Comment on "Realized variance and market microstructure noise" by Peter Hansen and Asger Lunde*

Yacine Aït-Sahalia, Per A. Mykland and Lan Zhang

October 6, 2005

We enjoyed reading the Hansen-Lunde paper (HL thereafter), and are pleased to be able to contribute some comments. We raise some issues which we feel are important. Some of them are addressed in our own work, while some are open questions. We certainly believe that this is an important area where there will be substantial further developments over the next few years.

1 Covariation Between the Efficient Price and the Noise

Exploring a possible correlation between the efficient price and the noise, as HL do, is an exciting and challenging task. By examining the volatility signature plots of trades and quotes, HL report that RV estimates based on quotes at very high frequency decrease. This is different from many earlier findings on volatility signature plots based on transaction prices. It is important to figure out how much of that is driven by the pre-processing of the data in the HL paper, and we discuss below the role that this might have played in delivering that result.

A different issue is that volatility signature plots give us a point estimate of RV at different sampling frequencies. A substantial increase or decrease of RV at high sampling frequencies can signify an upward bias or downward bias. But, what magnitude of change in RV is large enough to be considered as a signal for bias? This can be answered by superimposing the point-wise interval estimates of RV on the original plot. It seems that HL attempts to address this issue by providing a 95% confidence band of $\bar{RV}^{(1tick)}_{ACNW_{30}}$. However, the provided 95% confidence band seems to represent the variation between ask, bid and trades, rather than displaying the uncertainty

*Yacine Aït-Sahalia is Professor, Department of Economics and Bendheim Center for Finance, Princeton University and NBER, Princeton, NJ 08544-1021. E-mail: yacine@princeton.edu. Per A. Mykland is Professor, Department of Statistics, The University of Chicago, Chicago, IL 60637. E-mail: mykland@galton.uchicago.edu. Lan Zhang is Assistant Professor, Department of Finance, University of Illinois at Chicago, Chicago, IL 60607, E-mail: lanzhang@uic.edu, and Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: lzhang@stat.cmu.edu. Financial support from the NSF under grants SBR-0350772 (Aït-Sahalia) and DMS-0204639 (Mykland and Zhang) is gratefully acknowledged.

in the estimator $\bar{RV}^{(1tick)}_{ACNW_{30}}$ itself. In other words, we do not know whether the dipping at the higher sampling frequency in HL's Figure 1 is statistically significant or not. We know that the inconsistent estimator $\bar{RV}^{(1tick)}_{ACNW_k}$ would have a non-negligible bias and stochastic term, thus the point-wise confidence intervals could be quite wide. We explore this further below, see Table 1. We also explore the connection between this finding and the identifiability, or lack thereof, of certain noise models.

2 Identifiability, Consistency, and Efficiency

It is important at this point to remind ourselves what the rationale is for (nonparametrically) estimating certain quantities on the basis of high frequency data. The rationale is that these quantities, such as volatility within a fixed time period – say, one day, can be computed with a great deal of precision. This is the case when there is no microstructure noise, as seen in Jacod (1994) and Jacod and Protter (1998) where the asymptotic distribution can be found in great generality; for specific examples, see Zhang (2001) and Barndorff-Nielsen and Shephard (2002). But if, for whatever reason, one cannot find precise estimates of volatility for a given day, one is probably better off using a more parametric approach, for example, one could model the inter- and/or intraday behavior, using a GARCH model or in continuous time dealing with the noise as in Aït-Sahalia, Mykland, and Zhang (2005a).

So, what are the desirable (asymptotic) properties in an estimator? There are various ways of spelling out this requirement. One is *consistency*, which says that with enough data, the volatility (say) can be closely estimated with the estimator at hand. An even more basic requirement – on the model – is *identifiability*, which says that there exists some estimator which is consistent. Finally, there is *efficiency*, which means that one is making good use of the data at hand.

