Discussion of Aït-Sahalia and Barndorff-Nielsen and Shephard

Oliver Linton* and Ilze Kalnina†
London School of Economics

October 17, 2005

We discuss the issue of estimating quadratic variation of a discretely observed continuous time stochastic price process in the presence of measurement error induced by market microstructure. This issue has come to the forefront in recent months and is part of the research program described in both papers. It builds on the work of Barndorff-Nielsen and Shephard who established distribution theory for quadratic variation estimators without measurement error. We first review the main model and results.

An underlying continuous time log price X_t is observed discretely at times $t_1, \ldots, t_n$ in some fixed interval and is measured with error, i.e., one observes only $\{Y_{t_1}, \ldots, Y_{t_n}\}$ with

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i},\ i = 1, \ldots, n \tag{1}$$

where ε_{t_i} is an additive measurement error. The spacing of the data $\delta_i = t_{i+1} - t_i$ goes to zero with sample size and in most of the work is assumed to be deterministic and $O(n^{-1})$. In the strongest version of the model, ε_{t_i} are i.i.d. and independent of X_{t_i}. The parameter of interest is the quadratic variation of X on some interval. This is a classical measurement error model without feedback, and the usual realized volatility estimators of quadratic variation are inconsistent. In common with other measurement error models [Robinson (1986)] there has to be some difference between the signal and noise for identification, or some instrument that can purge out the measurement error [Bound, Brown, and Mathiowetz (2001)]. In this case it appears to come from the fact that the variance of the signal $\mathrm{var}[\Delta X_t]$ is small, $O(n^{-1})$, compared with the variance of the noise $\mathrm{var}[\Delta\varepsilon_t]$, which is $O(1)$ (where n

*Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: lintono@lse.ac.uk

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: i.kalnina@lse.ac.uk

is the full sample size). What makes this problem difficult is that it is the noise that dominates in the differenced scale. This can be overcome because aggregation over many periods raises the variance of the signal ΔX_t but has no effect on the variance of the noise $\Delta\varepsilon_t$. The authors have established under quite general conditions a class of estimators with convergence rates of order $n^{-1/4}$ for the MSRV (Multiple Scale Realized Volatility) estimators and order $n^{-1/6}$ for the simpler TSRV (Two Scale Realized Volatility) class. The estimators are explicitly defined and involve an additive bias correction. The assumptions have now been weakened in various directions to allow specifically for autocorrelated measurement error of quite general geometric mixing type [Aït–Sahalia, Mykland and Zhang 2005]. However, in the presence of autocorrelation the bias correction part of the estimator must be modified in a non-trivial way.

Note that model (1) only makes sense in the discrete time as a triangular array formulation since there can be no continuous time process $t \mapsto \varepsilon_t$ with ε_t i.i.d. for all t. In addition, there are some rather strong assumptions being made about the error term. The strongest set of assumptions are some subset of:

(A1) ε is independent of X

(A2) ε has constant scale regardless of the size of the observation interval

(A3) ε is homoskedastic and has no variation in mean

(A4) ε is i.i.d.

The figure shows the intraday price sequence for General Motors for a typical day in the 1990's; there is evidently a great deal of discreteness, which would appear to be inconsistent with assumptions A1-A4. A common statistical model for this type of discreteness is a rounding model, see for example Gottlieb and Kalay (1985). Suppose that we observe

