Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise*

Yacine Aït-Sahalia
Princeton University and NBER

Per A. Mykland
The University of Chicago

Lan Zhang
University of Illinois - Chicago

First Draft: November 2004. This Version: September 2, 2009.

Abstract

We analyze the impact of time series dependence in market microstructure noise on the properties of estimators of the integrated volatility of an asset price based on data sampled at frequencies high enough for that noise to be a dominant consideration. We show that combining two time scales for that purpose will work even when the noise exhibits time series dependence, analyze in that context a refinement of this approach based on multiple time scales, and compare empirically our different estimators to the standard realized volatility.

KEYWORDS: Market microstructure; Serial dependence; High frequency data; Realized volatility; Subsampling; Two Scales Realized Volatility; Multiple Scales Realized Volatility.

*We are grateful to Joel Hasbrouck for comments and help with the sequencing of trades. Financial support from the NSF under grants SBR-0350772 (Aït-Sahalia), DMS-0204639, DMS 06-04758, and SES 06-31605 (Mykland and Zhang) and the NIH under grant RO1 AG023141-01 (Zhang) is also gratefully acknowledged.

1. Introduction

When studying financial data, the notion that noise plays an essential role is an accepted fact of life, whether at the high frequency typical of transactions data or at the lower frequencies more commonly used in asset pricing. That this is a central issue is perhaps best demonstrated by the fact that two recent presidential addresses to the American Finance Association have been entitled "noise" (Black (1986)) and "frictions" (Stoll (2000)) respectively. So we work under the assumption that the observed log-price Y (either transaction or quoted) in high frequency financial data is the unobservable efficient log-price X plus some noise component ϵ due to the imperfections of the trading process,

$$Y_t = X_t + \epsilon_t. \tag{1.1}$$

Since X is defined implicitly (as opposed to explicitly, such as the sum of expected discounted dividends for instance) we have maintained the simple identifying assumption is that ϵ is independent of the X process. It is shown in Li and Mykland (2007) that this assumption can be substantially weakened (see also Jacod (1996) and Delattre and Jacod (1997)).

We are interested in the implications of such a data generating process for the estimation of the volatility of the efficient log-price process

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{1.2}$$

using discretely sampled data on the transaction price process at time intervals of length Δ. By ultra high frequency, we mean that we are in a situation where the data available are such that Δ will be measured in seconds rather than minutes or hours. Under these circumstances, the drift is of course irrelevant, both economically and statistically, and so we shall focus on functionals of the σ_t process and set $\mu_t = 0$. It is the case that transactions and quotes data series in finance are often observed at random time intervals (see Aït-Sahalia and Mykland (2003) for inference under these circumstances) but, throughout this paper, we will assume for simplicity that Δ is nonrandom when studying the asymptotic properties of our estimators. We make essentially no assumptions on the σ_t process: its driving process can of course be correlated with the Brownian motion W_t in (1.2), and it need not even have continuous sample paths.

The noise term ϵ summarizes a diverse array of market microstructure effects, which can be roughly divided into three groups. First, ϵ represents the frictions inherent in the trading process: bid-ask bounces, discreteness of price changes and rounding, trades occurring on different markets or networks, etc. Second, ϵ captures informational effects: differences in trade sizes or informational content of price changes, gradual response of prices to a block trade, the strategic component of the order flow, inventory control effects, etc. Third, ϵ encompasses measurement or data recording errors such as prices entered as zero, misplaced decimal points, etc., which are surprisingly prevalent in these types of data. As is clear from the laundry list of potential sources of noise, the data generating process for ϵ is likely to be quite involved. Therefore, robustness to departures from any assumptions on ϵ is desirable.

Models akin to (1.1) have been studied in the constant σ case by Zhou (1996), who proposes a bias correcting approach based on autocovariances. The behavior of this estimator has been studied by Zumbach et al. (2002). Hansen and Lunde (2006) study the Zhou estimator and extensions in the case where volatility is time varying but conditionally nonrandom. Related contributions have been made by Oomen (2006) and Bandi and Russell (2008).

If σ_t is modelled parametrically, as a constant, we showed in Aït-Sahalia et al. (2005) that incorporating ϵ explicitly in the likelihood of the observed log-returns Y provides consistent and asymptotically normal estimators of the parameters. But what distributional assumption to use for ϵ? Surprisingly, we found that misspecifying the marginal distribution of ϵ has no adverse consequences.

In the nonparametric case where σ_t is an unrestricted stochastic process, an important object of interest is the integrated volatility or quadratic variation of the process, $\langle X, X\rangle_T = \int_0^T \sigma_t^2 dt$, over a fixed interval T, typically one day in empirical applications. This quantity can then be used to hedge a derivatives' portfolio, forecast the next day's integrated volatility, etc. Without noise, the realized volatility (RV) estimator $[Y,Y]_T^{(\mathrm{all})} = \sum_{i=1}^n (Y_{t_{i+1}} - Y_{t_i})^2$ provides an estimate of the quantity $\langle X, X\rangle_T$, and asymptotic theory would lead one to sample as often as possible, or use all the data available, hence the "all" superscript. The sum $[Y,Y]_T^{(\mathrm{all})}$ converges to the integral $\langle X, X\rangle_T$, with a known distribution, a result which dates back to Jacod (1994) and Jacod and Protter (1998); see also e.g., Barndorff-Nielsen and Shephard (2002) and Mykland and Zhang (2006).

In Aït-Sahalia et al. (2005) and Zhang et al. (2005), we studied the corresponding problem when a relatively simple type of market microstructure noise, iid, is present. We showed there that the situation changes radically in the presence of market microstructure noise. In particular, computing RV using all the data available (say every second) leads to an estimate of the variance of the noise, not the quadratic variation that one seeks to estimate: $[Y,Y]_T^{(\mathrm{all})}$ has bias $2nE[\epsilon^2]$, which is an order of magnitude larger than the object we seek to estimate, $\langle X, X\rangle_T$. The divergence of the RV estimator as the number of observations n increases is illustrated in Figure 1, which shows the behavior of the RV estimator as a function of the sampling interval $\Delta = T/n$: as predicted by our theory, the plot shows divergence proportional to $1/n$.

Equivalently, since our theory predicts that $RV \approx 2nE\left[\epsilon^2\right]$ asymptotically in n, we expect that $\ln RV \approx \ln(2E\left[\epsilon^2\right]) + \ln n$ so that a regression of $\ln RV$ on $\ln n$ should have slope coefficient close to 1. Figure 2 shows the result: the estimated slope coefficient is 1.02 and the null value of 1 is not rejected. In theory, an estimate of $E[\epsilon^2]$ can be constructed using the intercept in that regression. In practice, the quality of the estimates derived from that regression could be adversely affected by the endogeneity of the regressor, cf. the data analysis in Hansen and Lunde (2006) and the theoretical development in Li and Mykland (2007). This difficulty seems to be of some importance for quote data, while transaction data seem more robust in this respect.

While a formal analysis of this phenomenon originated in our work cited above, the empirical message that emerges from this has long been known: do not compute RV at too high a frequency. This in fact formed the rationale for the recommendation in the literature to sample sparsely at some lower frequency. A sampling interval Δ_{sparse} is picked in the range from 5 to 30 minutes: see e.g., Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002) and Gençay et al. (2002). We denote the RV estimator corresponding to $\Delta_{\mathrm{sparse}} = T/n_{\mathrm{sparse}}$ as $[Y,Y]_T^{(\mathrm{sparse})}$.

If one insists upon sampling sparsely, we then showed in our earlier papers how to determine the optimal sparse frequency, instead of selecting it arbitrarily. But even if sampling sparsely at our optimally-determined frequency, one is still throwing away a large amount of data. For example, if $T = 1$ NYSE day and transactions occur every $\Delta = 1$ second, the original sample size is $n = T/\Delta = 23{,}400$. Sampling sparsely even at the highest frequency used by empirical researchers (once every 5 minutes) entails throwing away 299 out of every 300 observations: the sample size used is only $n_{\mathrm{sparse}} = 78$. This violates one of the most basic principles of statistics, and our objective when starting this research project was to propose a solution which made use of the full data sample, despite the fact that ultra high frequency data can be extremely noisy. [1]

Our approach to estimating the volatility is to use Two Scales Realized Volatility (TSRV). By evaluating the quadratic variation at two different frequencies, averaging the results over the entire sampling, and taking a suitable linear combination of the result at the two frequencies, one obtains a consistent and asymptotically unbiased estimator of $\langle X, X\rangle_T$. We start by briefly reviewing the rationale behind the TSRV estimator in Section 2.

In our earlier paper, however, we made the assumption that the noise term was iid. In Section 3, we document that dependence in the noise can be important in some empirical situations. So our main purpose in the following will be to propose a version of the TSRV estimator which can deal with such serial dependence in market microstructure noise. Both Aït-Sahalia et al. (2005) and Hansen and Lunde (2006) have considered such departures in the case of the MLE for σ constant and Zhou estimator for σ time varying, respectively.[2] [3]

Just like the marginal distribution of the noise is likely to be unknown, its degree of dependence is also likely to be unknown and so our approach will be nonparametric in nature. We develop the theory for a generalized, serial-dependence-robust, TSRV estimator in Section 4. In a nutshell, we will continue combining two different time scales, but rather than starting with the fastest possible time scale as our starting point, one now needs to be somewhat more subtle and adjust how fast the fast time scale is. Next, we analyze in Section 5 the impact of serial dependence in the noise on the distribution of the RV estimators, $[Y,Y]_T^{(\mathrm{all})}$ and $[Y,Y]_T^{(\mathrm{sparse})}$. We then discuss in Section 6 the Multiple Scales Realized Volatility (MSRV, Zhang (2006)), which achieves further asymptotic efficiency gains over TSRV. We did for TSRV and RV, we analyze the impact of serial dependence in the noise on that estimator, and see that this estimator does not need to be modified because of the dependence. Finally, we provide in Section 7 an empirical study of the TSRV and MSRV estimators, and compare them to RV. We examine in particular the robustness of TSRV to the choice of the two time scales, contrast it with RV's divergence as sampling gets more frequent and with RV's variability in empirical samples, and study the dependence of the estimators on various ways of pre-processing the raw high frequency data. Section 8 concludes.

2. The TSRV Estimator with IID Noise

Before showing how to extend TSRV to account for serial dependence in market microstructure noise, we first summarize the properties of TSRV under iid noise so that we can later on discuss the effect of dependence in

[1] Since the best achievable convergence rate under microstructure is of order $O_p(n^{-1/4})$ rather than the standard $O_p(n^{-1/2})$, the loss in keeping only one out of 300 observations is comparable to, in a standard statistical situation, keeping only 1 in $\sqrt{300} \approx 17$ data. The loss to subsampling is thus not quite as severe as it may at first seem, though is also most small. The preceding numbers ignore, of course, all asymptotic constants.

[2] We exclude here any form of correlation between the noise ε and the efficient price X in our analysis, something which has been stressed as potentially important by Hansen and Lunde (2006). As we discuss in Aït-Sahalia et al. (2006), however, 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, is not enough to make the noise identifiable.

[3] Another issue we do not address in the present paper is that of small sample corections to the asymptotics of the estimators. Recently, Gonçalves and Meddahi (2009) have 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 and it is used for assessing the accuracy of the bootstrap in comparison to the first order asymptotic approach. By contrast, we develop in Zhang et al. (2009) an Edgeworth expansion for nonstudentized statistics for the standard RV, TSRV and other estimators, but allow for the presence of microstructure noise.

the noise. The TSRV estimator is based on subsampling, averaging and bias-correction. The idea is to partition the original grid of observation times, $\mathcal{G} = \{t_0, ..., t_n\}$ into subsamples, $\mathcal{G}^{(k)}$, $k = 1, ..., K$ where $n/K \to \infty$ as $n \to \infty$. For example, for $\mathcal{G}^{(1)}$ start at the first observation and take an observation every 5 minutes; for $\mathcal{G}^{(2)}$, start at the second observation and take an observation every 5 minutes, etc. Then we average the estimators obtained on the subsamples. The idea is that the benefit of sampling sparsely, as in $[Y,Y]_T^{(\text{sparse})}$ vs. $[Y,Y]_T^{(\text{all})}$, can now be retained, while the variation of the estimator can be lessened by the averaging and the use of the full data sample.

Subsampling and averaging together gives rise to the estimator

$$[Y,Y]_T^{(\text{avg})} = \frac{1}{K}\sum_{k=1}^{K}[Y,Y]_T^{(\text{sparse},k)} \tag{2.1}$$

constructed by averaging the estimators $[Y,Y]_T^{(\text{sparse},k)}$ obtained by sampling sparsely on each of the K grids of average size $\bar{n} = n/K$.

Unfortunately, $[Y,Y]_T^{(\text{avg})}$ remains a biased estimator of the quadratic variation $\langle X, X\rangle_T$ of the true return process, although its bias $2\bar{n}E[\epsilon^2]$ now increases with the average size $\bar{n}$ of the subsamples, instead of the full sample size n as in $2nE[\epsilon^2]$. But $E[\epsilon^2]$ can be consistently approximated by $[Y,Y]_T^{(\text{all})}$:

$$\widehat{E[\epsilon^2]} = \frac{1}{2n}[Y,Y]_T^{(\text{all})}. \tag{2.2}$$

Thus a bias-adjusted estimator for $\langle X, X\rangle_T$ can be constructed as

$$\widehat{\langle X, X\rangle}_T^{(\text{tsrv})} = \underbrace{[Y,Y]_T^{(\text{avg})}}_{\text{slow time scale}} - \frac{\bar{n}}{n}\underbrace{[Y,Y]_T^{(\text{all})}}_{\text{fast time scale}} \tag{2.3}$$

and this is the TSRV estimator. Figure 3 summarizes this construction.

