation in the measurement error process, which is likely to exist in calendar time. Nevertheless, the measurement error in prices is approximately stationary under our conditions, which seems reasonable.

The term v in u induces a correlation between latent returns and the change in the measurement error, which can be of either sign depending on δ. Correlation between u and X is plausible due to rounding effects, price stickiness, asymmetric information, or other reasons (Bandi and Russell, 2006c; Hansen and Lunde, 2006; Diebold, 2006).[2] In the special case that $\sigma_t = \sigma$ and $\omega\left(t_i\right) = \omega$, we find

$$\text{corr}(\Delta X_{t_i}, \Delta u_{t_i}) \simeq \frac{\delta}{\sqrt{[2\delta^2 + 2\omega^2]}}.$$

In this case, the range of correlation is limited, although it is quite wide — one can obtain up to a correlation of $\pm 1/\sqrt{2}$ depending on the relative magnitudes of δ, ω.

An alternative model for endogenous noise has been developed by Barndorff-Nielsen et al. (forthcoming). In our notation, they have the endogenous noise part such that $\text{var}(v_{t_i}) = O\left(1/n\right)$, and an i.i.d., independent from X part with $\text{var}(\varepsilon_{t_i}) = O\left(1\right)$. They conclude robustness of their estimator to this type of endogeneity, with no change to the first order asymptotic properties compared to the case where $v_{t_i} = 0$.

The focus of this paper is on estimating increments in quadratic variation of the latent price process,[3] but estimation of parameters of the MS noise in our model is also of interest. We acknowledge that not all the parameters of our model are identifiable. In particular, the endogeneity parameter may not be identified unless one knows something about the distribution of ϵ and in particular that it is not Gaussian.[4] However, other parameters are identified. In Linton and Kalnina (2007) we provided a consistent estimator of α, see also Section 6 here for empirical implementation and discussion. Estimating the function $\omega\left(\tau\right)$ would allow us to measure the diurnal variation of the MS noise. In the benchmark measurement error model this is a constant $\omega\left(\tau\right) \equiv \omega$ that can be estimated consistently by $\sum_{i=1}^{n-1}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2/2n$ (Bandi and Russell, 2006b; Barndorff-Nielsen et al., forthcoming; Zhang et al., 2005a). In our model, instead of n^{-1}, the appropriate scaling is $n^{\alpha-1}$. Such an estimator would converge to $\delta^2 + \int \omega^2\left(u\right) du$. Hence, this estimator would converge asymptotically to the integrated variance of the MS noise. Following Kristensen (forthcoming), in the special case $\delta = 0$, we could also estimate $\omega\left(\cdot\right)$ at some fixed point τ using kernel smoothing,

$$\widehat{\omega}^2\left(\tau\right) = \frac{1}{2n^{1-\alpha}} \frac{\sum_{i=1}^{n} K_h\left(t_{i-1} - \tau\right)\left(\Delta Y_{t_{i-1}}\right)^2}{\sum_{i=1}^{n} K_h\left(t_{i-1} - \tau\right)\left(t_i - t_{i-1}\right)}.$$

When the observations are equidistant, this simplifies to $\widehat{\omega}^2\left(\tau\right) = \sum_{i=1}^{n} K_h\left(t_{i-1} - \tau\right)\left(\Delta Y_{t_{i-1}}\right)^2/2n^{-\alpha}$. In the above, h is a bandwidth that tends to zero asymptotically and $K_h(.) = K(./h)/h$, where $K(.)$ is a kernel function satisfying some regularity conditions. If we also allow for endogeneity ($\delta \neq 0$), $\widehat{\omega}^2\left(\tau\right)$ estimates $\omega^2\left(\tau\right)$ plus a constant, and so we still see the pattern of diurnal variation. See Section 6 for implementation.

[2] In a recent survey of measurement error in microeconometrics models, Bound et al. (2001) emphasize 'mean-reverting' measurement error that is correlated with the signal.

[3] There is a question about whether one should care about the latent price or the actual price. This has been raised elsewhere, see Zhang et al. (2005a). We stick with the usual practice here, acknowledging that the presence of correlation between the noise and efficient price makes this even more debatable, (Aït-Sahalia et al., 2006b). Also, note that we are following the literature and estimating the quadratic variation of the latent log-price and not the latent price.

[4] Suppose that $X_{i+1} = X_i + (\sigma/\sqrt{n})z_{i+1}$ and $Y_i = X_i + \rho z_i + \sigma_\epsilon\epsilon_i$, where z_i is standard normal and ε_i is i.i.d. with mean zero and variance one. Then $r_{i+1} = Y_{i+1} - Y_i = \left(\frac{\sigma}{\sqrt{n}} + \rho\right)z_{i+1} - \rho z_i + \sigma_\epsilon\epsilon_{i+1} - \sigma_\epsilon\epsilon_i$. We have $\text{var}[r_{i+1}] = 2\left(\rho^2 + \sigma_\epsilon^2\right) + \frac{2\rho\sigma}{\sqrt{n}} + \frac{\sigma^2}{n}$, $\text{cov}[r_{i+1}, r_i] = -\left(\rho^2 + \sigma_\epsilon^2\right) - \frac{\rho\sigma}{\sqrt{n}}$, and $\text{cov}[r_{i+j}, r_i] = 0, j > 1$. Therefore, from the covariogram we obtain $\sigma^2 = n(\text{var}[r_{i+1}] + 2\text{cov}[r_{i+1}, r_i])$ but we can only identify $\rho^2 + \sigma_\epsilon^2$ not the two quantities separately. There are just two equations in two unknowns and if ϵ_i is also Gaussian, then there is no more information. If there is a non-Gaussian distribution one can identify ρ using parametric restrictions. This is similar to the classical measurement error problem, (Maddala, 1977, p 296).

3. Estimation

We suppose that the parameter of interest is the quadratic variation of X on [0, 1], denoted $QV_X = \int_0^1 \sigma_t^2 \mathrm{d}t$. Let

$$[Y,Y]^n = \sum_{i=1}^{n-1} \left(Y_{t_{i+1}} - Y_{t_i}\right)^2$$

be the realized variation (often called realized volatility) of Y, and introduce a modified version of it (*jittered* RV) as follows,

$$[Y,Y]^{\{n\}} = \frac{1}{2}\left(\sum_{i=1}^{n-K}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2 + \sum_{i=K}^{n-1}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2\right). \tag{4}$$

This modification is useful for controlling the end effects that arise due to heteroscedasticity.