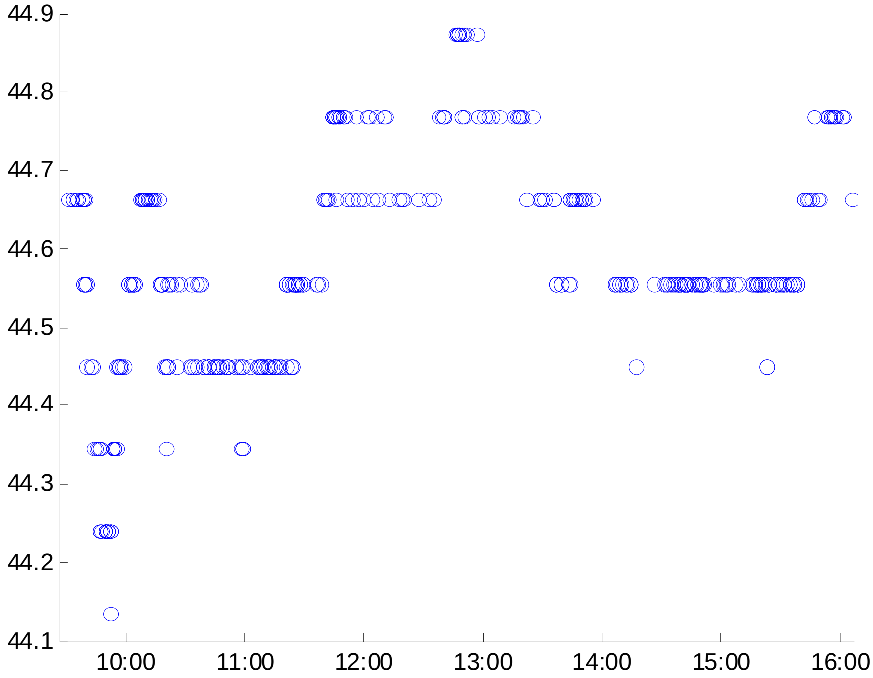
$$Y_{t_i} = d[X_{t_i}/d], \tag{2}$$

where $[x]$ denotes the closest integer to x and d is a parameter that measures the size of the discreteness (for simplicity we consider rounding on log-prices). Then we can write

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad \text{where } \varepsilon_{t_i} = d[X_{t_i}/d] - X_{t_i} \in [-d/2, d/2].$$

Although this is in the form (1), the ε_{t_i} violates some of the assumptions: ε_{t_i} is not independent of X_{t_i} and is also autocorrelated when X_{t_i} is. This model can be viewed as a simple example of a more general class of nonlinear measurement error models.

timeticks



1.pdf

Figure 1. Intraday price series for GM. August 3rd 1994

Errors due to discreteness also have another effect. The minimum price variation increment is always defined in terms of price, but we are modelling log-prices, thus the magnitude of error induced by discreteness varies depending on the level of the efficient price. Regarding (A2), while one might argue that the bid-ask spread does not depend on the frequency of trades, this is much more difficult to justify in the case of adjustment of the price due to a learning process of the market participants. It has also been shown in practice that these assumptions do not hold. For example, Hansen and Lunde (2006) have shown that the noise is correlated with the efficient price, and that the properties of the noise have changed substantially over time.

We discuss a weakening of the main assumptions to allow the measurement error to be small in an explicit pathwise asymptotic sense and only require a sort of local stationarity on the measurement error. We also allow for correlation between returns and the measurement error. The main consistency result holds in this more general setting although the rates of convergence are affected by a scale parameter denoted α.

Suppose that the observation times $t_i = i/n$ $i = 1, \dots n$ are equally spaced in $[0, 1]$ and that

$$\varepsilon_{t_i} = \mu(i/n) + n^{-\alpha/2}\sigma_\epsilon(i/n)\epsilon_{t_i}, \tag{3}$$

where ϵ_{t_i} are i.i.d. mean zero and variance one, $\alpha \in [0, 1)$, and $\sigma_\epsilon(.)$ and $\mu(.)$ are smooth functions [in fact, they could have a finite number of jump discontinuities]. This is a special case of the more general class of locally stationary processes of Dahlhaus (1997). For pedagogic reasons we suppose that

$$X_{t_{i+1}} = X_{t_i} + u_{t_{i+1}}/\sqrt{n}, \tag{4}$$

where u_{t_i} is i.i.d. and that $\mathrm{cov}(u_t, \epsilon_s) = \rho_{u\varepsilon} 1(s = t)$ [this process converges to a geometric Brownian motion as $n \to \infty$]. The parameter of interest is the quadratic variation of X on $[0, 1]$, denoted QV_X. The usual benchmark measurement error model here has $\alpha = 0$ and $\sigma_\epsilon(.)$ and $\mu(.)$ constant and $\rho_{u\varepsilon} = 0$. The generalization to 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. Correlation between u and ε is also plausible due to rounding effects or other reasons [Hansen and Lunde (2006), Diebold (2006)]. In a recent survey of measurement error in microeconometrics models, Bound, Brown, and Mathiowetz (2001) emphasize 'mean-reverting' measurement error that is correlated with the signal.