Unfortunately, the estimators that are used in HL, are *not* consistent (as was recognized by Zhou, see p. 47 in Zhou (1996) and p.114 in Zhou (1998).) Of course, perhaps consistency is not an absolute requirement; one could imagine situations where an estimator is inconsistent, but has only a small error. For the Zhou estimators used in the HL paper, however, Table 1 documents that the lack of consistency does generate substantial estimation errors.

Because their estimator is inconsistent, HL focus instead on unbiasedness and asymptotic unbiasedness. It is, of course, desirable that an estimator be approximately unbiased, but one should be aware that the property of unbiasedness alone can guarantee very little in some cases. To give an extreme example, suppose that (in the absence of noise) log prices have constant drift and diffusion coefficients, $dp_t = \mu dt + \sigma dW_t$. In this case, $E[(p_T - p_0)/T] = \mu$, but the unbiased estimator $(p_T - p_0)/T$ is hardly a good approximation for μ. In fact, this estimator over any finite time span T remains constant, irrespectively of how frequently sampling occurs. Results such as Corollary 3, therefore, have to be taken carefully, as they could indicate little about our ability to estimate precisely the quantity of interest. This is also the case for the results in Section 4.1 in HL.

We should also note that the theory in the HL paper does not allow for a leverage effect, i.e.,

number of observations m	100	1,000	2,000	5,000	10,000
	$\lambda = 1\%$				
Standard RV	200%	2,000%	4,000%	10,000%	20,000%
Zhou's 1 step corrected RV	28.3%	89.4%	126.5%	200.0%	282.8%
Two Scales RV	21.4%	14.6%	13.0%	11.1%	9.9%
	$\lambda = 0.5\%$				
Standard RV	100%	1,000%	2,000%	5,000%	10,000%
Zhou's 1 step corrected RV	14.4%	44.7%	63.2%	100.0%	141.4%
Two Scales RV	17.0%	11.6%	10.3%	8.8%	7.9%

Table 1: Root Mean Squared Error for three estimators.

Note: This table reports the RMSE as a fraction of IV, to first order asymptotically. λ is the noise to signal ratio and m is the sample size. For simplicity, it is assumed for the purpose of this calculation that σ^2 is constant. The RMSE for the standard RV estimator is $2m\lambda$ (the bias is the dominant quantity); for Zhou's 1 step corrected RV, it is $(8m)^{1/2}\lambda$; and for our TSRV estimator, it is $96^{1/6}m^{-1/6}\lambda^{1/3}$.

a correlation between the Brownian motions driving the asset price and that driving its stochastic volatility, which empirically is substantial (of the order of -0.75). For this, however, one needs the machinery of stochastic calculus, and we refer to Zhang, Mykland, and Aït-Sahalia (2002) and Zhang (2004) for its application to this problem. The same goes, to some extent, for the issue of allowing for a drift term $\mu_t dt$ in the efficient price.

In the case of independent noise, and noise with dependence in tick time, the volatility is identifiable even though HL's estimators are inconsistent. On the other hand, the volatility under the assumption that noise and efficient price are dependent seems to not be identifiable, as in the development right after Corollary 3. In other cases, the lack of consistency could also mask the lack of identifiability, as in Section 4.1. We shall elaborate more on identifiability in the next section.

Note that by using the methodology introduced by Zhang, Mykland, and Aït-Sahalia (2002), including a semimartingale model for the σ_t process and the device of letting $k \to \infty$ at an appropriate rate, Zhou (1996)-Zhou (1998)'s estimator can be improved to be consistent. A standard error for such an estimator can presumably be found by mimicking the methods we developed in our two scales paper, Zhang, Mykland, and Aït-Sahalia (2002). This is recognized by HL (section 4.2), but unfortunately, they do not use this or any other consistent version in their analysis, nor do they compute the variance of the estimator that would result.