Our estimator of QV_X makes use of the same principles as the TSRV estimator in Zhang et al. (2005a). We split the original sample of size n into K subsamples, with the jth subsample containing n_j observations. Introduce a constant β and c such that $K = cn^{\beta}$. The dependence of K on n is suppressed in the sequel. For consistency we will need $\beta > 1/2 - \alpha$. The optimal choice of β is discussed in the next section. By setting $\alpha = 0$, we get the condition for consistency in Zhang et al. (2005a), that $\beta > 1/2$.[5]

Let $[Y,Y]^{n_j}$ denote the jth subsample estimator based on a K-spaced subsample of size n_j,

$$[Y,Y]^{n_j} = \sum_{i=1}^{n_j-1}\left(Y_{t_{iK+j}} - Y_{t_{(i-1)K+j}}\right)^2,\ j = 1,\ldots,K,$$

and let

$$[Y,Y]^{\mathrm{avg}} = \frac{1}{K}\sum_{j=1}^{K}[Y,Y]^{n_j}$$

be the averaged subsample estimator. To simplify the notation, we assume that n is divisible by K and hence the number of data points is the same across subsamples, $n_1 = n_2 = \cdots = n_K = n/K$. Let $\overline{n} = n/K$.

Define the adjusted TSRV estimator (*jittered* TSRV) as

$$\widehat{QV}_X = [Y,Y]^{\mathrm{avg}} - \left(\frac{\overline{n}}{n}\right)[Y,Y]^{\{n\}}. \tag{5}$$

Compared to the TSRV estimator, this estimator does not involve any new parameters that would have to be chosen by the econometrician, so it is as easy to implement. The need to adjust the TSRV estimator arises from the fact that under our assumptions TSRV is not always consistent. The problem arises due to end-of-sample effects induced by heteroscedastic noise. For a simple example where the TSRV estimator is inconsistent, let us simplify the model to the framework of Zhang et al. (2005a), and introduce only heteroscedasticity in the noise, the exact form of which is to be chosen below. Let us evaluate the asymptotic bias of TSRV estimator.[6]

$$\begin{aligned} n^{1/6}E\left\{\widehat{QV}_X^{\mathrm{TSRV}} - QV_X\right\} &= n^{1/6}\left\{E[u,u]^{\mathrm{avg}} - \frac{\overline{n}}{n}E[u,u]^n\right\} + o(1) \\ &= c^{-1}n^{-1/2}\sum_{i=1}^{n-K}\left(\omega_{t_{i+K}}^2\epsilon_{t_{i+K}}^2 + \omega_{t_i}^2\epsilon_{t_i}^2\right) \\ &\quad - \left(c^{-1}n^{-1/2} - n^{-5/6}\right)\sum_{i=1}^{n-1}\left(\omega_{t_{i+1}}^2\epsilon_{t_{i+1}}^2 + \omega_{t_i}^2\epsilon_{t_i}^2\right) + o(1) \\ &= n^{-5/6}\sum_{i=1}^{n-1}\left(\omega_{t_{i+1}}^2\epsilon_{t_{i+1}}^2 + \omega_{t_i}^2\epsilon_{t_i}^2\right) \\ &\quad - c^{-1}n^{-1/2}\left\{\sum_{i=2}^{K}\omega_{t_i}^2\epsilon_{t_i}^2 + \sum_{i=n-K+1}^{n-1}\omega_{t_i}^2\epsilon_{t_i}^2\right\} + o(1). \end{aligned}$$

We see that the first and last K returns that are "ignored" by averaged subsampled realized volatility $[Y,Y]^{\mathrm{avg}} \sim [u,u]^{\mathrm{avg}}$ have to be off-set by a fraction of the noise of all returns, coming from $[Y,Y]^n \sim [u,u]^n$. For this bias correction to work, the volatility of the microstructure noise in the morning and afternoon has to be "close" to the volatility of the noise during the day. A simple counter-example that is motivated by our empirical Section 6.3 is a parabola on [0, 1], $\omega^2(i/n) = a + \left(\frac{i}{n} - 0.5\right)^2/100$, where a is any constant. In this case simple calculations give that TSRV estimator is inconsistent,

$$n^{1/6}\mathrm{E}(\widehat{QV}_X^{\mathrm{TSRV}} - QV_X) = -\frac{1}{300}n^{1/6} + o(1).$$

By contrast, *jittered RV*, $[Y,Y]^{\{n\}}$, mimics the structure of the volatility component that needs to be bias corrected for in $[Y,Y]^{\mathrm{avg}}$, which is

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n-K}\left(\omega_{t_{i+K}}^2\epsilon_{t_{i+K}}^2 + \omega_{t_i}^2\epsilon_{t_i}^2\right)$$

and so delivers a consistent estimator $\widehat{QV}_X$.

We remark that (5) is an additive bias correction and there is a nonzero probability that $\widehat{QV}_X < 0$. One can ensure positivity by replacing $\widehat{QV}_X$ by $\max\{\widehat{QV}_X, 0\}$, but this is not very satisfactory. Note, however, that we usually have $\widehat{QV}_X > \widehat{QV}_X^{\mathrm{TSRV}}$ (except for when first and last subsamples have all flat prices and so $\widehat{QV}_X = \widehat{QV}_X^{\mathrm{TSRV}}$), so the probability that $\widehat{QV}_X < 0$ is lower than the probability that $\widehat{QV}_X^{\mathrm{TSRV}} < 0$.

4. Asymptotic properties

The expansion for $[Y,Y]^{\mathrm{avg}}$ and $[Y,Y]^n$ both contain terms due to the correlation between the measurement error and the latent returns. The main issues can be illustrated using the expansion of $[Y,Y]^{\mathrm{avg}}$, conditional on the path of σ_t:

$$\begin{aligned} [Y,Y]^{\mathrm{avg}} &= \underbrace{QV_X}_{(a)} + 2\underbrace{\frac{\delta\gamma_n}{K}\int_0^1 \sigma_t \mathrm{d}t}_{(b)} + \underbrace{E[u,u]^{\mathrm{avg}}}_{(c)} \\ &\quad + O\left(\underbrace{\overline{n}^{-1/2}}_{(d)} + \underbrace{\sqrt{\frac{\overline{n}}{Kn^{2\alpha}}}}_{(e)}\right)Z, \end{aligned} \tag{6}$$

where $Z \sim N(0,1)$, while the terms in curly braces are as follows: (a) the probability limit of $[X,X]^{\mathrm{avg}}$, which we aim to estimate; (b) the bias due to correlation between the latent returns and the measurement error; (c) the bias due to measurement error; (d) the variance due to discretization; (e) the variance due to measurement error.