Allowing a non-constant scaling factor ($\alpha > 0$) seems natural from a statistical point of view since the ε_{t_i} represent outcomes that have happened in the small interval $[(i-1)/n, i/n]$; the scale of this distribution ought to reduce as the interval shrinks, i.e., as $n \to \infty$, at least for some of the components of the market microstructure noise. Many authors argue that measurement error is small; small is what the sampling interval is also argued to be and asymptotics are built off this assumption. Assumption (3) implies that $\mathrm{var}[\Delta\varepsilon_{t_i}] = O(n^{-\alpha})$, and nests both the i.i.d. case ($\alpha = 0$) and the diffusion case ($\alpha = 1$) but allows also for a range of meaningful intermediate cases. Indeed, Zhang, Mykland, and Aït-Sahalia (2005, p 11) implicitly allow this structure. However, their later work in Theorem 1 and 4 appear to explicitly rule this case out.

In the $\alpha = 0$ case, the measurement error persists at any frequency, so strictly speaking according to the asymptotics the measurement error is of the same magnitude as the true returns at the daily frequency. By taking $\alpha > 0$ we predict measurement error to be smaller magnitude than true returns at daily frequency.

Another advantage with the small measurement error assumption is that it provides a justification for the additivity as an approximation to a more general measurement error model. Suppose that for some function f with $f(0, 0) = 0$, $\partial f(0, 0)/\partial x = 1$ and $\partial f(0, 0)/\partial\varepsilon = 1$,

$$\Delta Y_t = f(\Delta X_t, \Delta\varepsilon_t). \tag{5}$$

This nests the additive model $f(x, z) = x + z$. Furthermore, when both $\Delta X, \Delta\varepsilon$ are 'small' $\Delta X + \Delta\varepsilon$ is a valid first order approximation so that quadratic variation computed from model (5) will behave to first order as if data came from model (1).

We just treat the TSRV estimators. Let

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

be the realized variation of Y, let $[Y,Y]^{\overline{n}}$ denote the subsample estimator based on a K-spaced subsample of size $\overline{n}$ with $K \times \overline{n} = n$, and let $[Y,Y]^{\overline{n}}_{avg}$ denote the averaged $\overline{n}$-subsample estimator. It can be shown that

$$[Y,Y]^{\overline{n}} = QV_X + O_p(\overline{n}^{-1/2}) + 2\overline{n}n^{-\alpha}\sigma_\epsilon^2 + 2n^{-(1+\alpha)/2}\overline{n}\sigma_\epsilon\rho_{u\varepsilon} + O_p(\overline{n}^{1/2}n^{-\alpha}), \tag{6}$$

where $\sigma_\epsilon^2 = \int_0^1 \sigma_\epsilon^2(u)du$ and $\sigma_\epsilon = \int_0^1 \sigma_\epsilon(u)du$. The cross-product term now yields an additional bias term due to the covariance between returns and measurement error. This term is slightly smaller than the main measurement error bias under our conditions i.e., $\sum_{i=1}^n \Delta X_{t_i}\Delta\varepsilon_{t_i} = n^{-(1+\alpha)/2}n\sigma_\epsilon\rho_{u\varepsilon} + O_p(n^{-\alpha/2})$. The validity of this expansion follows along the lines of Zhang, Mykland, and Aït-Sahalia (2005) with additional steps coming from the Taylor expansions that establish the smallness of $\mu(K(i+1)/n) - \mu(Ki/n)$ and $\sigma_\epsilon^2(K(i+1)/n) - \sigma_\epsilon^2(Ki/n)$ in relative terms for any K with $K/n \to 0$ fast enough. Then, averaging over all similar subsamples induces the usual variance reduction in both of the stochastic terms due to the measurement error. Define the bias corrected estimator (the TSRV)

$$\widehat{QV}_X = \frac{[Y,Y]^{\overline{n}}_{avg} - \left(\frac{\overline{n}}{n}\right)[Y,Y]^n}{1-\frac{\overline{n}}{n}}. \tag{7}$$

This is consistent for QV_X under some conditions on K. Suppose that $K = n^\beta$ and $\overline{n} = n^{1-\beta}$. It can be shown that

$$\widehat{QV}_X \simeq QV_X + O_p(n^{-(1-\beta)/2}) + O_p(n^{(1-2\beta)/2}n^{-\alpha}), \tag{8}$$

for which the error is minimized by setting $\beta = (2-2\alpha)/3$; this yields $\beta = 2/3$ when $\alpha = 0$ and convergence rate of $n^{-1/6}$. In general the required amount of averaging is less than $O(n^{2/3})$ because the noise is smaller, and the convergence rate is faster and is $n^{-(1+2\alpha)/6}$. As $\alpha \to 1$ the rate of convergence increases to $n^{-1/2}$ (but at $\alpha = 1/2$ lack of identification takes over). With dependent noise one should replace $[Y,Y]^n$ by $[Y,Y]^{\overline{\overline{n}}}_{avg}$ for some other $\overline{\overline{K}}, \overline{\overline{n}}$ set, but the above results would carry over like in Aït-Sahalia, Mykland, and Zhang (2005).