If the number of subsamples is optimally selected as $K^* = cn^{2/3}$, then TSRV has the following distribution:

$$\widehat{\langle X, X\rangle}_T^{(\text{tsrv})} \overset{\mathcal{L}}{\approx} \underbrace{\langle X, X\rangle_T}_{\text{object of interest}} + \frac{1}{n^{1/6}}\underbrace{[\underbrace{\frac{8}{c^2}E[\epsilon^2]^2}_{\text{due to noise}} + \underbrace{c\frac{4T}{3}\int_0^T \sigma_t^4 dt}_{\text{due to discretization}}]^{1/2}}_{\text{total variance}} Z_{\text{total}} \tag{2.4}$$

and the constant c can be set to minimize the total asymptotic variance above.

Unlike all the previously considered ones, this estimator is now correctly centered, and to the best of our knowledge is the first consistent estimator for $\langle X, X\rangle_T$ in the empirically relevant case where market microstructure noise is present and the volatility is non-constant. A small sample refinement to $\widehat{\langle X, X\rangle}_T$ can be constructed as follows

$$\widehat{\langle X, X\rangle}_T^{(\text{tsrv,adj})} = \left(1 - \frac{\bar{n}}{n}\right)^{-1}\widehat{\langle X, X\rangle}_T^{(\text{tsrv})}. \tag{2.5}$$

The difference from the estimator in (2.3) is of order $O_p(\bar{n}/n) = O_p(K^{-1})$, and thus the two estimators behave identically to the asymptotic order that we consider. The estimator (2.5), however, has the appeal of being unbiased to higher order.

If two scales are better than one, how about using three or more? This question has been studied in Zhang (2006) where it is shown that one can further improve efficiency by taking a weighted average of $[Y,Y]_T^{(\text{avg})}$ for multiple time scales. The resulting estimator, the MSRV, has rate of convergence $n^{-1/4}$, and is thus an improvement over the TSRV's rate of $n^{-1/6}$. This is the best possible rate even in the parametric case (when $\sigma_t = \sigma$, a constant, and the noise is normal), as established in Gloter and Jacod (2000). MSRV is further discussed in Section 6.

Following these papers, Barndorff-Nielsen et al. (2008) have shown that the TSRV and MSRV estimator are closely related to their "realized kernel" class of estimators based on autocovariances. They are also closely related to the "preaveraging" class of estimators studied by Jacod et al. (2009) and Podolskij and Vetter (2009). The three types of estimators (subsampling, realized kernel, and preaveraging) differ in their treatment of end effects.

Another possible derivation and generalization of our TSRV estimator is provided by Curci and Corsi (2005). They propose realized volatility measures based on a Discrete Sine Transform (DST) of high frequency returns data. The DST can diagonalize MA processes by providing an orthonormal basis decomposition of observed returns to disentangle the volatility signal of the underlying price process from the market microstructure noise. This approach delivers an estimator close to TSRV that combines two RV estimators and a new multi-frequency estimator that, like MSRV, combines multiple RV estimators.

3. Time Series Dependence in High Frequency Market Microstructure Noise

We now turn to examining empirically whether there is a need to relax the assumption that the market microstructure noise ϵ is iid. In other words, is it the case that every time a new price is observed, one observes it with an error that is independent of the previous one, no matter how close together those two successive prices might be?

3.1. The Data

Our data consist of transactions and quotes from the NYSE's TAQ database for the 30 Dow Jones Industrials Average (DJIA) stocks, over the last ten trading days of April 2004 (April 19-23 and 26-30). To save space, we will focus on four of the thirty stocks: 3M Inc. (trading symbol: MMM), American International Group (trading symbol: AIG), Intel (trading symbol: INTC) and Microsoft (trading symbol: MSFT). Of these, the first two are traded on the NYSE while the latter two are traded on the Nasdaq. Table 1 reports the basic summary statistics on these four stocks' transactions.

In our earlier paper where we introduced the TSRV estimator, we assumed that microstructure noise ϵ was iid. In that case, log-returns

$$Y_{\tau_i} - Y_{\tau_{i-1}} = \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t + \epsilon_{\tau_i} - \epsilon_{\tau_{i-1}} \tag{3.1}$$

follow an MA(1) process since the increments $\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$ are uncorrelated, $\epsilon \perp\!\!\!\perp W$ and therefore, in the simple case where σ_t is nonrandom (but possibly time varying),

$$E\left[\left(Y_{\tau_j} - Y_{\tau_{j-1}}\right)\left(Y_{\tau_i} - Y_{\tau_{i-1}}\right)\right] = \begin{cases} \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt + 2E\left[\epsilon^2\right] & \text{if } j = i \\ -E\left[\epsilon^2\right] & \text{if } j = i+1 \\ 0 & \text{if } j > i+1 \end{cases} \tag{3.2}$$

Under the simple iid noise assumption, log-returns are therefore (negatively) autocorrelated at the first order. We will examine below whether this is compatible with what we observe in the data, but for now note that this is consistent with the predictions of many simple reduced form market microstructure models. For instance, in the Roll (1984) model, $\epsilon_t = (s/2)Q_t$ where s is the bid/ask spread and Q_t, the order flow indicator, is a binomial variable that takes the values $+1$ and -1 with equal probability, generating first order autocorrelation in returns. French and Roll (1986) proposed to adjust variance estimates to control for such autocorrelation and Harris (1990) studied the resulting estimators. Zhou (1996) proposed a bias correcting approach based on the first order autocovariances; see also Hansen and Lunde (2006) who study the Zhou estimator.

We now turn to confronting this model to the data. The top panel of Figure 4 reports the autocorrelogram computed for the 3M and AIG transactions, respectively. That part of the plot shows a good agreement with the prediction of the iid noise model, namely the MA(1) structure in (3.2) for 3M and AIG.

However, Figure the middle panel of the same Figure 4 shows the corresponding result for Intel and Microsoft. It is clear that the MA(1) model, and consequently the iid noise model, does not fit those data well for these two stocks. Both stocks were added to the DJIA on November 1, 1999, becoming the first two companies traded on the Nasdaq to be included in the DJIA.

It is important to note however, that the difference between the two figures does not appear to be driven by the different market structures on the NYSE (a specialist market structure) compared to the Nasdaq (a dealers' market). In fact, the autocorrelogram pattern for the other 26 DJIA stocks is closer to that of Intel and Microsoft, not that of 3M and AIG. Table 2 reports the results of a cross-sectional OLS regressions of the autocorrelation coefficients of order 2-5 on the average time between transactions used as a measure of the liquidity of the stock, for the 30 DJIA stocks. These autocorrelation coefficients of order greater than 1 would be zero if the noise term were serially uncorrelated, as in (3.2). The table shows that the lower the time between successive transactions, the higher the observed autocorrelation in absolute value (the coefficients alternate signs because the autocorrelation coefficients do, as in the middle panel of Figure 4). In other words, based on these data, the more liquid the stock, the more likely we are to face departures from the iid assumption.

Another angle to understand the departure from the MA(1) autocorrelogram in the direction of an ARMA(1,1) is through the use of different sampling schemes. Griffin and Oomen (2008) provide an interesting analysis of the impact of tick vs. transaction sampling. Their results show that the nature of the sampling mechanism can generate fairly distinct autocorrelogram patterns for the resulting log-returns. Now, from a practical perspective, we can view the choice of sampling scheme as one more source of noise, this one attributable to the econometrician who is deciding between different ways to approach the same original transactions or quotes data: should we sample in calendar time? transaction time? tick time? something else altogether? Since the sampling mechanism is not dictated by the data, this argues for working under robust departures from the basic assumptions.

3.2. Example: A Simple Model to Capture the Noise Dependence

A simple model to capture the higher order dependence that we just documented in INTC and MSFT trades is

$$\epsilon_{t_i} = U_{t_i} + V_{t_i} \tag{3.3}$$

where U is iid, V is $AR(1)$ with first order coefficient ρ, $|\rho| < 1$, and $U \perp V$. Under this model, we have

$$E\left[\left(Y_{\tau_j} - Y_{\tau_{j-1}}\right)\left(Y_{\tau_i} - Y_{\tau_{i-1}}\right)\right] = \begin{cases} \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt + 2E\left[U^2\right] + 2\left(1-\rho\right)E\left[V^2\right] & \text{if } j = i \\ -E\left[U^2\right] - \left(1-\rho\right)^2 E\left[V^2\right] & \text{if } j = i+1 \\ -\rho^{j-i-1}\left(1-\rho\right)^2 E\left[V^2\right] & \text{if } j > i+1 \end{cases} \tag{3.4}$$

This model can easily be fitted to the data by the generalized method of moments. We use the first twenty autocovariances of the log-returns as moment functions, in order to estimate the three parameters $E[U^2]$, $E[V^2]$ and ρ. Their estimated values are $4.2\ 10^{-8}$, $3.5\ 10^{-8}$ and -0.68 for INTC and $2.9\ 10^{-8}$, $4.3\ 10^{-8}$ and -0.70 for MSFT. The bottom panel of Figure 4 shows the sample autocorrelogram and the corresponding one fitted by the model above, illustrating the generally good fit produced by this simple model.

Let us stress, however, that, while this simple model seems to capture fairly well the dependence in the stock data that we have examined, our theory is not tied to this particular specification of ϵ. It applies to fairly general dependence structures, as can be seen from Assumption 1 below. Finally, note also that while for consistency reasons the bottom panel of Figure 4 reports autocorrelations, the fitting is actually done on autocovariances as given in (3.4).

3.3. *Transactions or Quotes?*

The model (3.3) for the microstructure noise describes well a situation where the primary source of the noise beyond order one consists of further bid-ask bounces. In such a situation, the fact that a transaction is on the bid or ask side has little predictive power for the next transaction, or at least not enough to predict that two successive transactions are on the same side with very high probability (although Choi et al. (1988) have argued that serial correlation in the transaction type can be a component of the bid-ask spread, and extended the model of Roll (1984) to allow for it).

Figure 4 and the estimates just reported ($\rho = -0.7$) are evidence of negative autocorrelation at horizons of up to about 15 transactions. In trying to assess the source of the higher order dependence in the log-returns, a natural hypothesis is that this is due to the trade reversals: in transactions data and an orderly liquid market, one might expect that in most cases successive transactions of the same sign (buy or sell orders) will not move the price. The next recorded price move is then, more likely than not, going to be caused by a transaction that occurs on the other side of the bid-ask spread, and so we observed these reversals when the data consist of the transactions that lead to a price change.

To examine this hypothesis, we turn to quotes data, also from the TAQ database. The results are reported in Figure 5 and suggest that an important source for the AR(1) pattern with negative autocorrelation (the term V in (3.3)) will be trade reversals. The remaining autocorrelation exhibited in the quotes data can also be captured by model (3.3), but with a positive autocorrelation in the V term. This can capture effects such as the gradual adjustment of prices in response to a shock such as a large trade.

4. Extending the TSRV Estimator for Dependent Noise

In the previous section, we found that there are empirical situations (such as Intel or Microsoft transactions) where the assumption of iid market microstructure noise could be problematic. We now proceed to suitably extending the TSRV estimator to make it robust to departures from the iid noise assumption. The idea is to somewhat slow the fast time scale to reduce the degree of dependence that is induced by the noise.

4.1. The Setup

As above, we let Y be the logarithm of the transaction price, which is observed at times $0 = t_0, t_1, \ldots, t_n = T$. We assume that at these times, Y is related to a latent true price X (also in logarithmic scale) through equation (1.1). The latent price X is given by (1.2).

Assumption 1. *We assume that the noise process ϵ_{t_i} is independent of the X_t process, and that it is (when viewed as a process in index i) stationary and strong mixing with the mixing coefficients decaying exponentially. We also suppose that for some $\kappa > 0$, $E\epsilon^{4+\kappa} < \infty$.*

Definitions of mixing concepts can be found e.g., in Hall and Heyde (1980), p. 132. Note that by Theorem A.6 (p. 278) of Hall and Heyde (1980), there is a constant $\rho < 1$ so that, for all i,

$$\left|\mathrm{Cov}(\epsilon_{t_i}, \epsilon_{t_{i+l}})\right| \quad \leq \quad \rho^l \,\mathrm{Var}(\epsilon) \tag{4.1}$$

For the moment, we focus on determining the integrated volatility of X for one time period $[0, T]$. This is also known as the continuous quadratic variation $\langle X, X\rangle$ of X. In other words,

$$\langle X, X\rangle_T = \int_0^T \sigma_t^2 dt. \tag{4.2}$$

Our volatility estimators can be described by considering subsamples of the total set of observations. A realized volatility based on every j'th observation, and starting with observation number r, is given as

$$[Y, Y]_T^{(j,r)} = \sum_{0 \leq j(i-1) \leq n-r-j} (Y_{t_{ji+r}} - Y_{t_{j(i-1)+r}})^2.$$

Under most assumptions, this estimator violates the sufficiency principle, whence we define the *average lag j realized volatility* as

$$[Y, Y]_T^{(J)} = \frac{1}{J}\sum_{r=0}^{J-1}[Y, Y]_T^{(J,r)} = \frac{1}{J}\sum_{i=0}^{n-J}(Y_{t_{i+J}} - Y_{t_i})^2. \tag{4.3}$$

A generalization of TSRV can be defined for $1 \leq J < K \leq n$ as

$$\widehat{\langle X, X\rangle}_T^{(\mathrm{tsrv})} = \underbrace{[Y, Y]_T^{(K)}}_{\text{slow time scale}} - \frac{\bar{n}_K}{\bar{n}_J}\underbrace{[Y, Y]_T^{(J)}}_{\text{fast time scale}}, \tag{4.4}$$

thereby combining the two time scales J and K. Here $\bar{n}_K = (n - K + 1)/K$ and similarly for $\bar{n}_J$.