Should we observe the latent price without measurement error, (a) and (d) would be the only terms. In this case, of course, it is better to use $[X,X]^n$, since that has an error of smaller order $n^{-1/2}$. In the presence of the measurement error, however, both $[Y,Y]^{\mathrm{avg}}$ and $[Y,Y]^n$ are badly biased, the bias arising both from

[5] This condition is implicit in Zhang et al. (2005a) in Theorem 1 (page 1400) where the rate of convergence is $\sqrt{K/\overline{n}} = c\sqrt{n^{2\beta-1}}$.

[6] For the reader to be able to follow our calculations in the next few lines, she should use the exact definition of $\overline{n}$, $\overline{n} = \frac{n-K+1}{K}$ that Zhang et al. (2005a) use. For all other purposes differences between our and their definition are negligible.

correlation between the latent returns and the measurement error, and from the variance of the measurement error. The largest term is (c), which satisfies

$$E[u,u]^{\text{avg}} = 2\overline{n}n^{-\alpha}\left(\int_0^1 \omega^2(u)\,\mathrm{d}u + \delta^2\right) + O\left(n^{-\alpha} + \overline{n}^{-1}\right) = O\left(\overline{n}n^{-\alpha}\right),$$

i.e., it is of order $\overline{n}n^{-\alpha}$. So without further modifications, this is what $[Y,Y]^{\text{avg}}$ would be estimating. Should we be able to correct that, the next term would be $2(\delta\gamma_n/K)\int\sigma_t\mathrm{d}t$ arising from $E[X,u]^{\text{avg}}$. This second term is zero, however, if there is no correlation between the latent price and the MS noise, i.e., if $\delta = 0$. Interestingly when we use the TSRV estimator for bias correction of $E[u,u]^{\text{avg}}$, we also cancel this second term.

The asymptotic distribution of our estimator arises as a combination of two effects, measurement error and discretization effect. After correcting for the bias due to the measurement error (terms like b and c in Eq. (6)), we still have the variation due to the measurement error (term e in Eq. (6)). We can see that its contribution to the asymptotic distribution by observing how the estimator converges to the realized variance of the latent price X,

$$\sqrt{\frac{Kn^{2\alpha}}{\overline{n}}}\left(\widehat{QV}_X - [X,X]^{\text{avg}}\right) \Longrightarrow N\left(0, 8\delta^4 + 16\delta^2\int_0^1\omega^2(u)\,du + 8\int_0^1\omega^4(u)\,\mathrm{d}u\right), \quad (7)$$

The rate of convergence arises from $\text{var}[u,u]^{\text{avg}} = O\left(\overline{n}/Kn^{2\alpha}\right)$. Both parts of the noise u, which are v and ε, contribute to the asymptotic variance. The first part of the asymptotic variance roughly arises from $\text{var}[v,v]$, the second part from $\text{var}[v,\varepsilon]$ (which is nonzero even though the correlation between both terms is zero), and the third part from $\text{var}[\varepsilon,\varepsilon]$. If the measurement error is uncorrelated with the latent price, the first two terms disappear.

Should we observe the latent price without any error, we would still not know its quadratic variation due to observing the latent price only at discrete time intervals. This is another source of estimation error. From Theorem 3 in Zhang et al. (2005a) we have

$$\overline{n}^{1/2}\left([X,X]^{\text{avg}} - QV_X\right) \Longrightarrow MN\left(0, \frac{4}{3}\int_0^1\sigma_t^4\mathrm{d}t\right), \quad (8)$$

where $MN(0,S)$ denotes a mixed normal distribution with conditional variance S independent of the underlying normal random variable.

The final result is a combination of the two results (7) and (8), as well as the fact that they are asymptotically independent. The fastest rate of convergence is achieved by choosing K so that the variance from the discretization is of the same order as the variance arising from the MS noise, so set $\overline{n}^{-1/2} = \sqrt{\overline{n}/Kn^{2\alpha}}$. The resulting optimal magnitude of K is such that $\beta = 2(1-\alpha)/3$. The rate of convergence with this rule is $\overline{n}^{-1/2} = n^{-1/6-\alpha/3}$. The slowest rate of convergence is $n^{-1/6}$, and it corresponds to large MS noise case, $\alpha = 0$. The fastest rate of convergence is $n^{-1/3}$, which corresponds to $\alpha = 1/2$ case. If we pick a larger β (and hence more subsamples K) than optimal, the rate of convergence in (7) increases, and the rate in (8) decreases and so dominates the final convergence result. In this case the final convergence is slower and only the first term due to discretization appears in the asymptotic variance (see (9)). Conversely, if we pick a smaller β (and hence K) than optimal, we get a slower rate of convergence and only the second term in the asymptotic variance ("measurement error" in (9)), which is due to the MS noise.

We obtain the asymptotic distribution of $\widehat{QV}_X$ in the following theorem.

Theorem. *Suppose that $\{X_t, t\in[0,1]\}$ is a Brownian semimartingale satisfying (1). Suppose that $\{\mu_t, t\in[0,1]\}$ and $\{\sigma_t, t\in[0,1]\}$ are measurable and càdlàg processes, independent of the process $\{W_t, t\in[0,1]\}$. Suppose further that the observed price arises as in (2) with $\alpha\in[0,1/2)$. Let the measurement error u_{t_i} be generated by (3), with ϵ_{t_i} i.i.d. mean zero and variance one and independent of the Gaussian process $\{W_t, t\in[0,1]\}$ with $E|\epsilon_{t_i}|^{4+\eta} < \infty$ for some $\eta > 0$. Then,*

$$V(\sigma)^{-1/2}\overline{n}^{1/2}\left(\widehat{QV}_X - QV_X\right) \Longrightarrow N(0,1),$$

$$V(\sigma) = \underbrace{\frac{4}{3}\int_0^1\sigma_t^4\mathrm{d}t}_{\text{discretization}} + \underbrace{c^{-3}\left(8\delta^4 + 16\delta^2\int_0^1\omega^2(u)\,\mathrm{d}u + 8\int_0^1\omega^4(u)\,\mathrm{d}u\right)}_{\text{measurement error}} > 0 \quad a.s. \quad (9)$$

Remarks.