In conclusion, the TSRV estimator is consistent in this more general setting. Indeed, formulas (58) and (62) of Zhang, Mykland, and Aït-Sahalia (2005) remain valid under the sampling scheme (3) provided $E(\varepsilon^2)$ is interpreted according to our specification as half the integrated variance of ε_{t_i}, $\sigma_\epsilon^2/2n^\alpha$. Thus a data-based rule derived from their formula can be implemented for selecting the optimal K.

The magnitude of α might be of interest in itself: it measures the relative sizes of the signal to noise ratio. It also governs the best achievable rates for estimating QV_X. To estimate α we use the asymptotic relation (6) for n large assuming that $\alpha < 1$. Therefore, let

$$\widehat{\alpha} = -\frac{\ln\left(\frac{[Y,Y]^n}{n}\right)}{\ln(n)}.$$

This is a consistent estimator of α. Unfortunately, it is badly biased and the direction of the bias depends on whether σ_ϵ^2 is larger or smaller than 1/2. A bias corrected estimator is

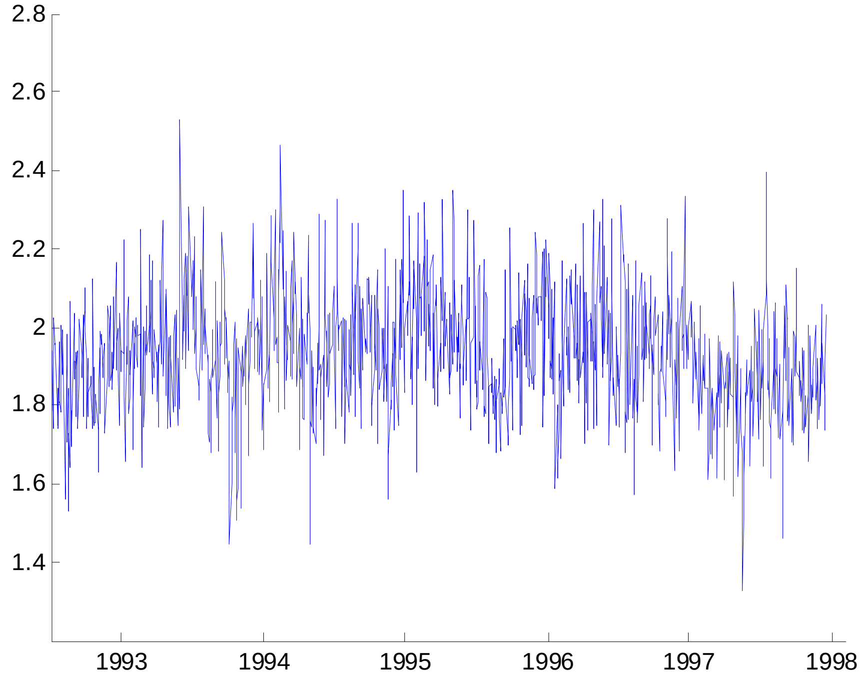
$$\widehat{\alpha}_{bc} = -\frac{\ln\left(\frac{[Y,Y]^n}{n}\right) - \ln\left(2\sigma_\epsilon^2\right)}{\ln(n)},$$

which performs much better in simulations but is infeasible due to the presence of σ_ϵ^2. Under some conditions one can obtain that

$$\sqrt{n}\log n\left(\widehat{\alpha}_{bc} - \alpha\right) \Longrightarrow N\left(0, (2 + \kappa_4/2)\sigma_\epsilon^4/(\sigma_\epsilon^2)^2\right),$$