We will continue to call this estimator the TSRV estimator, noting that the estimator we proposed in Zhang et al. (2005) is the special case where $J = 1$ and $K \to \infty$ as $n \to \infty$. The original TSRV produces a consistent estimator in the case where the ϵ_{t_i} are iid. For the optimal choice $K = O(n^{2/3})$,

$$\widehat{\langle X, X\rangle}_T^{(\mathrm{tsrv})} - \langle X, X\rangle_T = O_p(n^{-1/6}).$$

The problem with which we are concerned here is that these assumptions on the noise ϵ_{t_i} may be too restrictive. We shall see that in the case where J is allowed to be larger than 1, the problem of dependence of the ϵ_{t_i}'s will be eliminated and the generalized TSRV estimator given in (4.4) will be consistent for suitable choices of (J, K).

4.2. A Signal-Noise Decomposition

We have the following.

Lemma 1. *Under the assumptions above, let $n \to \infty$, and let $j = j_n$ be any sequence. Then*

$$\sum_{i=0}^{n-j} (X_{t_{i+j}} - X_{t_i})(\epsilon_{t_{i+j}} - \epsilon_{t_i}) = O_p(j^{1/2}).$$

The lemma is important because it gives rise to the sum of squares decomposition

$$[Y,Y]_T^{(J)} = [X,X]_T^{(J)} + [\epsilon,\epsilon]_T^{(J)} + O_p(J^{-1/2}).$$

Thus, if we look at linear combinations of the form (4.4), one obtains

$$\widehat{\langle X,X \rangle}_T^{(\mathrm{tsrv})} = \underbrace{[X,X]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}[X,X]_T^{(J)}}_{\text{signal term}} + \underbrace{[\epsilon,\epsilon]_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}[\epsilon,\epsilon]_T^{(J)}}_{\text{noise term}} + O_p(K^{-1/2}),$$

so long as

$$1 \le J \le K \quad \text{and} \quad K = o(n), \tag{4.5}$$

both of which will be assumed throughout.

4.3. Analysis of the Noise Term

It can be seen that when the ϵ's are independent $E[\text{noise term}] = 0$, so that the linear combination used in (4.4) is exactly what is needed to remove the bias due to noise. To analyze the more general case, and to obtain approximate distribution of the noise term, note that

$$[\epsilon,\epsilon]_T^{(J)} = \frac{1}{J}\sum_{i=0}^{n} c_i^{(J)} \epsilon_{t_i}^2 - \frac{2}{J}\sum_{i=0}^{n-J} \epsilon_{t_i}\epsilon_{t_{i+J}}, \tag{4.6}$$

where $c_i^{(J)} = 2$ for $J \le i \le n-J$, and $= 1$ for other i. By construction

$$\sum_i c_i^{(J)} = 2J\bar{n}_J, \tag{4.7}$$

so that for $J \le n/2$,

$$\begin{aligned} |E[\text{noise term}]| &\le 2E\epsilon^2 \left(\frac{1}{K}(n-K+1)\rho^K + \frac{\bar{n}_K}{\bar{n}_J}\frac{1}{J}(n-J+1)\rho^J \right) \\ &= O\left(\frac{n}{K}(\rho^K + \rho^J) \right) \end{aligned} \tag{4.8}$$

and, in the regular case where $\mathrm{Cov}(\epsilon_{t_0}, \epsilon_{t_K}) = o(\mathrm{Cov}(\epsilon_{t_0}, \epsilon_{t_J}))$

$$E\,[\text{noise term}] = 2\frac{n}{K}\,\mathrm{Cov}(\epsilon_{t_0}, \epsilon_{t_J})(1 + o(1)).$$

If $J \to \infty$ at even a quite slow rate when $n \to \infty$, the bias is negligible. Also, in the case of m-dependent ϵs, the bias becomes zero for finite J. We obtain:

Proposition 1. *Under assumption (4.12) below,*

$$\frac{K}{n^{1/2}}\left(\textit{noise term} - E\left[\textit{noise term}\right]\right) \xrightarrow{\mathcal{L}} \xi Z_{noise} \tag{4.9}$$

as $n \to \infty$, where Z_{noise} is standard normal. Further, in the case where both J and K go to infinity with n, we have that $\xi^2 = \xi^2_\infty$, where

$$\xi^2_\infty = 8\operatorname{Var}(\epsilon)^2 + 16\sum_{i=1}^{\infty}\operatorname{Cov}(\epsilon_{t_0}, \epsilon_{t_i})^2. \tag{4.10}$$

In the case where J does not go to infinity (the m-dependent case, say), then

$$\xi^2 = \xi^2_\infty + 4\alpha_0 + 8\sum_{i=1}^{\infty}\alpha_i \tag{4.11}$$

where

$$\alpha_i = \operatorname{Cov}(\epsilon_{t_0}, \epsilon_{t_{i+J}})\operatorname{Cov}(\epsilon_{t_i}, \epsilon_{t_J}) + \operatorname{Cum}(\epsilon_{t_0}, \epsilon_{t_i}, \epsilon_{t_J}, \epsilon_{t_{i+J}}).$$

Note that even when $J \to \infty$, one may be better off using (4.11) than ξ^2_∞ since the former is closer to the small sample variance, and since $J \to \infty$ quite slowly. (By contrast, $K \to \infty$ much more quickly, as we shall see).

4.4. Analysis of the Signal Term

As for the "signal term", we obtain that $[X, X]^{(K)}_T \to \langle X, X\rangle_T$ in probability as $n \to \infty$, provided $K = o(n)$. Obviously, for the signal term in $\widehat{\langle X, X\rangle}_T - \langle X, X\rangle_T$ to be scalable to be consistent (see equation (4.16) below), we need

$$\limsup_{n\to\infty}\frac{J}{K} \;<\; 1, \tag{4.12}$$

which is easily satisfied. In fact, as we shall see, one would normally take

$$\limsup_{n\to\infty}\frac{J}{K} \;=\; 0. \tag{4.13}$$

Specifically, we have in the following, which is proved in the same way as Theorem 2 (p. 1401) of Zhang et al. (2005):

Proposition 2. *Under (4.5),*

$$\left(\frac{K}{n}\left(1 + 2\frac{J^3}{K^3}\right)\right)^{-1/2}\left([X, X]^{(K)}_T - \frac{\bar{n}_K}{\bar{n}_J}[X, X]^{(J)}_T - \langle X, X\rangle_T\right) \xrightarrow{\mathcal{L}} \eta\sqrt{T}Z_{discrete}, \tag{4.14}$$

where, $Z_{discrete}$ is standard normal, and where in general, η^2 is given as the limit in Theorem 3 in Zhang et al. (2005) (i.e., the discretization variance η^2 has the same expression as when the noise is iid). In the special case where observations are equidistant,

$$\eta^2 = \frac{4}{3}\int_0^T \sigma_t^4\,dt. \tag{4.15}$$

The convergence in law is stable (see Chapter 3 of Hall and Heyde (1980)), the most important consequence of which is that Z_{discrete} is independent of η.

4.5. The Combined Estimator

Consider the adjusted estimator

$$\widehat{\langle X,X\rangle}_T^{(\text{tsrv,adj})} = \left(1-\frac{\bar{n}_K}{\bar{n}_J}\right)^{-1}\widehat{\langle X,X\rangle}_T^{(\text{tsrv})} \tag{4.16}$$

In the iid case of Zhang et al. (2005), this adjustment was introduced from small sample considerations (Section 4.2). Here, we also see that in the case where (4.12) is satisfied but (4.13) is not, this adjustment is needed for consistency. In the following, we analyze this estimator, and for the case when (4.13) holds, the same analysis applies to the original $\widehat{\langle X,X\rangle}_T^{(\text{tsrv})}$.

We obtain from equations (4.9) and (4.14) that

$$\begin{aligned}\widehat{\langle X,X\rangle}_T^{(\text{tsrv,adj})} &= \left(2E\epsilon^2\frac{n}{K}\operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_J})+\frac{n^{1/2}}{K}\xi Z_{\text{noise}}+\left(\frac{K}{n}\left(1+2\frac{J^3}{K^3}\right)\right)^{1/2}\eta\sqrt{T}Z_{\text{discrete}}\right)\\ &\quad\times\left(1-\frac{\bar{n}_K}{\bar{n}_J}\right)^{-1}(1+o_p(1)),\end{aligned} \tag{4.17}$$

where Z_{noise} and Z_{discrete} are asymptotically standard normal, and asymptotically independent.

It is easy to see that the optimal trade-off between the two variance terms results in a choice of $K=O(n^{2/3})$. The worst thing that can then happen to the bias term is then that this is of the order of $(n/K)\rho^J=n^{1/3}\rho^J$. Thus the bias is of small order relative to the variance provided one chooses $n^{1/3}\rho^J=o(n^{-1/6})$, *i.e.*, $\rho^J=o(n^{-1/2})$. Thus, one can safely assume that $J/K\sim 0$ (i.e., (4.13)), and it follows that (from the two previous propositions, the interaction term going away as in Lemma A.2 (p. 1408) of Zhang et al. (2005)):

Proposition 3. *The asymptotic behavior of the estimator $\widehat{\langle X,X\rangle}_T^{(tsrv,adj)}$ is given by*

$$\widehat{\langle X,X\rangle}_T^{(tsrv,adj)} = \langle X,X\rangle_T+\left(2\left(E\epsilon^2\right)\frac{n}{K}\operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_J})+\frac{n^{1/2}}{K}\xi Z_{noise}+\left(\frac{K}{n}\right)^{1/2}\eta\sqrt{T}Z_{discrete}\right)(1+o_p(1)). \tag{4.18}$$

Similar results have been developed for the multivariate asynchronous case in Zhang (2009). − The optimal K is as given above, and one chooses, ultimately, J so that

$$\operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_J})=o(n^{-1/2}). \tag{4.19}$$

Obviously, when ϵ is m-dependent, one can simply choose $J=m+1$. In terms of asymptotic variance, there is no unique optimal J. In order to minimize asymptiotic variance, J should follow, (4.13). Hence, so long as J falls into the range going from the lower bound defined by (4.19) and the upper bound defined by (4.13). For further optimization of J, either small sample or asymptotic expansion arguments would need to be invoked, and this is beyond the scope of this paper.

4.6. A Further Adjustment to the TSRV Estimator

It should be noted that when K is large, $[X,X]_T^{(K)}$ may be a slight underestimate of $\langle X,X\rangle_T$. To consider the issue, if σ_t^2 is constant, $\sigma_t^2=\sigma^2$, one gets that $\langle X,X\rangle_T=\sigma^2T$, whereas $[X,X]_T^{(K)}\approx\sigma^2T(n-K+1)/n$ (the approximation here is loose, but, for example, it is an equality in expectation when σ_t^2 is constant). Thus

$$[X,X]_T^{(K)}-\frac{\bar{n}_K}{\bar{n}_J}[X,X]_T^{(J)}\approx\sigma^2T\left(\frac{n-K+1}{n}-\frac{\bar{n}_K}{\bar{n}_J}\frac{n-J+1}{n}\right)=\sigma^2T\frac{(K-J)\bar{n}_K}{n}. \tag{4.20}$$

A further modification of our estimator is thus the area adjusted quantity

$$\widehat{\langle X,X\rangle}_T^{(\text{tsrv,aa})} = \frac{n}{(K-J)\bar{n}_K}\widehat{\langle X,X\rangle}_T^{(\text{tsrv})}. \tag{4.21}$$

Since, by (4.5), $\frac{n}{(K-J)\bar{n}_K} \sim \left(1-\frac{\bar{n}_K}{\bar{n}_J}\right)^{-1}$, we have:

Proposition 4. *The estimator $\widehat{\langle X,X\rangle}_T^{(tsrv,aa)}$ has the same asymptotics as $\widehat{\langle X,X\rangle}_T^{(tsrv,adj)}$ given in Proposition 3.*

The further adjustment therefore does no harm asymptotically. Because of its more careful treatment of small-sample unbiasedness, the area adjusted estimator (4.21) is the one we would most often recommend, especially for moderate sample size. It should be emphasized, however, that the bias-calculation is based on an assumption of constant σ and on borrowing information from the middle

In conclusion, the TSRV estimator that is robust to serial dependence in the noise, behaves as follows:

$$\widehat{\langle X,X\rangle}_T^{(\text{tsrv,aa})} \overset{\mathcal{L}}{\approx} \langle X,X\rangle_T + \frac{1}{n^{1/6}}\Big[\underbrace{\underbrace{\frac{1}{c^2}\xi^2}_{\text{due to noise}} + \underbrace{c\frac{4T}{3}\int_0^T \sigma_t^4 dt}_{\text{due to discretization}}}_{\text{total variance}}\Big]^{1/2} Z_{\text{total}} \tag{4.22}$$

whether it is taken in the form $\widehat{\langle X,X\rangle}_T^{(\text{tsrv,aa})}$ or $\widehat{\langle X,X\rangle}_T^{(\text{tsrv,adj})}$. Here $K \sim cn^{2/3}$, and ξ is given by (4.10), or, more generally, (4.11).

Feasible implementation of the procedure depends on having an estimate of the $\int_0^T \sigma_t^4 dt$. Several schemes are possible; one is discussed in Section 6 of Zhang et al. (2005), another is to use preaveraging (see Jacod et al. (2009) and Podolskij and Vetter (2009)).

5. RV Under Serial Dependence in the Noise

We now turn to an analysis of the standard RV estimator when the noise is serially dependent. First, we have than sparse sampling at a given n_{sparse} results in the same asymptotic distribution as when the noise is serially uncorrelated. Second, however, we find that dependence in the noise impacts the both the bias and the asymptotic variance of the RV estimator when all the data (all n) are used.