1. The quantity $V(\sigma)$ collapses to the expression in Zhang et al. (2005a) when $\omega(.)$ is constant.

2. If one could find a consistent estimator $\widehat{V}(\sigma)$ such that $\widehat{V}(\sigma) - V(\sigma) = o(1)$ a.s., then the above theorem can be strengthened along the lines of Barndorff-Nielsen and Shephard to a feasible CLT, i.e., $\widehat{V}(\sigma)^{-1/2}\overline{n}^{1/2}(\widehat{QV}_X - QV_X) \Longrightarrow N(0,1)$ from which one could obtain confidence intervals for QV_X. Without assuming $\delta = 0$ or constant $\omega(.)$, the procedure of Zhang et al. (2005a), p. 1404, would work to estimate $V(\sigma)$.

3. The main statement of the theorem can also be written as

$$n^{1/6+\alpha/3}\left(\widehat{QV}_X - QV_X\right) \Longrightarrow MN(0, cV(\sigma)),$$

where $V(\sigma) = V_1(\sigma) + c^{-3}V_2$, with $V_1(\sigma)$ being the discretization error, while MN denotes a mixed normal distribution with conditional variance $cV(\sigma)$ independent of the underlying normal random variable. We can use this to find the value of c that would minimize the conditional asymptotic variance, $c_{opt}(\sigma) = (2V_2/V_1(\sigma))^{1/3}$, provided $V_1(\sigma) > 0$, resulting in the asymptotic conditional variance $(3/2^{2/3})V_2^{1/3}V_1^{2/3}(\sigma)$. If one has consistent estimators $\widehat{V}_j(\sigma) - V_j(\sigma) = o(1)$ a.s., $j = 1,2$, then $\widehat{c}_{opt}(\sigma) = (2\widehat{V}_2(\sigma)/\widehat{V}_1(\sigma))^{1/3}$ is consistent in the sense that $\widehat{c}_{opt}(\sigma) - c_{opt}(\sigma) = o(1)$ a.s.

4. Suppose now that the measurement error is smaller than above and we have $\alpha\in[1/2,1)$ instead of $\alpha\in[0,1/2)$. Then, there is a consistency condition $\beta > 1/3$ that becomes binding and therefore optimal β allows the measurement error to converge faster than the discretization error. For $\beta = 1/3 + \Delta$ (where Δ small and positive) the rate of convergence is $\overline{n}^{-1/2} = n^{-(1-\beta)/2} = n^{-1/3+\Delta/2}$. Note that this is exactly the rate that occurs when there is no measurement error at all. So choose $\beta\in(1/3,1)$. The conclusion of the theorem becomes

$$V(\sigma)^{-1/2}n^{(1-\beta)/2}\left(\widehat{QV}_X - QV_X\right) \Longrightarrow N(0,1),$$

where $V_1(\sigma) = (4/3)\int\sigma_t^4\mathrm{d}t$. This can be shown by minor adjustments to the proofs.

5. What if $\alpha \geq 1$? This means that $[u,u]$ is of the same or smaller magnitude than $[X,X]$. In the case $\alpha = 1$ they are of the same order and identification breaks down. When $\alpha > 1$, realized volatility of observed prices is a consistent estimator of quadratic variation of latent prices, as measurement error is of smaller order. This is an artificial case and does not seem to appear in the real data.

How can we put this analysis in context? A useful benchmark for evaluation of the asymptotic properties of nonparametric estimators is the performance of parametric estimators. Gloter

and Jacod (2001) allow for the dependence of the variance of i.i.d. Gaussian measurement error ρ_n on n and establish the Local Asymptotic Normality (LAN) property of the likelihood, which is a precondition to asymptotic optimality of the MLE. For the special case $\rho_n = \rho$ they obtain a convergence rate $n^{-1/4}$, thus allowing one to conclude that the MSRV and realized kernels can achieve the fastest possible rate. They also show that the rate of convergence is $n^{-1/2}$ if ρ_n goes to zero sufficiently fast, which is the rate when there is no measurement error at all. Our estimator has a rate $n^{-1/3+\Delta}$ when there is no measurement error, which is also the rate of convergence when the noise is sufficiently small. Also, Gloter and Jacod have that for "large" noise, the rate of convergence depends on the magnitude of the noise, similarly to our results. The rate of convergence and the threshold for the magnitude of the variance of the noise is different, though.

5. Simulation study

In this section we explore the behavior of the estimator (5) in finite samples. We simulate the Heston (1993) model:

$$\mathrm{d}X_t = (\mu_t - v_t/2)\,\mathrm{d}t + \sigma_t \mathrm{d}W_t$$
$$\mathrm{d}v_t = \kappa\,(\theta - v_t)\,\mathrm{d}t + \gamma v_t^{1/2}\mathrm{d}B_t,$$

where $v_t = \sigma_t^2$, and W_t, B_t are independent standard Brownian motions.

For the benchmark model, we take the parameters of Zhang et al. (2005a): $\mu = 0.05, \kappa = 5, \theta = 0.04, \gamma = 0.5$. We set the length of the sample path to 23,400 corresponding to the number of seconds in a business day, the time between observations corresponding to one second when a year is one unit, and the number of replications to be 100,000.[7] We set $\alpha = 0$. We choose the values of ω and δ so as to have a homoscedastic measurement error with variance equal to 0.0005^2 (again from Zhang et al. (2005a)), and correlation between the latent returns and the measurement error equal to -0.1. For this we use the identity

$$\mathrm{corr}(\Delta X_{t_i}, \Delta u_{t_i}) = \frac{E(\sigma)}{\sqrt{2E(\sigma^2)}} \frac{\delta}{\sqrt{\delta^2+\omega^2}}$$

and the fact that for our volatility we have $E(\sigma) = \theta$, $\mathrm{var}(\sigma) = \theta\gamma^2/2\kappa$. We set $\beta = 2(1-\alpha)/3$. Fig. 1 shows the common volatility path for all simulations.