3 Identifiable Noise Models

In principle, one could build a complex model to relate microstructure noise to the efficient price . In practice, however, separating the signal from the noise is not that simple. There are several difficult issues concerning how to model the noise. First of all, the noise can only be distinguished from the efficient price under fairly careful modelling. In most cases, the assumption that the noise is stationary, alone, as in HL's Assumption 2, is not enough to make the noise identifiable. For example, one could write down an additive model for the observed (log) price process $\{p_t\}$:

$$p_{t_{i,m}} = p^*_{t_{i,m}} + u_{t_{i,m}},$$

and denote p^*_t and u_t are, respectively, signal and noise. This model, however, does not guarantee that one can disentangle the signal or the volatility of the signal. To see this, suppose that the efficient price can be written as

$$dp^*_t = \mu_t dt + \sigma_t dW_t,$$

where the drift coefficient μ_t and the diffusion coefficient σ_t can be random, and W_t is a standard Brownian motion. If one assumed that u_t is also an Itô process, say,

$$du_t = \nu_t dt + \gamma_t dB_t,$$

then p_t is also an Itô process of the form

$$dp_t = (\mu_t + \nu_t)dt + \omega_t dV_t,$$

where $\omega_t^2 = \sigma_t^2 + \gamma_t^2 + 2\sigma_t\gamma_t d < W, B >_t /dt$ (by the Kunita-Watanabe inequality, see, for example, Protter (2004)).

Unless one imposes additional constraints, it is therefore not possible to distinguish signal and noise in this model, and the integrated variance (quadratic variation) of the process should be taken to be $\int_0^T \omega_t^2 dt$. One could, of course, require $\mu_t = 0$, as is done in Aït-Sahalia, Mykland, and Zhang (2005a), but estimability is only possible as $T \to \infty$, and given a parametric model or similar. Note that HL makes the assumption that the drift in the efficient price is zero. This assumption is necessary for their unbiasedness considerations, but for the purposes of asymptotics in a fixed time interval T such as a day, it does not matter. This is for the same reason that a consistent separation of efficient price is not possible in a fixed time interval, so long as the noise is also an Itô process.

The same statement, broadly interpreted (replace integrated volatility of any process X with its quadratic variation), holds true for general semimartingales, see Theorem I.4.47 (p. 52) of Jacod and Shiryaev (2003). One can in some cases extend the concept of quadratic variation to non-semimartingales, such as the process discussed in Section 4.1 of HL, see equation (1) below. As we shall see, even in this case the noise is not separable from the signal except under additional assumptions.

We tentatively conclude, therefore, that the development just after HL's Corollary 3 mostly holds for models where one cannot, in fact, distinguish signal and noise. (Corollary 3 in itself is more general, but does not address identifiability).

What makes this problem particularly difficult is that a substantial fraction of continuous processes of interest here are Itô processes. And many Itô processes have stationary solution. One can easily construct a stationary diffusion process with given marginal distribution and exponential autocorrelation function. By superposition, and by taking limits, one can, for example, construct a Gaussian Itô process with mean zero and with any autocovariance function on the form $\pi(s) = \int_0^\infty e^{-us}\nu(du)$, where ν is any finite measure on $[0, \infty)$.

The only case where one can hope to distinguish between efficient price and noise, is if the noise u_t is not an Itô process. One way for this to occur is if the u_t are independent for different t, and hence the autocovariance function satisfies $\pi(s) = 0$ for $s \neq 0$. This is the model in HL's Section 3.

It should be emphasized that consistency is not guaranteed if the noise is not a semimartingale. This is the case for the noise used in HL's Section 4.1. To see that, assuming that the function ψ is sufficiently smooth, by Itô's formula,

$$u(t) = B(t)\psi(0) - B(t-\theta_0)\psi(\theta_0) + \int_{t-\theta_0}^{t} B(s)\psi'(t-s)ds, \tag{1}$$

and the time-lagged Brownian motion precludes semimartingaleness, see, for example, Definition I.4.21 (p. 43) of Jacod and Shiryaev (2003). Consistency does not follow even in this case. (If one assumed that $\psi(0) = \psi(\theta_0)$, one would be able to estimate this quantity along with the integrated volatility.)