Specifically, the traditional RV estimator, $[Y,Y]_T^{(\text{sparse})}$, computed at a sparse sampling frequency $\Delta_{\text{sparse}} = T/n_{\text{sparse}}$, has the following behavior:

$$[Y,Y]_T^{(\text{sparse})} \overset{\mathcal{L}}{\approx} \langle X,X\rangle_T + \underbrace{2n_{\text{sparse}}E\epsilon^2}_{\text{bias due to noise}} + [\underbrace{\underbrace{4n_{\text{sparse}}E\epsilon^4}_{\text{due to noise}} + \underbrace{\frac{2T}{n_{\text{sparse}}}\int_0^T \sigma_t^4 dt}_{\text{due to discretization}}}_{\text{total variance}}]^{1/2} Z_{\text{total}}. \tag{5.1}$$

The reason why this last expression is as in the iid case is as follows. Essentially, the asymptotic variance of $\sum_i \epsilon_{t_i}\epsilon_{t_{i+J}}$ behaves as if the quantities were uncorrelated if J goes to infinity, and so when we have $n_{\text{sparse}} = n/J$ go to infinity with n effectively the log-returns involved in $[Y,Y]_T^{(\text{sparse})}$ are separated by enough time interval for the dependence in ϵ not not to matter. Then $\sum_i \epsilon_{t_i}\left(\epsilon_{t_{i+J}} - E\left[\epsilon_{t_{i+J}}|\mathcal{F}_i\right]\right)$ is a martingale, with variance

$\sum_i E\left[\epsilon_{t_i}^2\left(\epsilon_{t_{i+J}} - E\left[\epsilon_{t_{i+J}}|\mathcal{F}_i\right]\right)^2\right]$, which is approximately $n_{\text{sparse}} E\left[\epsilon^2\right]^2$ under exponential mixing, while the remainder term $\sum_i \epsilon_{t_i}\left(\epsilon_{t_{i+J}} - E\left[\epsilon_{t_{i+J}}|\mathcal{F}_i\right]\right)$ becomes negligible under exponential mixing.

By contrast, when all the observations are used, as in $[Y, Y]_T^{(\text{all})}$, the asymptotic variance of RV is influenced by the dependence of the noise. The asymptotics of $[Y, Y]_T^{(\text{all})}$ are, to first order, like that of $[\epsilon, \epsilon]$. The mean of the latter is $2nE\epsilon^2$. As for the asymptotic variance, from standard formulas for mixing sums, we have

$$\begin{aligned}\Omega_\infty &= \text{AVAR}\left[\sqrt{n}\left(\frac{[\epsilon,\epsilon]}{n} - 2E\epsilon^2\right)\right] \\ &= \text{Var}((\epsilon_1-\epsilon_0)^2) + 2\sum_{i=1}^{\infty}\text{Cov}((\epsilon_1-\epsilon_0)^2, (\epsilon_{i+1}-\epsilon_i)^2). \qquad (5.2)\end{aligned}$$

This gives, by contrast, that the RV estimator using all the data, $[Y, Y]_T^{(\text{all})}$, computed at the highest sampling frequency $\Delta = T/n$, has the following behavior:

$$[Y, Y]_T^{(\text{all})} \overset{\mathcal{L}}{\approx} \langle X, X\rangle_T + \underbrace{2n(E\epsilon^2 + E\epsilon_{t_0}\epsilon_{t_1})}_{\text{bias due to noise}} + \underbrace{[\underbrace{4n\,\Omega_\infty}_{\text{due to noise}} + \underbrace{\frac{2T}{n}\int_0^T \sigma_t^4\,dt}_{\text{due to discretization}}]}_{\text{total variance}}{}^{1/2} Z_{\text{total}}. \qquad (5.3)$$

where $\Omega_\infty = E\epsilon^4$ when the noise is iid; otherwise, dependency in ϵ gives rise to Ω_∞ in (5.2).

An alternate expression for Ω_∞ can be obtained by noting that

$$\text{Cov}((\epsilon_1-\epsilon_0)^2, (\epsilon_{i+1}-\epsilon_i)^2) = 2\,\text{Cov}(\epsilon_1-\epsilon_0, \epsilon_{i+1}-\epsilon_i)^2 + \text{Cum}(\epsilon_1-\epsilon_0, \epsilon_1-\epsilon_0, \epsilon_{i+1}-\epsilon_i, \epsilon_{i+1}-\epsilon_i)$$

(and similarly for the variance).

6. MSRV: Multiple Scales Realized Volatility

6.1. *Review of the MSRV Estimator*

We have seen that TSRV provides the first consistent and asymptotic (mixed) normal estimator of the quadratic variation $\langle X, X\rangle_T$, that it can be made to work even if market microstructure noise is serially dependent, and that it has the rate of convergence $n^{-1/6}$. At the cost of higher complexity, it is possible to generalize TSRV to multiple time scales, by averaging not on two time scales but on multiple time scales. The resulting estimator, *multiple scale realized volatility (MSRV)*, has the form of

$$\widehat{\langle X, X\rangle}_T^{(\text{msrv})} = \underbrace{\sum_{i=1}^M a_i\, [Y, Y]_T^{(K_i)}}_{\text{weighted sum of } M \text{ slow time scales}} + 2\underbrace{\widehat{E\epsilon^2}}_{\text{fast time scale}}. \qquad (6.1)$$

where $\widehat{E\epsilon^2}$ is given as before in (2.2).

The estimator was introduced by Zhang (2006). It was shown there that for suitably selected weights a_i's, $\widehat{\langle X, X\rangle}_T^{(\text{msrv})}$ converges to the true $\langle X, X\rangle_T$ at rate $n^{-1/4}$ when the noise is iid. We here show that a similar result holds under dependent noise. The class of weights are as in the earlier paper, given by

$$a_i = \frac{i}{M^2}h\left(\frac{i}{M}\right) - \frac{1}{2M^2}\frac{i}{M}h'\left(\frac{i}{M}\right), \qquad (6.2)$$

where h is a continuously differentiable real-value function with derivative h', and satisfying the following two conditions:

$$\int_0^1 xh(x)dx = 1 \text{ and } \int_0^1 h(x)dx = 0. \tag{6.3}$$

The key step to the asymptotic analysis is the decomposition

$$\begin{aligned}
\widehat{\langle X, X\rangle}_T^{(\text{msrv})} &= \underbrace{\sum_{i=1}^M a_i[X,X]_T^{(K_i)}}_{\text{signal}} + \underbrace{\sum_{i=1}^M a_i U_{n,K_i}}_{\text{noise}} + \underbrace{2\sum_{i=1}^M a_i[X,\epsilon]_T^{(K_i)}}_{\text{signal-noise interaction}} \\
&+ \underbrace{\sum_{i=1}^M a_i E_{n,K_i} + 2E\epsilon^2}_{\text{end points of noise}} + O_p(n^{-1/2}),
\end{aligned} \tag{6.4}$$

where $U_{n,K} = -\frac{2}{K}\sum_{i=K}^n \epsilon_{t_i}\epsilon_{t_{i-K}}$, and $E_{n,K} = -\frac{1}{K}\sum_{j=0}^{K-1}\epsilon_{t_j}^2 - \frac{1}{K}\sum_{j=n-K+1}^n \epsilon_{t_j}^2$. Conditions (6.3) ensure that the first term in (6.4) will be asymptotically unbiased for $\langle X, X\rangle_T$.

6.2. *The Asymptotics of MSRV When the Noise is Serially Dependent*

We now study the MSRV estimator under the (dependence) Assumption 1 from Section 4.1. The only extra bias (due to dependency) in equation (6.4) comes from the U_{n,K_i}. Under our mixing assumption,

$$|E[U_{n,K_i}]| \leq \frac{2}{K}\operatorname{Var}(\epsilon)\sum_{j=K_i}^n \rho^i \leq \frac{2}{K}\operatorname{Var}(\epsilon)\frac{\rho^{K_i}}{1-\rho}$$

Thus, when the a_i follow (6.2), the absolute value of the extra bias in (6.4) becomes

$$\begin{aligned}
\left|E\left[\sum_{i=1}^M a_i U_{n,i}\right]\right| &\leq \sum_{i=1}^M \frac{2}{M^2}\left|h(\frac{i}{M})\right|\frac{\rho^i}{1-\rho} + \text{corresponding } h' \text{ term} \\
&= O(M^{-1})
\end{aligned} \tag{6.5}$$

To the extent that the MSRV estimator converges at the rate $O_p(M^{-1/2})$, the bias induced by the dependence of the ϵ's is therefore irrelevant asymptotically.

An inspection of the terms in (6.4) shows that the rate of convergence does, indeed, remain of order $O_p(M^{-1/2}) = O_p(n^{-1/4})$ under Assumption 1. As in the TSRV case, however, the asymptotic (random) variance now changes due to the dependence of the ϵ's. To compute that variance when the market microstructure noise is serially dependent, note first that the four terms in (6.4) are asymptotically independent. This is by the same methods as we use in the following. Also, the behavior of the signal term is, obviously, unchanged.

We compute the covariances of the individual terms, and obtain:

Proposition 5. *Assume the conditions of Theorem 4 in Zhang (2006), except that the noise process ϵ, instead of being iid, is now a dependent process following Assumption (1), and with autocorrelation function $\gamma(l) =$*

$\operatorname{Cov}(\epsilon_0, \epsilon_{t_l})$. *Assume that $M = M_n$ satisfies $M_n/n^{1/2} \to 0$. In this case, the following expression has a limit:*

$$
\begin{aligned}
& c\frac{4}{3}T\eta^2 \int_0^1 dx \int_0^x h(y)h(x)y^2(3x-y)dy \\
& + c^{-1}\frac{1}{M^3}\sum_{J=1}^{M}\sum_{K=1}^{M} h(\frac{J}{M})h(\frac{K}{M})\,\langle X, X\rangle_T \sum_{l=-J}^{K}(\gamma(l)+\gamma(l+J-K)-\gamma(l-K)-\gamma(l+J))(\min(l+J,K)-\max(0,l)) \\
& + c^{-1}\frac{4}{M^3}\sum_{J=1}^{M}\sum_{K=1}^{M} h(\frac{J}{M})h(\frac{K}{M}) \sum_{l=-(n-K)}^{n-J}(\min(n,n+l)-\max(J+l,K)+1)^+ \operatorname{Cov}(\epsilon_{t_0}\epsilon_{t_{-J}}, \epsilon_{t_l}\epsilon_{t_{l-K}}) \qquad (6.6) \\
& + c^{-1}\frac{2}{M^3}\sum_{J=1}^{M}\sum_{K=1}^{M} h(\frac{J}{M})h(\frac{K}{M}) \sum_{l=-(J-1)}^{K-1}(\min(J+l,K)-\max(0,l)+1)^+ \operatorname{Cov}(\epsilon_{t_0}^2, \epsilon_{t_l}^2).
\end{aligned}
$$

Here, η^2 is given by equation (30) in Zhang (2006). Furthermore, $n^{1/4}(\widehat{\langle X, X\rangle}_T^{(msrv)} - \langle X, X\rangle_T)$ converges stably to a mixed normal distribution with mean zero and (random) variance given as the limit of (6.6).

We have stated a prelimiting expression in (6.6) since this is closer to the small sample variance.

In the special case of model (3.3), we have

$$
\gamma(l) = \begin{cases} 2E\left[U^2\right] + 2\left(1-\rho\right)E\left[V^2\right] & \text{if } l = 0 \\ -E\left[U^2\right] - \left(1-\rho\right)^2 E\left[V^2\right] & \text{if } |l| = 1 \\ -\rho^{|l|-1}\left(1-\rho\right)^2 E\left[V^2\right] & \text{if } |l| > 1 \end{cases} \qquad (6.7)
$$

to be inserted in (6.6). It can also be noted that, if one sets $\rho = 0$ then the expression (6.6) reduces to the asymptotic variance of MSRV in the iid case, see Remark 3 below.

We conclude our analysis of MSRV with three additional comments:

Remark 1. *Recall that for TSRV we replaced (2.3) with (4.4), thereby "jumping" to frequencies (J, K) over the very fastest one $(1, K)$ at which the serial dependence in the noise manifests itself. By letting both J and K go to infinity with n, we were effectively able to eliminate the serial dependence within each subgrid. However, the asymptotic variance of TSRV is affected by the serial dependence across subgrids coming from the averaging over RV from different subgrids in both $[Y, Y]_T^{(K)}$ and $[Y, Y]_T^{(J)}$ in (4.4), hence the asymptotic variance in Proposition 3, also in (4.22), which is different than in the iid noise case.*

Remark 2. *The asymptotic distribution of MSRV is also affected by the dependence of the noise. Unlike the TSRV case, there is no benefit to adjusting the MSRV estimator in the presence of serial dependence in the noise. This is because under (6.3), the weights a_i in (6.2) already assign most of the mass on the interval $[1, M]$ with $M = O(n^{1/2})$ to subintervals of the form $[cM, M]$ where c is a positive constant. Therefore, the very fastest frequencies of observations (those close to $m = 1$) already play a small role in MSRV even under iid noise.*

Remark 3. *The result specializes to Theorem 4 of Zhang (2006) in the case of iid noise. As choice of weight function (kernel) h, one can either use the noise optimal choice (25) (p. 1027) from the earlier paper, or one can seek to optimize the overall variance (6.6) as a function of h. This would be analogous to the kernel design question which is the subject of Barndorff-Nielsen et al. (2008) (in the context of their autocovariance based estimator).*

7. Empirical Analysis

With these theoretical results in hand, we now turn to a comparison of the empirical performance of the RV, TSRV and MSRV estimators, study the impact of the selection of the fast and slow time scales on the TSRV estimators and the improvement due to MSRV relative to TSRV in the context of transactions data for Intel and Microsoft in the last ten days of April 2004.

In practice, our estimators require that a choice be made for K in the basic TSRV estimator, and for (J, K) in the dependence-robust TSRV estimator. Our asymptotic formulae give the rate at which K needs to grow with n, and a relative rate of convergence for J, which leaves open the question of how to choose the constant c in front of those rates. This is of course an issue that is not unique to these estimators but rather shared by essentially every estimator that has some nonparametric feature. In this particular problem, fortunately, J and K have a natural interpretation in terms of numbers of observations that translate into sampling intervals that are meaningful for the asset price series under consideration.