where $\sigma_\epsilon^4 = \int_0^1 \sigma_\epsilon^4(u)du$ and κ_4 is the fourth cumulant of ϵ_{t_i}. In the normal homoskedastic case the limiting distribution is $N(0,2)$. Note that the consistency of $\widehat{\alpha}$ is robust to serial correlation in ϵ_{t_i} - if there is serial correlation then σ_ϵ^2 in $\widehat{\alpha}_{bc}$ should be replaced by the long run variance of ϵ_{t_i} and the limiting variance becomes more complicated. Below we show some estimated α values for daily IBM stock series during the 1990's. The typical values are quite large and greater than one in most cases, although there is a great deal of variation over time. We have tried a number of different implementations including varying $\overline{n}$, replacing $[Y,Y]^n/n$ by $[Y,Y]_{avg}^{\overline{n}}/\overline{n}$ but obtain similar results in all cases.

alphas



2.pdf

Figure 2. Estimated daily $\widehat{\alpha}$ for GM, for every day during period Jan 4, 1993 - May 29, 1998

Is there some way of estimating α better and estimating also σ_ϵ^2 or even $\sigma_\epsilon^2(u)$? The obvious estimator of σ_ϵ^2 doesn't work here because

$$\begin{aligned} n^{\widehat{\alpha}}\frac{[Y,Y]^n}{2n} &= \exp(\log(n^{\widehat{\alpha}}/n^{\alpha}))n^{\alpha}\frac{[Y,Y]^n}{2n} = \exp((\widehat{\alpha}-\alpha)\log n)n^{\alpha}\frac{[Y,Y]^n}{2n} \\ &= \exp(-\ln\left(2\sigma_\epsilon^2\right))2\sigma_\epsilon^2 + o_p(1) = 1 + o_p(1). \end{aligned}$$

When $\alpha = 0$ it is possible to consistently estimate the function $\sigma_\epsilon^2(u)$ but otherwise not.

We make one final set of remarks about the rounding model. It may well fit into the general framework described after (3) but there is one area of concern. The error term and its difference may not necessarily be geometric mixing. Since the process X_{t_i} is not mixing we can't expect ε_{t_i} to be mixing. Then, although ΔX_{t_i} is mixing there is no guarantee that $\Delta\varepsilon_{t_i}$ is mixing due to the nonlinearity involved (rounding and differencing are not commutable operators). Lack of mixingness would potentially cause the bias correction method not to work since sums like $\sum(\Delta\varepsilon_{t_i})^2/n$ may not converge in probability. However, we have simulated various processes of this type and our simulations suggest better news, specifically, the correlogram of the differenced noise decays rapidly. Furthermore, $\sum(\Delta\varepsilon_{t_i})^2$ diverges in probability but $\sum(\Delta\varepsilon_{t_i})^2/n$ converges to zero in probability. The terms $\sum_{i=1}^n \Delta X_{t_i}\Delta\varepsilon_{t_i}$ grows very slowly. This suggests that the TSRV and MSRV are consistent under this sampling scheme. It may be that the rounding model has become less significant over time as the NYSE has moved from 1/8ths to 1/100ths as the smallest possible unit of pricing. Nevertheless, this model can be viewed as a representative of a more general class of nonlinear measurement error models so the issues raised in its analysis may carry over to other situations.

In conclusion, the results for TSRV type estimators of quadratic variation remain valid in quite general settings and without parametric assumptions. This is partly due to the simple explicit structure of the estimator. Estimators like bi-power variation can be similarly analyzed under measurement error but are likely to require stronger assumptions on the distribution of the noise. We think this is an interesting and lively area of research and we congratulate the authors on their work.

References

[1] Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005). Ultra high frequency volatility estimation with dependent microstructure noise. Unpublished paper: Department of economics, Princeton University

[2] Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. In the Handbook of Econometrics, Eds. J.J. Heckman and E. Leamer vol. 5. 3705-3843.

[3] Dahlhaus, R. (1997): "Fitting time series models to nonstationary processes," Annals of Statistics 25, 1-37.

[4] Diebold, F.X. (2006). On Market Microstructure Noise and Realized Volatility. Discussion of Hansen and Lunde (2006).

[5] Gottlieb, G., and A. Kalay (1985). Implications of the Discreteness of Observed Stock Prices. The Journal of Finance XL, 135-153.

[6] Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). Journal of Business and Economic Statistics 24. Forthcoming

[7] Robinson, P.M. (1986): "On the errors in variables problem for time series." Journal of Multivariate Analysis 19, 240-250.

[8] 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. Forthcoming