4 Inference in Tick Time

All the difficulties described above go away if one assumes that u_t is independent for different t's, and independent of the efficient price. Since this is an overly restrictive assumption, both HL (in Section 4.2) and ourselves (in Aït-Sahalia, Mykland, and Zhang (2005b)) have fallen back on modeling dependence in tick time. The way this works is that if t_i is tick number i, one assumes that the process u_{t_i} is stationary, and let the autocovariance function be given by $\pi(k) = cov(u_{t_i}, u_{t_{i+k}})$. As discussed both here and in Aït-Sahalia, Mykland, and Zhang (2005b), this provides a scheme for obtaining reasonable estimates of the integrated volatility of the efficient price.

From a conceptual point of view, it also makes sense to sample in tick time, since this permits sampling to adapt to the activity of the market. For transactions, it is also natural to think of the stock price as being a measurement of the efficient price and, as such, subject to measurement error, just like any other measurement.

The situation with bid and ask quotes is more complex. As HL note, taking every adjustment as a tick will result in understating the volatility by a factor of about $1/\sqrt{2}$, and it may be better to take the transaction times as the tick times, even when dealing with quote data. This issue may be part of what leads to falling estimates of volatility at very high frequencies, it should not affect the analysis of transaction data though. This does not, of course, preclude that correlation between signal and noise can exist.

An issue of a perhaps rather academic nature comes up in connection with the modelling of dependency (within the noise, and between noise and signal). This appears to require a triangular array formulation: as ticks become more frequent, so does the dependence occur over ever smaller clock time intervals. This is particularly so for HL's definition of correlation, which is only given as a limit, and it is not quite clear how to set up (even a triangular array) model to achieve such an effect.

Triangular array formulations is a valid and time honored way of asymptotically capturing phenomena which would otherwise be unavailable in the limit. One of the more well known instances of this is the "local to unity" asymptotics in an AR process. But here is a second worry, which pertains to the signal-noise correlation which was found by the authors. The data themselves do not live on a triangular array. The conceptual intent of triangular array asymptotics would seem to preclude its usage on different sampling frequencies in this same dataset. The definition in Theorem 1 requires that the magnitude of $e_{i,n}$ go up as the tick frequency increases. This framework is clearly plausible when used for different frequencies in different datasets, but it is hard to believe that this occurs as one increases the sampling frequency in these given datasets. There is a problem of modelling here which seems to be an open question for researchers to address.

5 Some Comments on the Empirical Analysis

The paper analyzes a vast amount of empirical data, the returns on the 30 DJIA stocks over the five years 2000-2004, including an interesting cointegration analysis. To better understand how different estimators behave, we looked at a small subset of the HL data, which was kindly made available to us: the first stock in the sample (Alcoa, ticker symbol: AA) and one month (January 2004). Clearly, any conclusions drawn from such a limited analysis are to be taken with some caution, but it is certainly possible for the patterns to hold more generally.

Our first concern is the degree of pre-processing of the data that HL performed, prior to any estimator being applied. As the table included here shows, HL on average create a "clean" set of transactions by discarding about half of the starting sample ("raw"). In light of the percentage of the number of transactions being deleted, we think we are no longer talking about a light touch cleaning of the data, but rather major surgery on the data! And, as we will see, this is not without consequences.

We downloaded the raw data for AA in January 2004 directly from the TAQ database. For this particular stock and month, there were no transaction prices reported at 0 and this particular sample appears to be free of major data entry errors. Discarding all price bouncebacks of magnitude 0.5% or greater eliminated fewer than 10 observations each day.[1] So a minimal amount of data cleaning would discard a very tiny percentage of the raw transactions. By contrast, the empirical analysis in HL is conducted after eliminating transactions on the following basis: (i) deleting ex-ante all observations originating from regional exchanges while retaining only those originating on the NYSE; (ii) aggregating all transactions time-stamped to the same second; and (iii) eliminating transactions that appear to have taken place outside a posted bid/ask spread at the same time stamp.