Our view is that one should start with plausible values of the constants c (say, J corresponding to 1 mn and K to 5 mn for the type of liquid stocks we are studying in this paper) and then compute the estimators for values of the constants c ranging for a few minutes around those center points. This is what we will do in our empirical analysis. As we can see from the figures below, the resulting TSRV estimates are quite robust to varying J and K in those ranges.

7.1. *Comparison of the RV and TSRV Estimators*

In our empirical analysis of the different estimators, we start by comparing our TSRV estimator to the traditional RV estimator. In particular, we establish that TSRV solves the two main problems associated with RV, namely the divergence of RV as the sampling interval gets small and the variability of RV. The comparison is reported for Intel in Figure 6 and for Microsoft in Figure 7, where we compare RV computed at different sampling frequencies with TSRV computed for a range of values of K of 2 to 8mn around the typical 5mn value.

Besides the well-known divergence of RV as $\Delta \to 0$, the two figures also demonstrate the large difference in variability of both estimates. Without the benefits of the double averaging in Figure 1, what these two series of plots show is that computing RV at, say, 4mn as opposed to 5mn or 6mn can result in substantially different daily estimates. And the computation of day-by-day estimates is how RV is actually used. In existing empirical applications, RV has typically been employed in the empirical literature at an arbitrary sparse frequency: in light of the variability of RV as a function of the sparse sampling interval Δ_{sparse}, whatever particular choice is made can matter.

7.2. *Robustness of TSRV to the Choice of Slow Time Scale*

Both RV and TSRV require that the econometrician make a choice. In RV, one needs to select the sparse sampling frequency at which to compute the estimator. In TSRV, one needs to select the number of subgrids K over which to average the slow time scales sum of squares. In Zhang et al. (2005), we showed how to compute an optimal value K^* for the slow time scale parameter K when the noise term is assumed to be iid, but in practical application it would be beneficial to be able to dispense with that computation. For that, we would need to establish that the TSRV estimator is empirically robust to departures from the optimal K^*. So, we now examine and compare the robustness of the two estimators to the selection of their respective free parameter.

The left panels in Figure 8 show RV, computed for Intel and Microsoft, as an average of the RV values over the last ten days in April 2004 for different sparse sampling frequency (which is the choice parameter for RV). The right panels report the robustness of TSRV to the choice of K. The right panels show that the estimator is numerically very robust to a range of choices of K. In other words, the value of the TSRV estimator is largely unaffected by the choice of K within a reasonable range corresponding, in sampling interval scale, to 2 to 10 mn.

7.3. Robustness of TSRV to the Choice of Slow and Fast Time Scales

When the noise is serially dependent, the TSRV estimator defined in (4.4) depends on the choice of both the slow time scale (K) and the fast time scale (J). We find that the time-dependent TSRV is quite robust to the choice of (J, K). Figure 9 shows that the value of the estimator is essentially identical within the range of values considered, which correspond in sampling interval scale to a few seconds to 2 minutes for J and 5 minutes to 10 minutes for K.

7.4. The Improvement in MSRV over TSRV

It is possible to improve upon TSRV by considering MSRV. Both are consistent estimators of $\langle X, X \rangle_T$, but MSRV has the faster convergence rate $n^{-1/4}$ vs. $n^{-1/6}$ for TSRV. The trade-off involves the additional computational burden, since the number of slow time scales to be computed for MSRV is $M = O(n^{1/2})$ and n can be large in empirical applications: for instance $n = 23,400$ for a stock that trades on average once a second for a full day.

We now examine the difference between TSRV and MSRV in the context of our empirical application. Figure 10 shows that both methods produce close estimates, especially when compared to the differences exhibited earlier between RV and TSRV. As in the case of TSRV shown in 9, the MSRV estimator is not sensitive to the specific choice of M within the range considered (up to 500).

7.5. Robustness to Data Cleaning Procedures

One aspect that is sometimes briefly mentioned, but often not emphasized, in empirical papers using high frequency financial data is the fact that the raw data is typically pre-processed to eliminate data errors, outliers, etc. In addition, empirical applications of RV can involve pre-filtering of the data of various types, but we focus here on the impact of data cleaning procedures that typically take place before any actual RV computation is performed.

It turns out that the impact of the specific data-cleaning procedures used to pre-process the raw data can have a large impact on RV estimators. We illustrate this effect by considering different cutoffs to determine which outliers to eliminate before calculating the RV estimator. First, we eliminate the obvious data errors (such as a transaction price reported as zero, transaction times that are out of order, etc.).

Second, we seek to eliminate outliers of various sizes. This is where things get trickier. For that purpose, an outlier is a “bounceback”: 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. Certainly, we do not expect such large “roundtrips” to represent meaningful transactions. The question is how large is large, and so we are led to study the dependence of the RV and TSRV estimators on three different cutoffs that could conceivably be adopted, 0.1% and 1% respectively in log-returns terms, and no cutoff (no raw bounceback return is larger than 2% in our sample, so that any cutoff larger than this would make no difference). The analysis reported above is all based on the intermediary cutoff of 1%.

The left panels in Figure 11 show the large impact of the cutoff on the RV estimator. As shown in the right panel, where all three curves are close together, TSRV is much less sensitive to the specific cutoff used. This is due to the structure of TSRV as a difference of two estimators: large returns in the data are part of the slow time scale calculation, but then subtracted out in the fast time scale one.

Since the cutoff level is essentially arbitrary, we can view such outliers as a form of market microstructure noise, and the robustness of TSRV to different ways of pre-processing the data is therefore a desirable property.

8. Conclusions

Market microstructure noise contained in high frequency financial data can exhibit serial correlation, as was suggested in Aït-Sahalia et al. (2005) and Hansen and Lunde (2006). We showed that combining two or more time scales for the purpose of estimating integrated volatility will work even in the situation where the microstructure noise exhibits time series dependence, thereby making the TSRV construction robust to this departure from the basic assumptions in Zhang et al. (2005).

In most data error situations, one might expect that progress will lead the issue to become somehow less salient over time. But in this instance, the measurement errors we face in ultra high frequency data are compounded by the institutional evolution of the equity markets. While changes such as the passage to decimalization contribute to reducing the amount of noise in the data, by reducing the rounding errors, the emergence of competing electronic networks means that multiple transactions can be executed (and ultimately reported in our database) on different exchanges at the same time, thereby increasing the potential for slight time reporting mismatches and other forms of data error.

Indeed, the data generated by the individual market venues find their way to the public in various ways. The principal Electronic Communication Networks (ECNs), such as INET and its precursors and Archipelago, have high speed dissemination directly to their subscribers. These dissemination systems run on a telecommunications protocol known as "frame relay", which is quite fast. Most other market data, however, reaches the trading public (and ultimately us econometricians) either through Nasdaq's dissemination or CTS/CQS. The Consolidated Tape Association administers the CTS (the consolidated trade system) and CQS (the consolidated quote system). Virtually all US trades are reported to CTS, but the path may be indirect. Island (not INET) may report a trade to the National (formerly Cincinnati) Stock Exchange, which will then report it to CTS, which then broadcasts it to us. The general problem is that trading activity is fast relative to the CTS speed of collection and dissemination.

Furthermore, Nasdaq has had long-standing issues with late and delayed trade reports. In principle, a Nasdaq member has up to 30 (in the past, 90) seconds to report a trade and anecdotal evidence suggests that some dealers were/are using this leeway to its greatest extent. Since this practice was not uniform across dealers and across time, the sequencing can be suspect. And the sequencing across exchanges may be unreliable over very short time intervals: a trade on one exchange followed (and time-stamped to the same second as) a trade in the same stock on a different exchange, may not in fact have occurred in that order.

While the consolidated tape feed (which we see on TAQ) is probably the best source of data available, we may not be seeing the trades in the order in which they occurred, and the emergence and further development of alternative networks on which to trade the same stocks makes the issue of market microstructure noise in the data an increasing, not decreasing, one. ECNs represent over 30% of Nasdaq trading volume and are increasing their market share in NYSE-listed issues as well (see e.g., Barclay et al. (2003)). Clearly, the decentralization of trading, combined with the increased frequency of trading, create challenges for the data collection which ultimately affect the estimation of a quantity as basic as the daily integrated volatility of the price. So there are reasons to believe that the issue of controlling for market microstructure noise in high frequency financial econometrics will be with us for some time.

References

Aït-Sahalia, Y., Mykland, P. A., 2003. The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica 71, 483–549.

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2006. Comments on "Realized variance and market microstructure noise ". Journal of Business and Economic Statistics 24, 162–167.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of exchange rate realized volatility. Journal of the American Statistical Association 96, 42–55.

Bandi, F. M., Russell, J. R., 2008. Microstructure noise, realized volatility and optimal sampling. Review of Economic Studies 75, 339–369.

Barclay, M. J., Hendershott, T., McCormick, D. T., 2003. Competition among trading venues: Information and trading on Electronic Communications Networks. Journal of Finance 58, 2637–2665.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O. E., 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.

Black, F., 1986. Noise. Journal of Finance 41, 529–543.

Choi, J. Y., Salandro, D., Shastri, K., 1988. On the estimation of bid-ask spreads: Theory and evidence. The Journal of Financial and Quantitative Analysis 23, 219–230.

Curci, G., Corsi, F., 2005. A discrete sine transform approach for realized volatility measurement. Tech. rep., University of Lugano.

Delattre, S., Jacod, J., 1997. A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. Bernoulli 3, 1–28.

French, K., Roll, R., 1986. Stock return variances: The arrival of information and the reaction of traders. Journal of Financial Economics 17, 5–26.

Gençay, R., Ballocchi, G., Dacorogna, M., Olsen, R., Pictet, O., 2002. Real-time trading models and the statistical properties of foreign exchange rates. International Economic Review 43, 463–491.

Gloter, A., Jacod, J., 2000. Diffusions with measurement errors: I - local asymptotic normality and II - optimal estimators. Tech. rep., Université de Paris-6.

Gonçalves, S., Meddahi, N., 2009. Bootstrapping realized volatility. Econometrica 77, 283–306.

Griffin, J., Oomen, R., 2008. Sampling returns for realized variance calculations: Tick time or transaction time? Econometric Reviews 27, 230–253.

Hall, P., Heyde, C. C., 1980. Martingale Limit Theory and Its Application. Academic Press, Boston.

Hansen, P. R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161.

Harris, L., 1990. Statistical properties of the Roll serial covariance bid/ask spread estimator. Journal of Finance 45, 579–590.

Jacod, J., 1994. Limit of random measures associated with the increments of a Brownian semimartingale. Tech. rep., Université de Paris-6.

Jacod, J., 1996. La variation quadratique du Brownien en présence d'erreurs d'arrondi. Astérisque 236, 155–162.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Processes and Their Applications 119, 2249–2276.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. Annals of Probability 26, 267–307.

Li, Y., Mykland, P. A., 2007. Are volatility estimators robust with respect to modeling assumptions? Bernoulli 13, 601–622.

Mykland, P. A., Zhang, L., 2006. ANOVA for diffusions and Itô processes. Annals of Statistics 34, 1931–1963.

Oomen, R. C., 2006. Properties of realized variance under alternative sampling schemes. Journal of Business and Economic Statistics 24, 219–237.

Podolskij, M., Vetter, M., 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli, forthcoming .

Roll, R., 1984. A simple model of the implicit bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Stoll, H., 2000. Friction. Journal of Finance 55, 1479–1514.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12, 1019–1043.

Zhang, L., 2009. Estimating covariation: Epps effect and microstructure noise. Journal of Econometrics, forthcoming .

Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2009. Edgeworth expansions for realized volatility and related estimators. Journal of Econometrics, forthcoming .

Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business & Economic Statistics 14, 45–52.

Zumbach, G., Corsi, F., Trapletti, A., 2002. Efficient estimation of volatility using high frequency data. Tech. rep., Olsen & Associates.

Appendix: Proofs

A. Proof of Lemma 1

First, $\sum_{i=0}^{n-J}(X_{t_{i+J}} - X_{t_i})(\epsilon_{t_{i+J}} - \epsilon_{t_i}) = \sum_{i=0}^{n}(-c_{i+J} + c_i)\epsilon_{t_i}$ where $c_i = X_{t_i} - X_{t_{i-J}}$ for $J \leq i \leq n$, and $= 0$ otherwise. Second,

$$\begin{aligned}
E\left(\left(\sum_{i=0}^{n}(-c_{i+J}+c_i)\epsilon_{t_i}\right)^2 |X\right) &= \mathrm{Var}(\sum_{i=0}^{n}(-c_{i+J}+c_i)\epsilon_{t_i}|X) \\
&\leq E\epsilon^2\left(\sum_i(-c_{i+J}+c_i)^2 + 2\sum_{l\geq 1}\rho^l|\sum_i(-c_{i+J}+c_i)(-c_{i+J+l}+c_{i+l})|\right) \\
&\leq E\epsilon^2\sum_i(-c_{i+J}+c_i)^2(1+2\sum_{l\geq 1}\rho^l) \\
&\leq 4J[X,X]_T^{(J)}E\epsilon^2(1+2\rho/(1-\rho))
\end{aligned} \tag{A.1}$$

where the last two transitions follow by the Cauchy-Schwarz Inequality. The lemma follows by the Markov inequality. This finishes the proof.