First, we construct different models to see the effect of varying α and the number of observations within a day. We take the values of δ and ω that arise from the benchmark model, and then do simulations for the following combinations of α and n. When interpreting the results, we should also take into account that both of these parameters change the size of the variance of the measurement error. We measure the proximity of the finite sample distribution to the asymptotic distribution by the percentage errors of the interquartile range of $\overline{n}^{1/2}(\widehat{QV}_X - QV_X)$ compared to $1.3\sqrt{V}$, the value predicted by the distribution theory. We note that this is not the same as the MSE or variance of the estimator: it can be that a very efficient estimator can be poorly approximated by its limiting distribution and vice versa. This

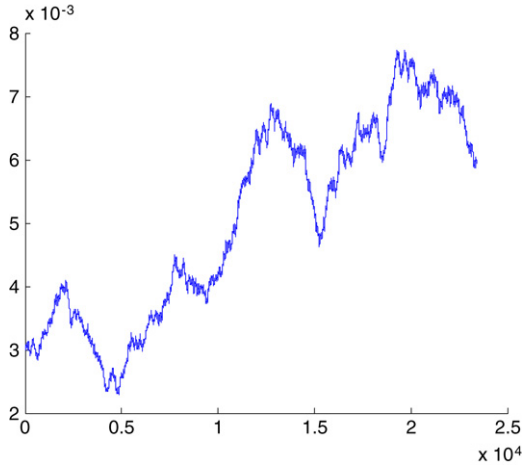


Fig. 1. The common volatility path for all simulations.

Table 1
Choices of K

$(2V_2/V_1)^{1/3} n^{\frac{2}{3}(1-\alpha)}$	Asymptotically optimal rate and c	Tables 2 and 3
$n^{\frac{2}{3}(1-\alpha)}$	Variation of above	Tables 4 and 5
$n^{\frac{2}{3}}$	Variation of above	Table 6
$\left(\frac{3RV^2}{2RQ}\right)^{1/3} n^{1/3}$	Bandi and Russell (2006a, Eq. 24)	Table 7

measure is easiest to interpret if we work with a fixed variance, i.e., when we condition on the volatility path. Hence, we simulate the volatility path for the largest number of observations, 23,400, and perform all simulations using this one sample path of volatility. The last parameter to choose is K, the number of subsamples. This is the only parameter that an econometrician has to choose in practice. We examine four different values as in Table 1 (the expressions are all rounded to the closest integer):

Table 2 contains the interquartile range errors (IQRs), in per cent, with the asymptotically optimal rate and constant (in terms of minimizing asymptotic mean squared error) for K. That is, we use $K = (2V_2/V_1)^{1/3} n^{2(1-\alpha)/3}$, rounded to the nearest integer, where V_1 and V_2 are discretization and measurement errors from (9). Table 3 contains the values of K.

First of all, for small values of α, the percentage errors decrease with n as predicted by the theory. However, we do see some large errors, and from the values of K in Table 3 we can guess this is due to the asymptotically optimal rule selecting very low c_{opt}. In fact, for the volatility path used here, $c_{opt} = (2V_2/V_1)^{1/3} = 0.0242$. Hence, another experiment we consider is an arbitrary choice $c = 1$. The next two tables (Tables 4 and 5) contain the percentage errors and values of K that result from using $K = n^{2(1-\alpha)/3}$.

The performance of this choice is much better. We can see from Table 4 that for small values of α, the asymptotic approximation improves with sample size. The sign of the error changes as α increases for given n, meaning that the actual IQR is below that predicted by the asymptotic distribution for small α and small n but this changes into the actual IQR being above the asymptotic prediction.

Another variant that does not include the unobservable α would be to use $K = n^{2/3}$.

Finally, we consider a method proposed by Bandi and Russell (2006a), which requires some discussion. They establish the exact mean squared error of TSRV under the assumptions of the independent additive noise model, and in addition they assume asymptotically constant volatility, i.e., $\int_{t_{i-1}}^{t_i} \sigma_u^2 \mathrm{d}u = \int_0^1 \sigma_u^2 \mathrm{d}u/n$ for each i, as well as $E(\epsilon^4) = 3E^2(\epsilon^2)$. Two assumptions are not satisfied in our simulation setup, the independence between

[7] Note that in the theoretical part of the paper we had for brevity taken interval $[0, 1]$. For the simulations we need the interval $[0, 1/250]$. Suppose the parameter of interest is $\int_0^\tau \sigma_t^2 dt$, the quadratic variation of X on $[0, \tau]$. In that case the asymptotic conditional variance of the theorem becomes

$$V(\sigma) = \frac{4}{3}\tau \int_0^\tau \sigma_t^4 \mathrm{d}t + c^{-3}\left(8\tau^2\delta^4 + 16\delta^2 \int_0^\tau \omega^2(u)\,\mathrm{d}u + 8\tau^{-1}\int_0^\tau \omega^4(u)\,\mathrm{d}u\right).$$

This follows by simple adjustments in the proofs. We take $\tau = 1/250$.

Table 2
IQR percentage error with $K = (2V_2/V_1)^{1/3} n^{\frac{2}{3}(1-\alpha)}$

n	α										
	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	96	186	145	120	145	114	95	78	65	54	N/A
390	94	135	110	200	156	128	143	111	89	71	59
780	67	90	108	137	107	181	151	162	119	100	76
1560	55	74	67	86	94	125	205	161	119	125	92
4680	48	47	56	58	74	96	99	117	201	144	151
5850	44	51	57	57	66	81	76	135	98	160	163
7800	45	46	52	53	68	70	90	94	109	175	134
11700	40	44	45	52	53	59	81	78	141	208	148
23400	36	40	43	46	49	58	61	79	106	123	196

Table 3
$K = (2V_2/V_1)^{1/3} n^{\frac{2}{3}(1-\alpha)}$

n	α										
	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	3	2	2	2	1	1	1	1	1	1	0
390	4	3	3	2	2	2	1	1	1	1	1
780	7	5	4	3	3	2	2	1	1	1	1
1560	11	8	7	5	4	3	2	2	2	1	1
4680	22	17	13	10	7	5	4	3	2	2	1
5850	26	19	14	11	8	6	5	3	3	2	1
7800	31	23	17	13	9	7	5	4	3	2	2
11700	41	30	22	16	12	9	6	5	3	2	2
23400	65	47	33	24	17	12	9	6	4	3	2