Our view is that those unpleasant data features are precisely the types of things that any estimator that is robust to microstructure noise ought to deal with, without prior intervention: perhaps those transactions time-stamped to the same second did not occur exactly in that order, especially if they took place on different exchanges; perhaps the regional exchanges are less liquid, or less efficient at providing price discovery, but still provide meaningful information as to what the volatility of the price process is for someone contemplating a trade there; perhaps the depth of the quotes is insufficient to fulfill the order, and the transaction has a non-negligible price impact; perhaps the matching of the quotes to the transactions that is used to eliminate transactions (which in their vast majority did actually take place) is not perfect. In the end, this is what market microstructure noise is all about!

Figure 1 examines the consequence of this pre-processing of the data. The left column in the figure shows the results using the "raw" data (after eliminating bouncebacks), while the right column uses the HL "clean" sample. The first row in the figure shows the TSRV estimator, computed for different numbers of subgrids over which our averaging takes place. The second row reports a signature plot for the standard, uncorrected, RV estimator for different number of lags between successive observations (that is, different sparse frequencies). The third row reports the same result as in the second row but now for the autocovariance-corrected estimator used in the HL paper, also as a function of the number of lags used to compute the RV and autocovariance correction. Daily results are averaged over all 20 trading days in January 2004.

[1] We defined in Aït-Sahalia, Mykland, and Zhang (2005b) a "bounceback" as a log-return from one transaction to the next that is both greater in magnitude than an arbitrary cutoff, and is followed immediately by a log-return of the same magnitude but of the opposite sign, so that the price returns to its starting level before that particular transaction.

Trading Day Jan. 2004	Number of Transactions ("RAW")	Transactions Retained ("CLEAN")	% of Sample Discarded
02	4,516	2,855	37%
05	6,495	3,770	42%
06	8,683	3,532	59%
07	6,990	3,575	49%
08	5,661	2,288	60%
09	11,949	3,402	72%
12	9,969	3,528	65%
13	8,551	3,274	62%
14	5,879	3,169	46%
15	6,736	3,798	44%
16	6,853	2,934	57%
20	6,455	3,639	44%
21	5,789	3,352	42%
22	7,029	3,447	51%
23	6,685	3,216	52%
26	5,044	2,753	45%
27	4,857	3,176	35%
28	7,604	3,071	60%
29	7,854	3,289	58%
30	5,521	3,331	40%

Table 2: Transactions on AA Stock, January 2004.

Note: This table reports the number of transactions of AA stock for each trading day in January 2004. The column labeled "RAW" reports the number of transactions that occurred between 9:30am and 4:00pm EST, obtained from the TAQ database. The column labeled "CLEAN" reports the number of transactions retained for the empirical analysis in HL. The last column reports the percentage of transactions from the raw data that are discarded when going to the clean dataset that is used for the empirical analysis in the paper.

Incidentally, the comparable plots in HL report the RV and Zhou estimators also estimated on a daily basis, and then the results averaged over every trading day in 2004. While this time series averaging has the advantage of delivering plots that visually appear to be very smooth, it is not clear to us that this is how such estimators would be used in practice. Our view is that the whole point of using nonparametric measurements of *stochastic* volatility, estimated on a *day-by-day* basis, is that one believes that the quantity of interest can change meaningfully every day, at least for the purposes for which it is to be used (such as adjusting a position hedge). While some averaging is perhaps necessary, computing an average of the day-by-day numbers over an entire year seems to be at odds with the premise of the exercise.

We argue that this pre-processing of the data has multiple consequences: first, it reduces the sample size available for inference, which inevitably increases the variability of the estimators, i.e., decreases the precision with which we can estimate the quadratic variation. This is clear from the figure, comparing the left and right columns for each row, especially for RV and the Zhou/HL estimator.

Second, the “clean” data are smoothed to the point where the estimator analyzed by HL looks in fact very close to the basic uncorrected RV (compare the second and third rows in the right column of the figure.) So in this case at least, the autocovariance correction does not appear to do much. In fact, the signature plot for RV computed from the “clean” data subset exhibits an atypical behavior: as the sampling frequency increases, even to the highest possible level, the value of RV decreases. This is certainly at odds with the empirical evidence available across many different markets, time periods (and papers!)