B. Proof of Proposition 1

By (4.7),

$$\begin{aligned}
E\left(\frac{1}{K}\sum_{i=0}^{n}c_i^{(K)}\epsilon_{t_i}^2 - \frac{\bar{n}_K}{\bar{n}_J}\frac{1}{J}\sum_{i=0}^{n}c_i^{(J)}\epsilon_{t_i}^2\right)^2 &= \mathrm{Var}\left(\frac{1}{K}\sum_{i=0}^{n}c_i^{(K)}\epsilon_{t_i}^2 - \frac{\bar{n}_K}{\bar{n}_J}\frac{1}{J}\sum_{i=0}^{n}c_i^{(J)}\epsilon_{t_i}^2\right) \\
&\leq \sum_{i=0}^{n}\left(c_i^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}c_i^{(J)}\right)^2\mathrm{Var}(\epsilon^2) \\
&\quad + 2\sum_{l=1}^{n}\sum_{i=0}^{n-l}\left(c_i^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}c_i^{(J)}\right)\left(c_{i+l}^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}c_{i+l}^{(J)}\right)\mathrm{Cov}(\epsilon_{t_i}^2, \epsilon_{t_{i+l}}^2) \\
&\leq \sum_{i=0}^{n}\left(c_i^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}c_i^{(J)}\right)^2\left(\mathrm{Var}(\epsilon^2) + 2\sum_{l=0}^{n}\mathrm{Cov}(\epsilon_{t_0}^2, \epsilon_{t_l}^2)\right) \\
&= O\left(\sum_{i=0}^{n}\left(c_i^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}c_i^{(J)}\right)^2\right).
\end{aligned} \tag{B.1}$$

where the second to last transition is due to the Cauchy-Schwarz inequality, and the final one follows from our moment and mixing assumptions, again in view of Theorem A.6 (p. 278) of Hall and Heyde (1980). Under (4.5) and by tedious calculation, one obtains that the r.h.s. of (B.1) is no larger than the order $O(J/n)$. (In fact, this is the exact order under the condition (4.12) below.) Thus

$$\text{noise term} = -2\frac{1}{K}\sum_{i=0}^{n-K}\epsilon_{t_i}\epsilon_{t_{i+K}} + 2\frac{\bar{n}_K}{\bar{n}_J}\frac{1}{J}\sum_{i=0}^{n-J}\epsilon_{t_i}\epsilon_{t_{i+J}} + O_p\left(\left(\frac{J}{n}\right)^{1/2}\right),$$

whence, finally,

$$\begin{aligned}
\frac{K}{n^{1/2}}(\text{noise term} - E[\text{noise term}]) &= -2\frac{1}{\sqrt{n}}\sum_{i=0}^{n-K}\epsilon_{t_i}\epsilon_{t_{i+K}} + 2\frac{1}{\sqrt{n}}\sum_{i=0}^{n-J}\epsilon_{t_i}\epsilon_{t_{i+J}} + o_p(1) \\
&\xrightarrow{\mathcal{L}} \xi Z_{\text{noise}}
\end{aligned} \tag{B.2}$$

as $n \to \infty$, where Z_{noise} is standard normal, by the same methods as in Chapter 5 of Hall and Heyde (1980) (we here have a triangular array of sums, but the arguments go through nonetheless). By uniform integrability, the asymptotic variance is on the form $\xi^2 = 4\alpha_0' + 8\sum_{i=1}^{\infty}\alpha_i'$ where α_i' is the limit as $K \to \infty$ (and $J \to \infty$ if such is the case) of

$$\begin{aligned}
\operatorname{Cov}(\epsilon_{t_0}(\epsilon_{t_K}-\epsilon_{t_J}),\epsilon_{t_i}(\epsilon_{t_{i+K}}-\epsilon_{t_{i+J}})) &= \operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_i})\operatorname{Cov}(\epsilon_{t_K}-\epsilon_{t_J},\epsilon_{t_{i+K}}-\epsilon_{t_{i+J}})\\
+\operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_{i+K}}-\epsilon_{t_{i+J}})&\operatorname{Cov}(\epsilon_{t_K}-\epsilon_{t_J},\epsilon_{t_i})+\operatorname{Cum}(\epsilon_{t_0},\epsilon_{t_i},\epsilon_{t_K}-\epsilon_{t_J},\epsilon_{t_{i+K}}-\epsilon_{t_{i+J}}).
\end{aligned}$$

For fixed J, this means that

$$\alpha_i' = 2\operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_i})^2+\operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_{i+J}})\operatorname{Cov}(\epsilon_{t_i},\epsilon_{t_J})+\operatorname{Cum}(\epsilon_{t_0},\epsilon_{t_i},\epsilon_{t_J},\epsilon_{t_{i+J}}).$$

The limit for $J \to \infty$ is $2\operatorname{Cov}(\epsilon_{t_0},\epsilon_{t_i})^2$. This shows the result.

C. Proof of Proposition 5

Consider first the behavior of the $[X,\epsilon]^{(K)}$ term. The conditional covariance behaves as follows:

$$\begin{aligned}
\operatorname{Cov}\left([X,\epsilon]^{(J)},[X,\epsilon]^{(K)}\middle| X\text{ process}\right) &= \frac{1}{JK}\sum_{j=J}^{n}\sum_{k=K}^{n}(X_{t_j}-X_{t_{j-J}})(X_{t_k}-X_{t_{k-K}})Cov(\epsilon_{t_j}-\epsilon_{t_{j-J}},\epsilon_{t_k}-\epsilon_{t_{k-K}})\\
&= \frac{1}{JK}\sum_{j=J}^{n}\sum_{k=K}^{n}(X_{t_j}-X_{t_{j-J}})(X_{t_k}-X_{t_{k-K}})\\
&\quad\times(\gamma(j-k)+\gamma(j-k-(J-K))-\gamma(j-k+K)-\gamma(j-k-J))\\
&\approx \frac{1}{JK}\sum_{j=J}^{n}\sum_{k=K}^{n}(\langle X,X\rangle_{\min(t_j,t_k)}-\langle X,X\rangle_{\max(t_{j-J},t_{k-K})})^+\\
&\quad\times(\gamma(j-k)+\gamma(j-k-(J-K))-\gamma(j-k+K)-\gamma(j-k-J))
\end{aligned}$$

so that

$$\begin{aligned}
\operatorname{Cov}\left([X,\epsilon]^{(J)},[X,\epsilon]^{(K)}\middle| X\text{ process}\right) &= \frac{1}{JK}\sum_{l=-J}^{K}\sum_{k=K}^{n}(\langle X,X\rangle_{\min(t_{k-l},t_k)}-\langle X,X\rangle_{\max(t_{k-l-J},t_{k-K})})^+\\
&\quad\times(\gamma(l)+\gamma(l+J-K)-\gamma(l-K)-\gamma(l+J))\\
&= \frac{1}{JK}\sum_{l=-J}^{K}(\gamma(l)+\gamma(l+J-K)-\gamma(l-K)-\gamma(l+J))\\
&\quad\times\sum_{k=K}^{n}(\langle X,X\rangle_{\min(t_{k-l},t_k)}-\langle X,X\rangle_{\max(t_{k-l-J},t_{k-K})})^+.
\end{aligned}\tag{C.1}$$

For the final summation in (C.1), note that this is a telescope sum, of the form (where a and b depend on J, K, and l, and where one can take $a \le b$)

$$\begin{aligned}
\sum_{k=K}^{n}(\langle X,X\rangle_{t_{k-a}}-\langle X,X\rangle_{t_{k-b}}) = \sum_{k=K-a}^{n-a}\langle X,X\rangle_{t_k}-\sum_{k=K-b}^{n-b}\langle X,X\rangle_{t_k} &= \sum_{k=n-b+1}^{n-a}\langle X,X\rangle_{t_k}-\sum_{k=K-b}^{K-a-1}\langle X,X\rangle_{t_k}\\
\approx (b-a)\langle X,X\rangle_T&
\end{aligned}\tag{C.2}$$

since $\langle X,X\rangle_0 = 0$. Specifically, it is easy to see that $a = \max(0,l)$ while $b = \min(l+J,K)$. Thus

$$\begin{aligned}
\operatorname{Cov}\left([X,\epsilon]^{(J)},[X,\epsilon]^{(K)}\middle| X\text{ process}\right) &\approx \langle X,X\rangle_T\frac{1}{JK}\sum_{l=-J}^{K}(\gamma(l)+\gamma(l+J-K)-\gamma(l-K)-\gamma(l+J))\\
&\quad\times(\min(l+J,K)-\max(0,l))
\end{aligned}\tag{C.3}$$

At the same time,

$$\begin{aligned}
\mathrm{Cov}(U_{n,J},U_{n,K}) &= \frac{4}{JK}\,\mathrm{Cov}(\sum_{j=J}^{n}\epsilon_{t_j}\epsilon_{t_{j-J}},\sum_{k=K}^{n}\epsilon_{t_k}\epsilon_{t_{k-K}}) \\
&= \frac{4}{JK}\sum_{j=J}^{n}\sum_{k=K}^{n}\mathrm{Cov}(\epsilon_{t_j}\epsilon_{t_{j-J}},\epsilon_{t_k}\epsilon_{t_{k-K}}) \\
&= \frac{4}{JK}\sum_{j=J}^{n}\sum_{k=K}^{n}\mathrm{Cov}(\epsilon_{t_0}\epsilon_{t_{-J}},\epsilon_{t_{k-j}}\epsilon_{t_{k-j-K}}) \\
&= \frac{4}{JK}\sum_{l=-(n-K)}^{n-J}(\min(n,n+l)-\max(J+l,K)+1)^{+}\,\mathrm{Cov}(\epsilon_{t_0}\epsilon_{t_{-J}},\epsilon_{t_l}\epsilon_{t_{l-K}}), \qquad \text{(C.4)}
\end{aligned}$$

where

$$\mathrm{Cov}(\epsilon_{t_0}\epsilon_{t_{-J}},\epsilon_{t_l}\epsilon_{t_{l-K}}) = \gamma(l)\gamma(l-(J-K)) + \gamma(l-K)\gamma(l+J) + \mathrm{Cum}(\epsilon_{t_0},\epsilon_{t_{-J}},\epsilon_{t_l},\epsilon_{t_{l-K}})$$

Finally,

$$\begin{aligned}
\mathrm{Cov}(E_{n,J},E_{n,K}) &\approx \frac{2}{JK}\,\mathrm{Cov}\left(\sum\nolimits_{j=0}^{J-1}\epsilon_{t_j}^2,\ \sum\nolimits_{k=0}^{K-1}\epsilon_{t_k}^2\right) \\
&= \frac{2}{JK}\sum_{l=-(J-1)}^{K-1}(\min(J+l,K)-\max(0,l)+1)^{+}\,\mathrm{Cov}(\epsilon_{t_0}^2,\epsilon_{t_l}^2) \qquad \text{(C.5)}
\end{aligned}$$

As before, $\mathrm{Cov}(\epsilon_{t_0}^2,\epsilon_{t_l}^2) \;=\; 2\gamma(l)^2 + \mathrm{Cum}(\epsilon_{t_0},\epsilon_{t_0},\epsilon_{t_l},\epsilon_{t_l})$. Following equation (31) (p. 1030) in Zhang (2006), and (C.3)-(C.5) above, we obtain the combined expression given in equation (6.6). Note that in the special case of model (3.3), with Gaussian U and V, the fourth cumulant above is zero, and $\gamma(l)$ is given by (6.7). The arguments involving stable convergence are as in Zhang (2006).

Descriptive Statistics	3M	AIG	Intel	Microsoft
Transactions				
Average number of transactions per day	$2,820$	$3,435$	$13,018$	$14,299$
Average time between transactions (seconds)	8.3	6.8	1.8	1.6
Min log-return from transactions	-0.019	-0.028	-0.044	-0.083
Max log-return	0.019	0.028	0.044	0.082
Average daily first order autocorrelation	-0.41	-0.40	-0.60	-0.63
Average daily second order autocorrelation	0.017	0.08	0.21	0.25
Average daily third order autocorrelation	0.009	-0.01	-0.12	-0.17
Quotes				
Average number of quote revisions per day	$12,824$	$13,507$	$22,275$	$22,661$
Average time between quote revisions	1.8	1.7	1.1	1.0
Min log-return from quote revisions	-0.031	-0.044	-0.016	-0.013
Max log-return	0.034	0.044	0.016	0.013
Average daily first order autocorrelation	-0.49	-0.49	-0.24	-0.23
Average daily second order autocorrelation	0.001	0.001	0.07	0.02
Average daily third order autocorrelation	0.005	0.004	0.03	0.02

Table 1. **Descriptive Statistics**

For the purpose of counting transactions, only transactions leading to a price change are counted. Identical quotes are counted as a single one when reporting the number of quote revisions. Log-returns from quotes are computed using a bid-ask midpoint, weighted by the respective depth of the two sides. Autocorrelations of log-returns are reported in transaction time and quote time, respectively. Averages are computed over the last ten trading days in April 2004 (April 19-23 and 26-30). Minima and maxima are computed over the full ten day sample. All descriptive statistics for the transactions data are reported prior to any data processing, except for the removal of obvious data errors such as prices or quotes reported as zero. The estimates to be computed in the rest of the paper from transaction prices are based on data cleaned to remove any price "bounceback", defined as a price jump of size greater than a cutoff of 1%, immediately followed by a jump of equal magnitude but opposite sign (see Section 7.5 below). The raw quotes data are pre-processed to remove any sets of quotes whose bid or ask price deviate from the closest transaction price recorded by more than 5% (except in instances where the transaction price itself moves by that amount). The data are from the TAQ database.

Autocorrelation Order	Constant	Avg Time Between Transactions	R^2
2	0.25 (8.0)	−0.015 (−3.9)	0.35
3	−0.16 (−6.5)	0.012 (4.2)	0.39
4	0.11 (7.1)	−0.009 (−4.9)	0.46
5	−0.08 (−6.7)	0.008 (5.2)	0.49

Table 2. **Regressions of Higher Order Autocorrelations on Stock Liquidity**

This table reports the results of cross-sectional OLS regressions of the autocorrelation coefficients of order 2-5 on the average time between transactions used as a measure of the liquidity of the stock, for the 30 DJIA stocks. These autocorrelation coefficients would be zero if the noise term were serially uncorrelated. The autocorrelation coefficients are computed for each stock as the average of the daily autocorrelations over the last ten trading days in April 2004. t-statistics are in parentheses.