Table 4
IQR percentage error with $K = n^{\frac{2}{3}(1-\alpha)}$

n	α										
	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	−21	−16	−13	−7	−7	−3	−1	4	8	13	13
390	−15	−12	−7	−3	−3	1	3	6	7	12	14
780	−13	−11	−4	−2	0	0	4	5	6	11	14
1560	−9	−7	−2	−1	1	3	5	7	8	13	12
4680	−5	−3	−1	−2	1	0	3	5	6	7	11
5850	−4	−3	1	3	5	5	2	4	8	8	8
7800	−2	−2	0	1	3	2	5	3	6	8	10
11700	−3	0	0	2	2	5	4	2	6	3	8
23400	−2	1	2	1	3	4	2	6	6	6	8

Table 5
$K = n^{\frac{2}{3}(1-\alpha)}$

n	α										
	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
195	34	28	24	20	17	14	12	10	8	7	6
390	53	44	36	29	24	20	16	13	11	9	7
780	85	68	54	44	35	28	22	18	14	11	9
1560	135	105	82	64	50	39	31	24	19	15	12
4680	280	211	159	120	91	68	52	39	29	22	17
5850	325	243	182	136	102	76	57	43	32	24	18
7800	393	292	216	161	119	88	66	49	36	27	20
11700	515	377	276	202	148	108	79	58	42	31	23
23400	818	585	418	299	214	153	109	78	56	40	29

the noise and the latent returns, as well as the assumption $\int_{t_{i-1}}^{t_i} \sigma_u^2 \mathrm{d}u = \int_0^1 \sigma_u^2 \mathrm{d}u/n$ for each i (see Fig. 1). Therefore, this should be considered as another ad hoc selection method in our simulation setup. We note that this bandwidth choice results in an inconsistent estimator in our framework and in the framework of Zhang et al. (2005a) (i.e., when $\alpha = 0$, $\beta > 1/2$ is required for consistency). Note that the choice K^{BR} was derived for $\widehat{QV}^{\mathrm{TSRV}}$ without *jittering*, but this end-of-sample adjustment, though theoretically crucial, is negligible in simulations and, as we will see in the next section, also in real data. Table 7 contains the IQR percentage errors and values of K that result from using $K^{BR} = \left(3RV^2/2RQ\right)^{1/3} n^{1/3}$, where RV is the realized variance, $RV = \sum (\Delta Y_{low})^2$ and RQ is the realized quarticity, $RQ = \frac{S}{3} \sum (\Delta Y_{low})^4$. Here, Y_{low} is low frequency (15 min) returns, which gives $S = 24$ to be the number of low frequency observations during one day.

We see that the IQR errors of this choice get worse with sample size for small α, which reflects the inconsistency predicted by the theory. On the other hand the errors are small and improve with n for large α, i.e., when the noise is small. The performance is generally better than with asymptotically optimal K, except for cases that have both large n and small α, including the case $\alpha = 0$ usually considered in the literature. We notice that K^{BR} rule gives better results than the asymptotically optimal rule when it chooses

Table 6
IQR percentage error with $K = n^{\frac{2}{3}}$

n	α											
	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	K
195	−23	−23	−24	−23	−23	−21	−23	−24	−23	−24	−23	34
390	−17	−19	−19	−17	−19	−20	−18	−16	−16	−18	−18	53
780	−14	−15	−12	−15	−14	−12	−15	−15	−16	−14	−13	85
1 560	−12	−9	−10	−10	−12	−11	−11	−9	−11	−12	−9	135
4 680	−7	−2	−7	−5	−5	−7	−6	−5	−5	−6	−5	280
5 850	−6	−6	−6	−6	−6	−6	−5	−7	−6	−5	−4	325
7 800	−5	−6	−4	−4	−3	−4	−5	−4	−5	−6	−5	393
11 700	−2	−6	−3	−3	−3	−4	−2	−5	−6	−2	−3	515
23 400	−2	−2	−3	−2	−1	−2	−1	−3	−4	−2	−4	818

Table 7
IQR percentage error with $K^{BR} = \phi = \left(\frac{3RV^2}{2RQ}\right)^{1/3} n^{1/3}$

n	α											
	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	K^{BR}
195	55	46	34	29	27	21	22	19	16	18	15	6
390	67	49	37	28	23	20	17	18	15	15	14	8
780	94	65	48	32	26	22	19	16	16	14	12	10
1 560	124	81	54	36	27	24	15	14	14	13	13	13
4 680	243	146	91	54	34	24	18	16	12	14	8	18
5 850	263	155	92	53	35	24	18	11	11	11	12	20
7 800	300	182	97	60	33	26	15	13	10	11	9	22
11 700	381	223	125	68	39	24	17	11	12	9	8	25
23 400	539	305	163	86	47	28	15	13	8	8	8	32

a larger K, which is in most cases, but not all. In comparison to rules $K = n^{2(1-\alpha)/3}$ and $K = n^{2/3}$ (Tables 4 and 6, respectively), the performance of this choice is still disappointing, especially for small α. We conclude that in this setting the K^{BR} rule is not always the best choice according to our criterion.

It has been noted elsewhere that the asymptotic approximation can perform poorly, see Gonçalves and Meddahi (forthcoming) and Zhang et al. (2005b).

From Tables 2, 4 and 6 we see that magnitude of noise does not affect the quality of the asymptotic approximation. Although we see the interquartile range error having some relationship with α in Table 4 and especially Table 2, this is purely driven by changes in K. This is evidenced by Table 6 where the rule for K does not depend on α and the respective error is close to constant for the same number of observations and different α. Another conclusion here is that a good rule for K does not necessarily have to depend on α, which is convenient for practical purposes.

In a second set of experiments we investigate the effect of varying ω, which controls the variance of the second part of the measurement error, for the largest sample size. Denoting by ω_b^2 the value of ω^2 in the benchmark model, we construct models with $\omega^2 = \omega_b^2, 4\omega_b^2, 8\omega_b^2, 10\omega_b^2$, and $20\omega_b^2$. The corresponding interquartile errors are 0.96%, 1.26%, 1.93%, 2.29%, and 4.64%.