Third, as a consequence of this decrease in the signature plot, the Zhou/HL estimator taken at lag order 1 seems to underestimate the quadratic variation, with a point estimate (averaged over all 20 days) that is about 0.0002 when the value appears (on the basis of TSRV) to be between 0.00025 and 0.0003.

Fourth, the data cleaning performed by HL may have changed the autocorrelation structure of returns. Figure 2 shows this, by comparing the autocorrelogram from the “raw” and “clean” datasets. The “clean” dataset results in a first order autocorrelation (which is indicative of the inherent i.i.d. component of the noise of the data) that is about a *quarter* of the value obtained from the raw data, while at the same time seeming to introduce spuriously higher positive autocorrelation at orders 3 and above. So, at least for the data we analyze, the pre-processing of the data is far from being inconsequential. The main manifestation of the noise, namely the first order autocorrelation coefficient, has been substantially altered.

To conclude, this (limited) empirical analysis of a (small) subset of the HL data does raise some concerns about the impact of the pre-processing of the data, and the empirical behavior of the estimator used in the HL paper.

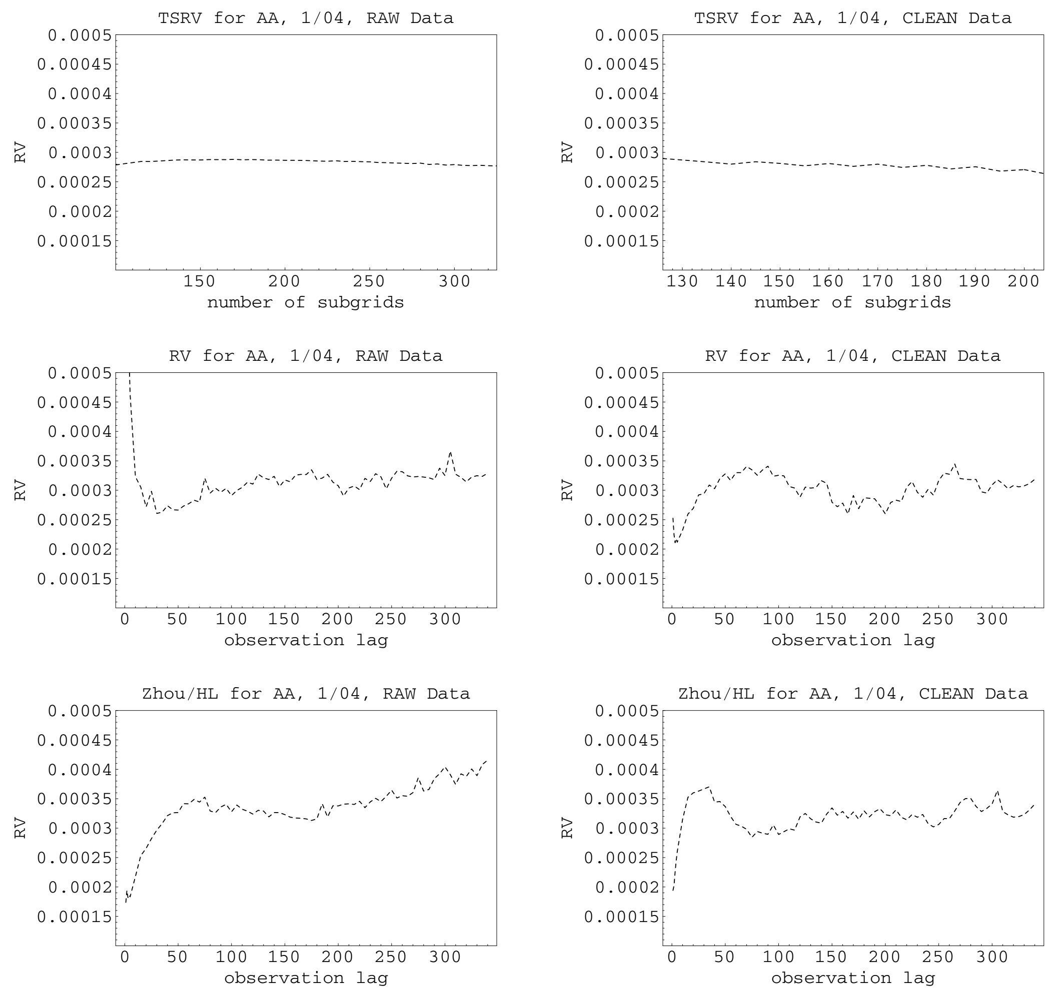

Figure 1: Comparison of Estimators for the AA Transactions, January 2004.