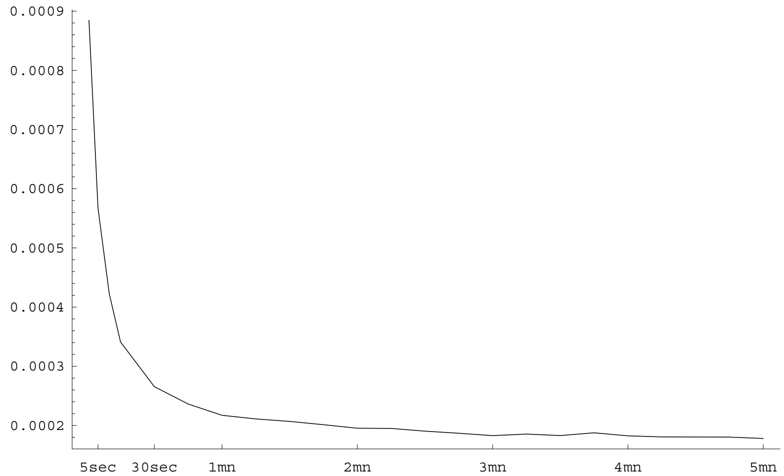


Fig. 1. This figure shows the RV estimator $[Y,Y]_T^{(\text{all})}$ plotted against the sampling interval $\Delta = T/n$. The RV estimator in the figure is computed for an average of the 30 Dow Jones Industrial Average stocks, averaged again over the last ten trading days in April 2004; the objective of the double averaging is to reduce the variability of the estimator in order to display its bias. Since $\Delta = T/n$, the plot illustrates the divergence of RV as $n \to \infty$ predicted by our theory: RV has bias $2nE\left[\varepsilon^2\right]$.

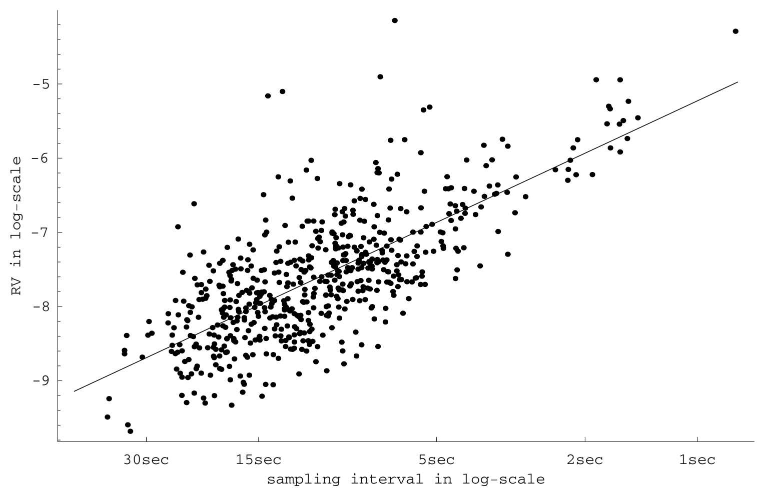


Fig. 2. This figure shows a regression of $\ln RV$ against $\ln n$, plotted in log-log scale. Each data point in the plot represents a triplet (one stock, one day, j) from the 30 DJIA stocks, the last 10 trading days in April 2004, and $j = 1$ or 2 depending upon whether all the observations are used on one out of two. For ease of interpretation, the sample size n on the x-axis is translated into an average sampling interval on the basis of 1 trading day $=$ 6.5 hours $= 23,400$ one-second time intervals.

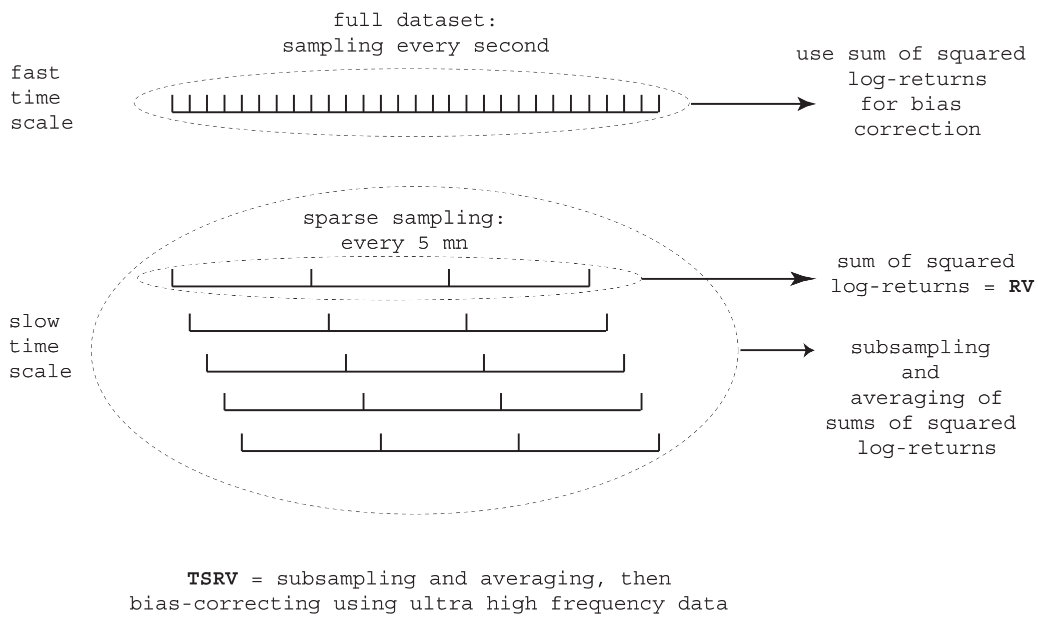


Fig. 3. This figure describes the construction of the TSRV estimator.

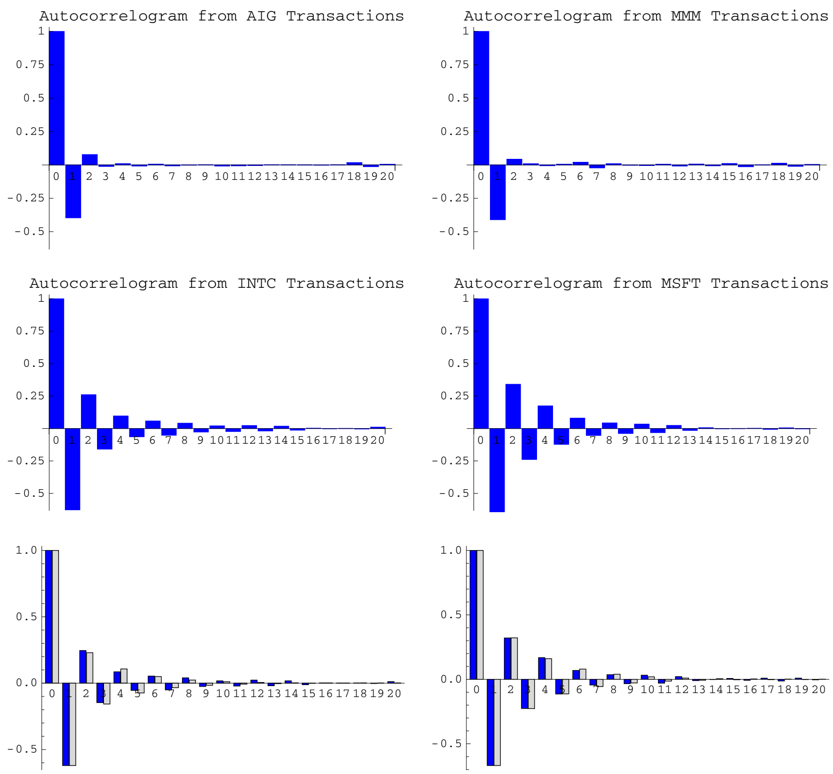


Fig. 4. Top and middle panels: Log-return autocorrelograms from transactions for American International Group, Inc. (AIG), 3M Co. (MMM), Intel (INTC) and Microsoft (MSFT), last ten trading days in April 2004. Bottom panel: log-return autocorrelogram from the same transactions for Intel and Microsoft, superimposed with the autocorrelogram fitted from the basic i.i.d. plus AR(1) model for the noise.

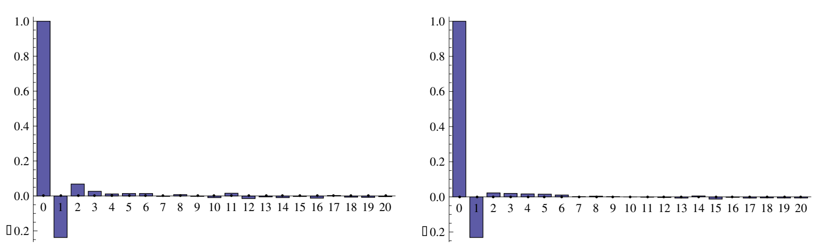


Fig. 5. Log-return autocorrelogram from quote revisions for Intel and Microsoft, last ten trading days in April 2004.

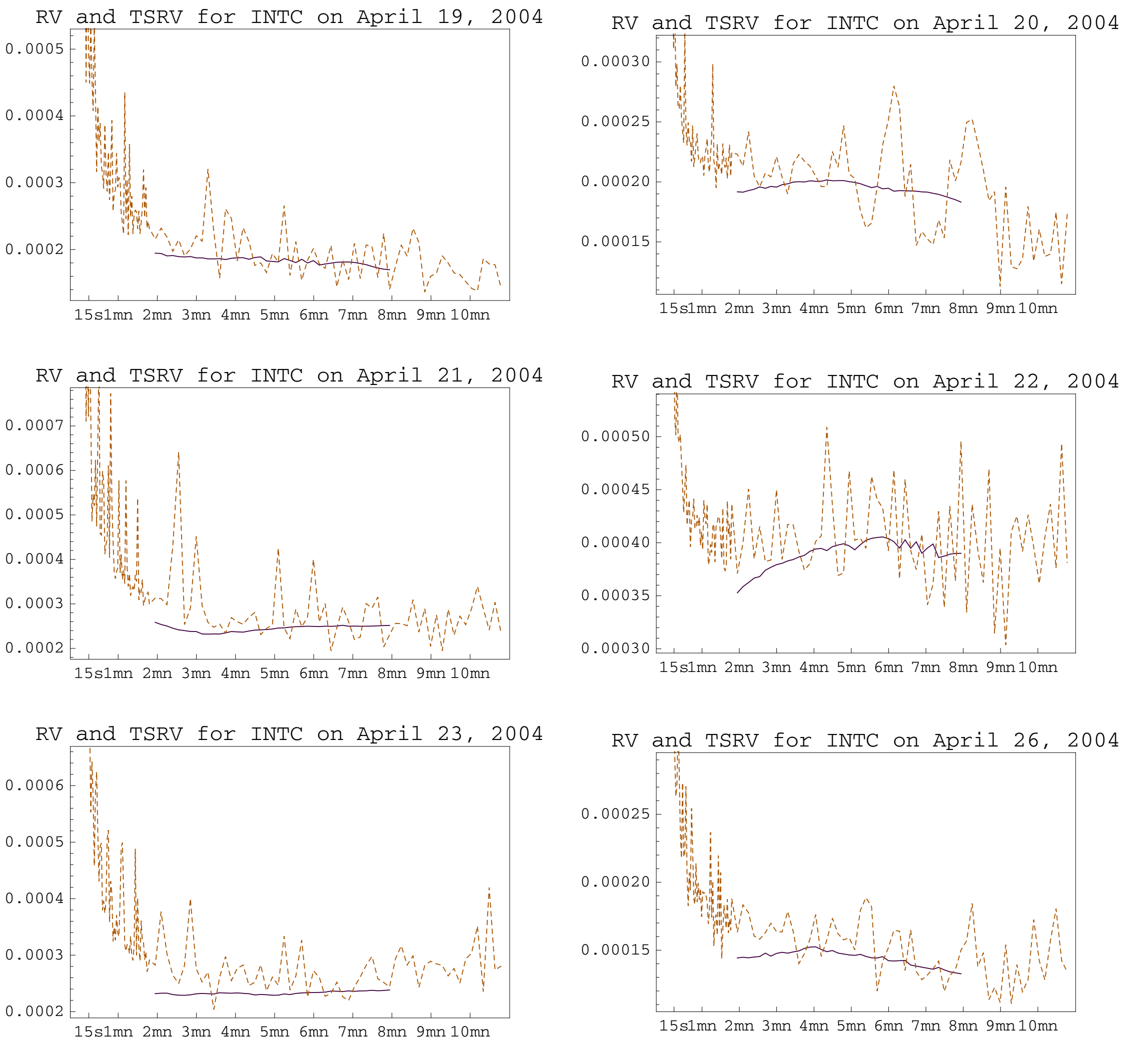


Fig. 6. Comparison of the RV (dashed line) and TSRV (solid line) estimators for Intel, computed on a daily basis, from transaction data.