In a third set of experiments we investigate the effect of varying δ, which controls the size of the correlation of the latent returns and measurement error. Denoting by δ_b^2 the value of δ^2 in the benchmark model, we construct models with δ^2 being from $0.01 \times \delta^2$ to $20 \times \delta^2$. The exact values of δ^2, as well as the corresponding correlation between returns and increments of the noise, and the resulting interquartile errors are reported in Table 8. We can see that when the number of observations is 23,400, there is no strong effect from the correlation of the latent returns and measurement error on the approximation of the asymptotic interquartile range of the estimator.

Table 8
Effect of δ^2 on the estimates $\widehat{QV}_x$

δ^2/δ_b^2	$\mathrm{corr}(\Delta X_{t_i}, \Delta u_{t_i})$	IQR error
0.01	−0.0010	0.0133
0.05	−0.0051	0.0128
0.1	−0.0102	0.0049
0.25	−0.0254	0.0182
0.5	−0.0506	0.0037
1	−0.1000	0.0136
2	−0.1909	0.0100
4	−0.3280	0.0090
10	−0.4869	0.0130
20	−0.5351	0.0105

6. Empirical analysis

To illustrate the above ideas, we perform a small empirical analysis. We discuss estimation of α, $\omega(.)$, and the quadratic variation of the latent price. The endogeneity parameter δ is unfortunately nonparametrically unidentified and so cannot be estimated. Its sole purpose is in allowing for flexible size and sign of endogeneity, with respect to which our estimator of quadratic variation is robust.

Fig. 5 in the Appendix B shows the volatility signature of the data we use, which is IBM transaction data, year 2005. The plot indicates that market microstructure noise is prevalent at the frequencies of 10–15 min and higher. Since the volatility signature plot does not become negative, one cannot find evidence of endogeneity using the method of HL (2006). As pointed out already by HL (2006), this does not mean there is no endogeneity.

6.1. The data

We use IBM transactions data for the whole year 2005. We employ the data cleaning procedure as in HL (2006), main paper and rejoinder. First, we use transactions from NYSE exchange only as this is the main exchange for IBM. Second, we use only transactions from 9:30AM to 4:00PM. Third, for transactions with the same time stamp, we use the average price. Fourth, we remove outliers as follows. If the price is too much above the ask price or too much below the bid, we remove it. Too high means more than spread above the ask, and too low means more than spread below

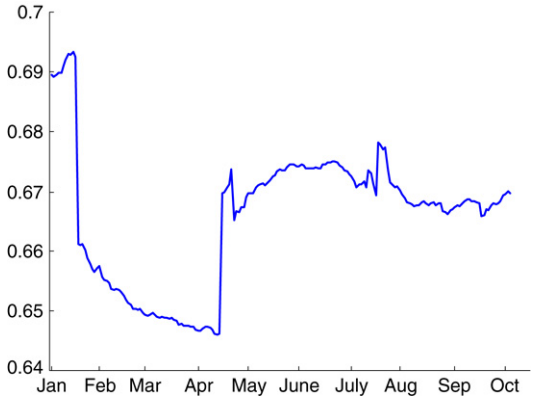


Fig. 2. Estimated α over a rolling window of 60 days (approx. 3 months). X axis shows the date of the first day in the window.

the bid. Fifth, we remove days with less than 5 h of trading (there were none). For discussion of the advantages of this procedure see HL (2006). The mean number of transactions per day in our cleaned data set is 4484 (for comparison, there are 4680 intervals of 5 s in the 6.5 h between 9:30 and 16:00).

6.2. Estimation of α

The parameter that governs the magnitude of the microstructure noise, α, can be consistently estimated. Recall that the leading term of realized volatility $[Y, Y]^n$ is $[u, u]^n$ i.e.,

$$\begin{aligned}[Y, Y]^n &= \sum_{i=1}^{n-1}(u_{t_{i+1}} - u_{t_i})^2 + o_p(n^{1-\alpha}) \\ &= n^{-\alpha}\sum_{i=1}^{n-1}(\omega_{t_{i+1}}\epsilon_{t_{i+1}} - \omega_{t_i}\epsilon_{t_i} + \delta\sqrt{n}(W_{t_{i+1}} - W_{t_i}))^2 \\ &\quad + o_p(n^{1-\alpha}) \\ &= n^{1-\alpha}c + o_p(n^{1-\alpha})\end{aligned}$$

for some positive constant c. It follows that

$$\log([Y, Y]^n/n) = -\alpha \log n + \log c + o_p(\log n).$$

We therefore estimate α by

$$\widehat{\alpha} = -\frac{\log([Y, Y]^n/n)}{\log(n)}, \tag{10}$$

see Linton and Kalnina (2007).

Although this is a consistent estimator for α, it has a bias that decays slowly. To reduce the bias, we estimate α over windows of 60 days instead of 1 day, i.e., we take our fixed interval [0, 1] to represent 3 months instead of 1 day. Fig. 2 shows the estimates over the whole year 2005 where we roll the 60 day window by 1 day. We see that $\widehat{\alpha}$ varies between 0.64 and 0.7 with an average value of 0.67.

Although this is a consistent estimator for α, it is not precise enough to give a consistent estimator of n^{α}. As a consequence, this estimator cannot be used for consistent inference for $\widehat{QV}_x$. In Linton and Kalnina (2007) we provide a sharper bias adjusted version of $\widehat{\alpha}$, $\widehat{\alpha}^{adj}$, but the adjusted estimator is not feasible as it requires knowledge of $\omega(\tau)$. This last parameter can only be consistently estimated if $\alpha = 0$ and $\delta = 0$. The lack of precision in $\widehat{\alpha}$ also prevents us from developing a test of the null hypothesis $\alpha = 0$. Therefore, the deviations of $\widehat{\alpha}$ we see in Fig. 2 provide only a heuristic evidence that the true α is positive.