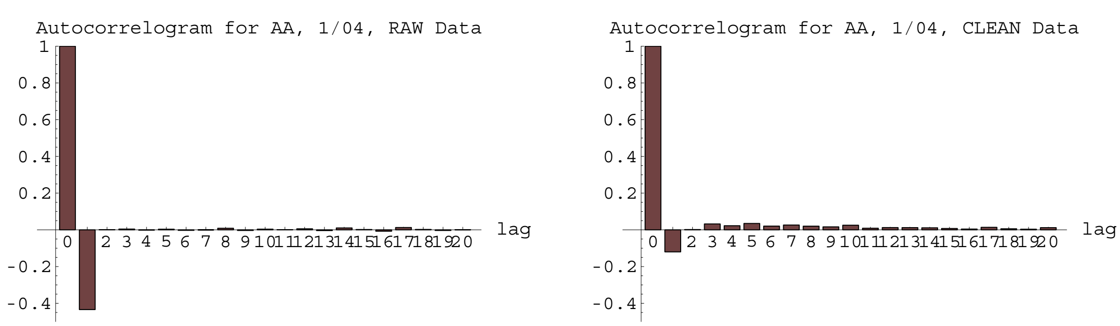

Figure 2: Autocorrelogram of log-returns from the AA Transactions, January 2004.

6 Small Sample Behavior

The paper is silent about the small sample behavior of the estimators considered. This is particularly important in light of the fact, now documented, that RV-type quantities sometimes do not behave exactly as their asymptotic distributions predict. Recently, Goncalves and Meddahi (2005) developed an Edgeworth expansion for the basic RV estimator when there is no noise. Their expansion applies to the studentized statistic based on the standard RV estimator and it is used for assessing the accuracy of the bootstrap in comparison to the first order asymptotic approach. Complementary to that work is the paper Zhang, Mykland, and Aït-Sahalia (2005), where we develop an Edgeworth expansion for nonstudentized statistics for the standard RV, TSRV and other estimators, but allow for the presence of microstructure noise.

It would be interesting to know how the autocovariance-corrected RV estimators behave in small samples and, if need be, investigate the feasibility of Edgeworth corrections for them.

References

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005a), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

— (2005b), "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise," Tech. rep., Princeton University.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

Goncalves, S. and Meddahi, N. (2005), "Bootstrapping realized volatility," Tech. rep., Université de Montréal.

Jacod, J. (1994), "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," Tech. rep., Université de Paris VI.

Jacod, J. and Protter, P. (1998), "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307.

Jacod, J. and Shiryaev, A. N. (2003), *Limit Theorems for Stochastic Processes*, New York: Springer-Verlag, 2nd ed.

Protter, P. (2004), *Stochastic Integration and Differential Equations: A New Approach*, New York: Springer-Verlag, 2nd ed.

Zhang, L. (2001), "From Martingales to ANOVA: Implied and Realized Volatility," Ph.D. thesis, The University of Chicago, Department of Statistics.

— (2004), "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," Tech. rep., Carnegie-Mellon University.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2002), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association, forthcoming*, –, –.

— (2005), "Edgeworth Expansions for Realized Volatility and Related Estimators," Tech. rep., Princeton University.

Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 14, 45–52.

— (1998), "F-consistency, De-volatization and Normalization of High Frequency Financial Data," in *Nonlinear Modelling of High Frequency Financial Time Series*, eds. Dunis, C. L. and Zhou, B., New York: John Wiley Sons Ltd., pp. 109–123.