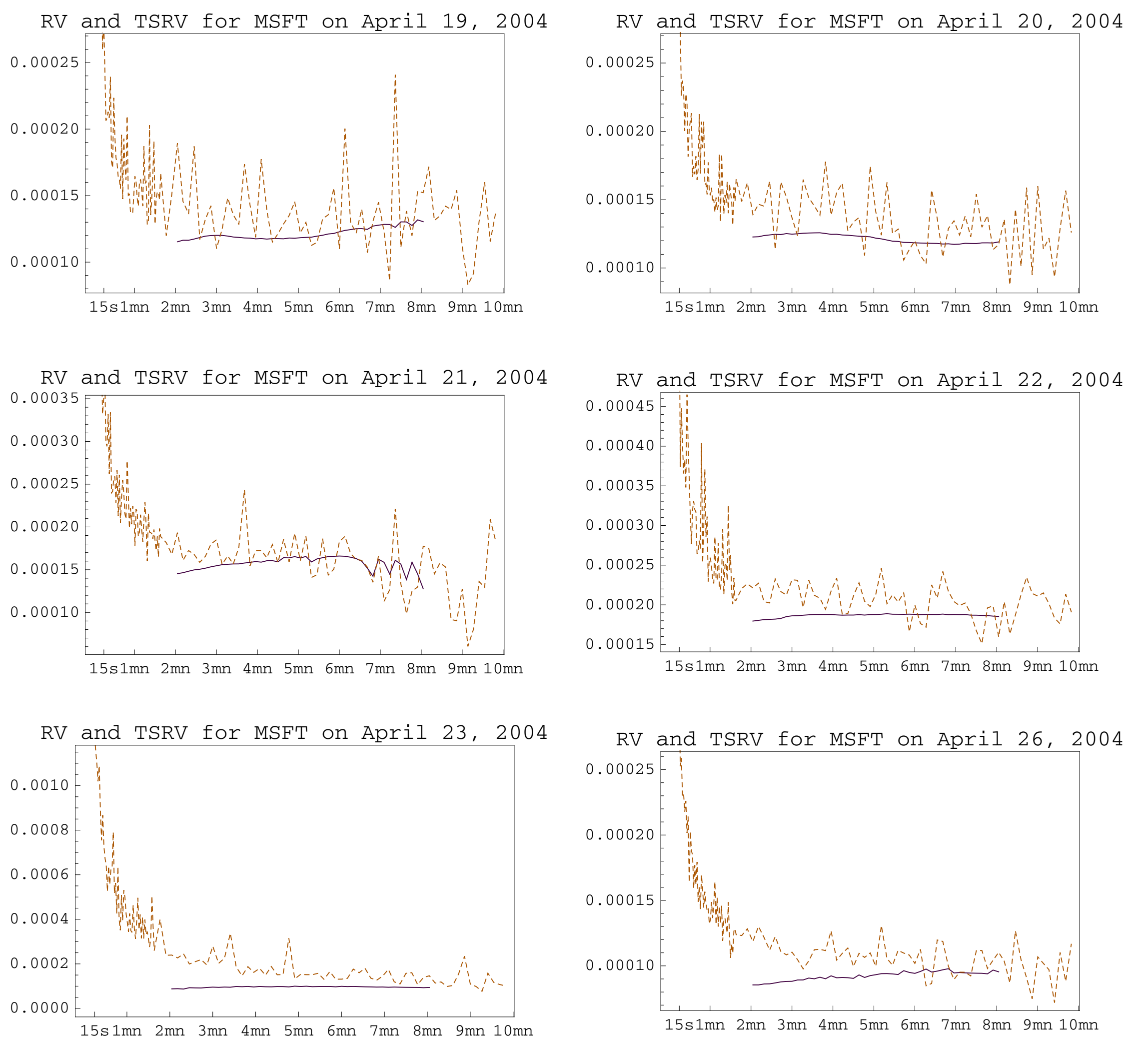


Fig. 7. Comparison of the RV (dashed line) and TSRV (solid line) estimators for Microsoft, computed on a daily basis.

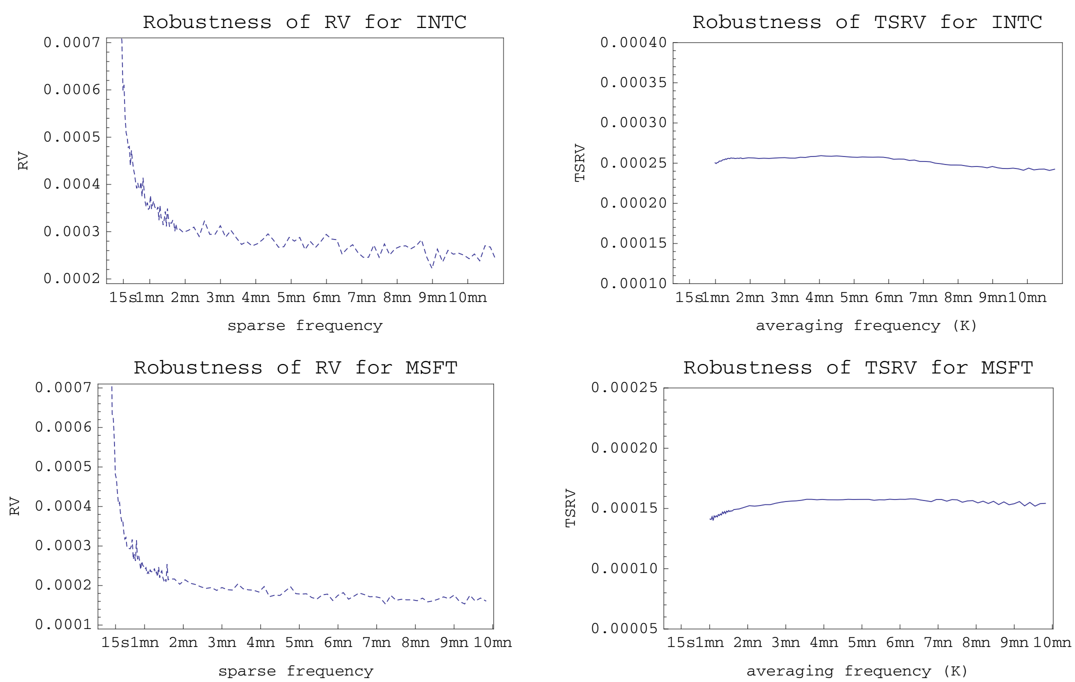
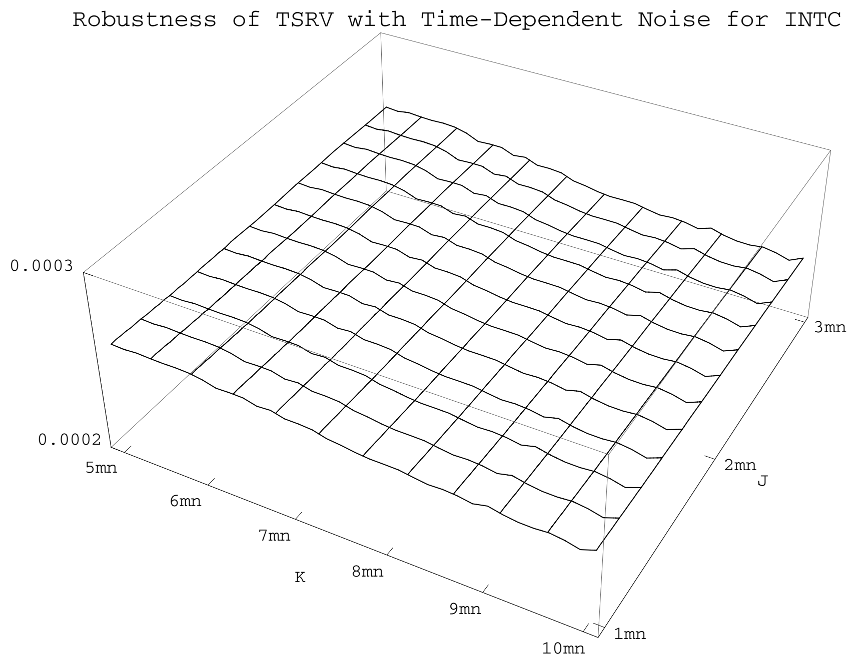


Fig. 8. Comparison of the RV and TSRV estimators for Intel and Microsoft, averaged over the last ten trading days of April 2004. The left panels demonstrate the dependence of RV as a function of the sparse sampling interval, while the right panels study the robustness of TSRV with respect to the choice of the averaging frequency as represented by the number of subgrids K. The estimators were calculated from transaction data.



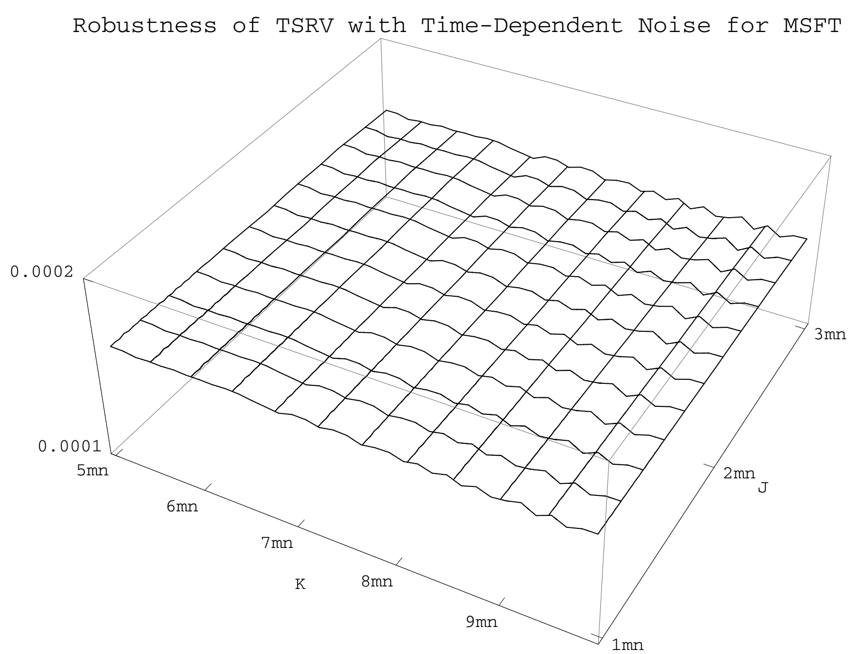


Fig. 9. Robustness of the TSRV estimator for Intel and Microsoft over the choice of the two time scales J (fast) and K (slow), averaged over the last 10 trading days of April 2004.

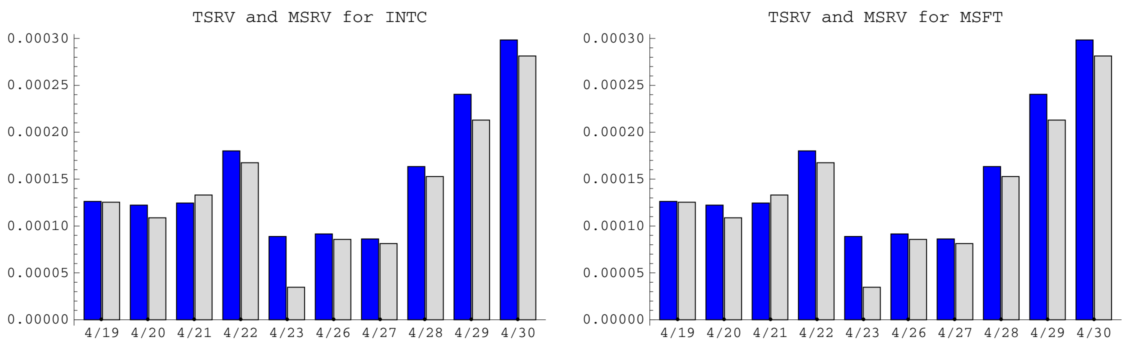


Fig. 10. Comparison of the TSRV and MSRV estimators for Intel and Microsoft, for each of the last ten trading days of April 2004. For each stock and time period, the left bar is TSRV and the right bar is MSRV. The estimators were calculated from transaction data.

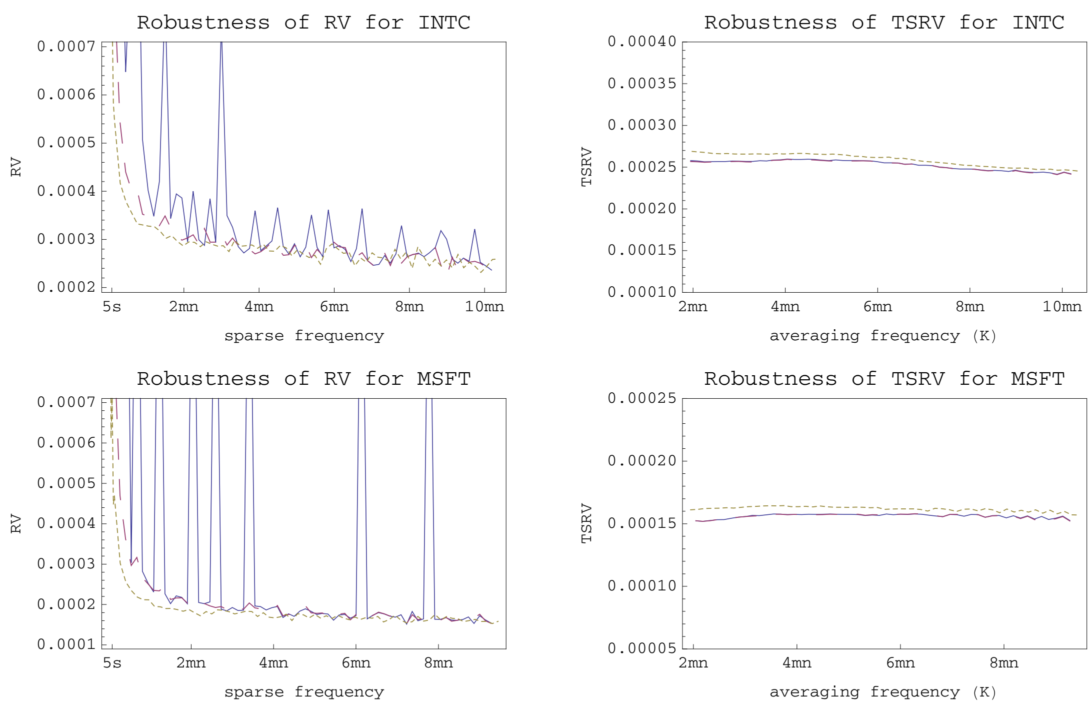


Fig. 11. Dependence of the RV and TSRV estimators for Intel and Microsoft, averaged over the last ten trading days of April 2004 on the degree of pre-processing of the raw data. In each panel, the three curves correspond respectively to the raw data (solid line), the data where immediate price bouncebacks of 1% or more are eliminated (large dashes) and the data where immediate price bouncebacks of 0.1% or more are eliminated (short dashes). In the case of TSRV, the results for the raw data and the elimination of 1% bouncebacks are virtually indistinguishable.