6.3. Estimation of Scedastic function $\omega(.)$

Now we estimate the function $\omega(\tau)$ that allows us to measure the diurnal variation of the MS noise. In the benchmark measurement error model this is a constant $\omega(\tau) \equiv \omega$ that can be estimated consistently by $\sum_{i=1}^{n-1}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2/2n$ (Bandi and Russell, 2006c; Barndorff-Nielsen et al., forthcoming; Zhang et al., 2005a). In the special case $\alpha = 0$ and $\delta = 0$ this estimator would converge asymptotically to the integrated variance of the MS noise, $\int \omega^2(\tau)\,d\tau$. We can estimate the function $\omega^2(.)$ at a specific point τ using a simple generalization of the approach of Kristensen (forthcoming) to the case with market microstructure noise. For equidistant observations, the estimator is

$$\widehat{\omega}^2(\tau) = \frac{\sum_{i=1}^{n} K_h(t_{i-1} - \tau)\left(\Delta Y_{t_{i-1}}\right)^2}{2n^{-\alpha}}. \tag{11}$$

We pick a random day, say 77th, which corresponds to 22nd of April. Assume $\alpha = 0$ and $\delta = 0$ and note that if these assumptions are not true, the level will be incorrect, while the diurnal variation will still be correct. Fig. 3b shows the estimated function $\widehat{\omega}^2(\tau)$ using calendar time with 30 s frequency. We see that the variance of MS noise is far from being constant, and is closer to U-shape. Higher $\widehat{\omega}^2(\tau)$ at the beginning of the day and low values around 13:00 are displayed by virtually all days in 2005, while higher values of $\widehat{\omega}^2(\tau)$ at the end of the day are less common. Hence, overall, we confirm the findings of the empirical market microstructure literature that the intraday patterns are of U or reverse J shape (see references in the introduction).

6.4. Estimation of quadratic variation

Our theory predicts that original TSRV estimator is asymptotically as good as our *jittered* version if intraday volatility pattern is "close enough" to constant volatility. Visual inspection of the estimated volatilities in the previous section suggest that there is some deviation from constant volatility, so one might call for adjustment to the TSRV estimator. How important is this adjustment in practice?

We check empirically the effect of *jittering* on daily point estimates of quadratic variation using IBM data in 2005. Fig. 4a shows a plot of relative differences

$$\frac{\widehat{QV}_X - \widehat{QV}_X^{\mathrm{TSRV}}}{\widehat{QV}_X^{\mathrm{TSRV}}}$$

for every day in 2005 where we use tick time sampling (with 1-tick and $K = n^{2/3}$). The plot for 5 min calendar time sampling (CTS) is very similar. The mean of these relative differences over all days is 0.0009. Fig. 4b shows means of this relative difference for CTS, across different frequencies.[8] We see that, on average, for high frequencies, *jittering* makes very little difference. For lower frequencies the change is more visible. This arises from the fact that the *jittering* changes the TSRV estimator on two subsamples only (see Eq. (12)). The more subsamples there are, the less important our adjustment (this can also be achieved for any fixed frequency by using a larger number of subsamples than our choice $K = n^{2/3}$).

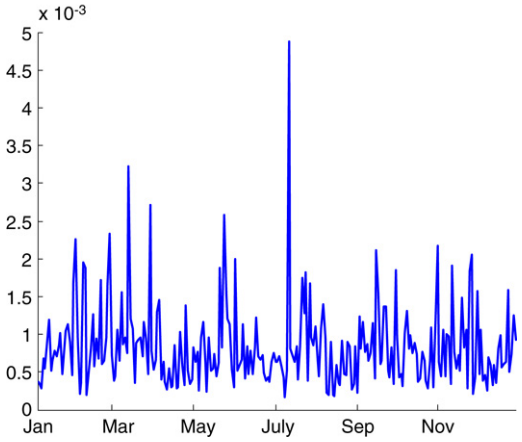
Another important observation is that *jittering* always increases the value of QV estimates, since we can write

$$\begin{aligned}\widehat{QV}_X^{\mathrm{TSRV}} &= \widehat{QV}_X + \frac{1}{2}\left(\sum_{i=1}^{K-1}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2 + \sum_{i=n-K+1}^{n}\left(Y_{t_{i+1}} - Y_{t_i}\right)^2\right) \\ &> \widehat{QV}_X. \end{aligned} \tag{12}$$

[8] This average excludes October 27. On this day our estimator, when calculated on frequencies above 7 min, became several times bigger than TSRV estimator.

(a) Squared returns. (b) Estimated function $\omega^2(.)$.

Fig. 3. IBM transactions data, 22nd of April 2005.



(a) Daily differences, 1-tick sampling. (b) Average daily differences, CTS.

Fig. 4. What is the relative difference $\frac{\widehat{QV}_X - \widehat{QV}_X^{TSRV}}{\widehat{QV}_X^{TSRV}}$ from our adjustment to the TSRV estimator?

The more there is variation in the beginning of the day and the end of the day, the larger is the adjustment. This implies that *jittering* partly alleviates the problem that the usual TSRV estimator can sometimes become negative. With our data set, the only negative value (though very small) we saw was on February 28 when we calculated TSRV estimator with 10 min CTS frequency. The *jittered* version was positive.

We conclude that for most applications our estimator is very close to the TSRV estimator, and so for practical applications plain TSRV estimator can be used, without adjustment for heteroscedastic market microstructure noise. As a result, as far as point estimates are concerned, the existing empirical studies of TSRV estimator are still valid in our theoretical framework. See, for example, investigations of forecasting performance in Aït-Sahalia and Mancini (forthcoming), Andersen et al. (2006), Bandi et al. (2007), and Ghysels and Sinko (2006).

7. Conclusions and extensions

In this paper we showed that the TSRV estimator is consistent for the quadratic variation of the latent (log) price process when the measurement error is correlated with the latent price, although some adjustment is necessary when the measurement error is heteroscedastic. We also showed how the rate of convergence of the estimator depends on the magnitude of the measurement error.

Inference for TSRV estimator is robust to endogeneity of the measurement error. Provided the suggested adjustment to the estimator is implemented to preserve consistency, inference is also robust to heteroscedasticity of the noise. However, since the rate of convergence depends on the magnitude of the noise, inference is not robust to possible deviations from assumptions about this magnitude. We plan to investigate this question further.

Other examples where the inference question needs to be solved include autocorrelation in measurement error (as in Aït-Sahalia et al. (2006a)), or other generalizations to the independent additive error model (Li and Mykland, 2007). Gonçalves and Meddahi (forthcoming) have recently proposed a bootstrap methodology for conducting inference under the assumption of no noise and shown that it has good small sample performance in their model. Zhang et al. (2005b) have developed Edgeworth expansions for the TSRV estimator, and it would be very interesting to use this for analysis of inference using bootstrap. The results we have presented may be generalized to cover MSRV estimators and to allow for serial correlation in the error terms, although in both cases the notation becomes very complicated.

Appendix A

We assume for simplicity that $\mu \equiv 0$ in the sequel. Drift is not important in high frequencies as it is of order dt, while the diffusion