Lectures 1 & 2

How to estimate volatility in the presence of market microstructure noise

Yacine Aït-Sahalia
Princeton University

based on joint work with Per A. Mykland and Lan Zhang

- A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, *Journal of the American Statistical Association*, 2005, 100, 1394-1411.

- How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise, *Review of Financial Studies*, 2005, 18, 351-416.

1. Introduction

- Observed transaction price = unobservable efficient price + some noise component due to the imperfections of the trading process

$$\tilde{X}_\tau = X_\tau + U_\tau$$

- U summarizes a diverse array of market microstructure effects, either informational or not: bid-ask bounces, discreteness of price changes, 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.

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

$$dX_t = \mu_t dt + \sigma_t dW_t$$

using discretely sampled data on the transaction price process at times 0, Δ,..., $N\Delta = T$.

- We study the two situations where σ_t is parametric (which can be reduced to $\sigma_t = \sigma$, a fixed parameter to be estimated), and σ_t is nonparametric (i.e., a stochastic process).

- Without noise, the realized volatility of the process estimates the quadratic variation $T^{-1}\int_0^T \sigma_t^2 dt$.

- In theory, sampling as often as possible will produce in the limit a perfect estimate of that quantity.

- We show, however, that the situation changes radically in the presence of market microstructure noise. For instance, if noise is present but unaccounted for, then there exists a finite optimal sampling frequency:
  - A log-return over a tiny time interval Δ is mostly composed of market microstructure noise, while the volatility of the price process is proportional to Δ.
  - As Δ increases, the amount of noise remains constant, since each price is measured with error, while the informational content of volatility increases.
  - Running counter to this effect is the basic statistical principle that sampling more frequently cannot hurt.
- What we show is that these two effects compensate each other and result in a finite optimal Δ (in the RMSE sense).

- But even if sampling optimally, one is throwing away a large amount of data. We then address the question of what to do about the presence of the noise U:
  - The usual response to the presence of microstructure noise has been to reduce the sampling frequency to some arbitrary level, say every 5 or 30 minutes even though the raw data may be available every second.
  - In the parametric case, we show that modelling U explicitly through the likelihood restores the first order statistical effect that sampling as often as possible is optimal.
  - But, more surprisingly, this is true even if one misspecifies the distribution of U.

– This robustness result argues for incorporating U when estimating continuous time models with high frequency financial data, even if one is unsure about the true distribution of the noise term.

– We also study the same questions when the observations are sampled at random time intervals Δ, which are an essential empirical feature of transaction-level data.

- Then we move on to the nonparametric or stochastic volatility case:

  – We show that ignoring the noise is even worse than in the parametric case, in that the quadratic variation no longer estimates a mixture of the price volatility and the noise, but now estimates exclusively the variance of the noise.

  – We propose a solution based on subsampling and averaging, which again makes use of the full data.

2. Outline

- The Parametric Case: Constant Volatility
  - What happens if market microstructure noise is ignored
  - Correcting for the presence of the noise: likelihood

- The Nonparametric Case: Stochastic Volatility
  - What happens if the noise is ignored
  - Correcting for the presence of the noise: subsampling and averaging

3. The Baseline Case: Constant σ, No Microstructure Noise

- With $U \equiv 0$, the log-returns $Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}}$ are iid $N(0, \sigma^2\Delta)$. The MLE for σ^2 coincides with the realized volatility of the process,

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{N} Y_i^2,$$

- $T^{1/2}\left(\hat{\sigma}^2 - \sigma^2\right) \underset{T \longrightarrow \infty}{\longrightarrow} N(0, 2\sigma^4\Delta)$

- Thus selecting Δ as small as possible is optimal for the purpose of estimating σ^2.

4. When the Observations Are Noisy But the Noise is Ignored

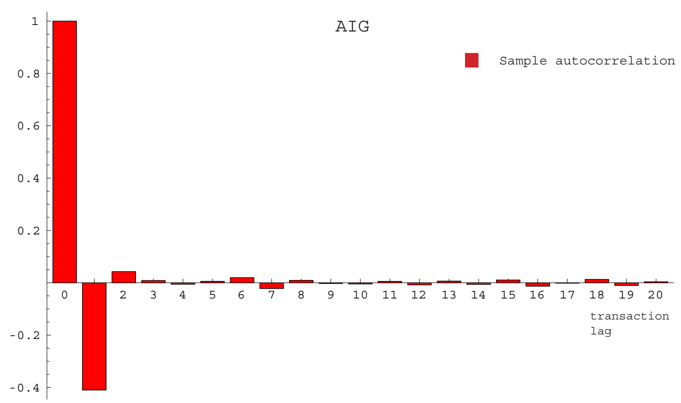
- Suppose now that market microstructure noise is present

- But the presence of the $U's$ (iid, mean 0, variance a^2) is ignored when estimating σ^2.

- In other words, we use the same likelihood as before even though the true structure of the observed log-returns Y is given by an MA(1) process since

$$Y_i = \sigma\left(W_{\tau_i} - W_{\tau_{i-1}}\right) + U_{\tau_i} - U_{\tau_{i-1}} \equiv \varepsilon_i + \eta\varepsilon_{i-1}$$

- What does the data say? Here is the autocorrelation structure for AIG, last 10 trading days in April 2004:



Theorem 1: In small samples (finite T), the bias and variance of the estimator $\hat{\sigma}^2$ are given by

$$E\left[\hat{\sigma}^2\right] - \sigma^2 = \frac{2a^2}{\Delta}$$

$$\text{Var}\left[\hat{\sigma}^2\right] = \frac{2\left(\sigma^4\Delta^2 + 4\sigma^2\Delta a^2 + 6a^4 + 2\,\text{Cum}_4\left[U\right]\right)}{T\Delta} - \frac{2\left(2a^4 + \text{Cum}_4\left[U\right]\right)}{T^2}$$

where $Cum_4(U)$ denotes the fourth cumulant of the random variable U :

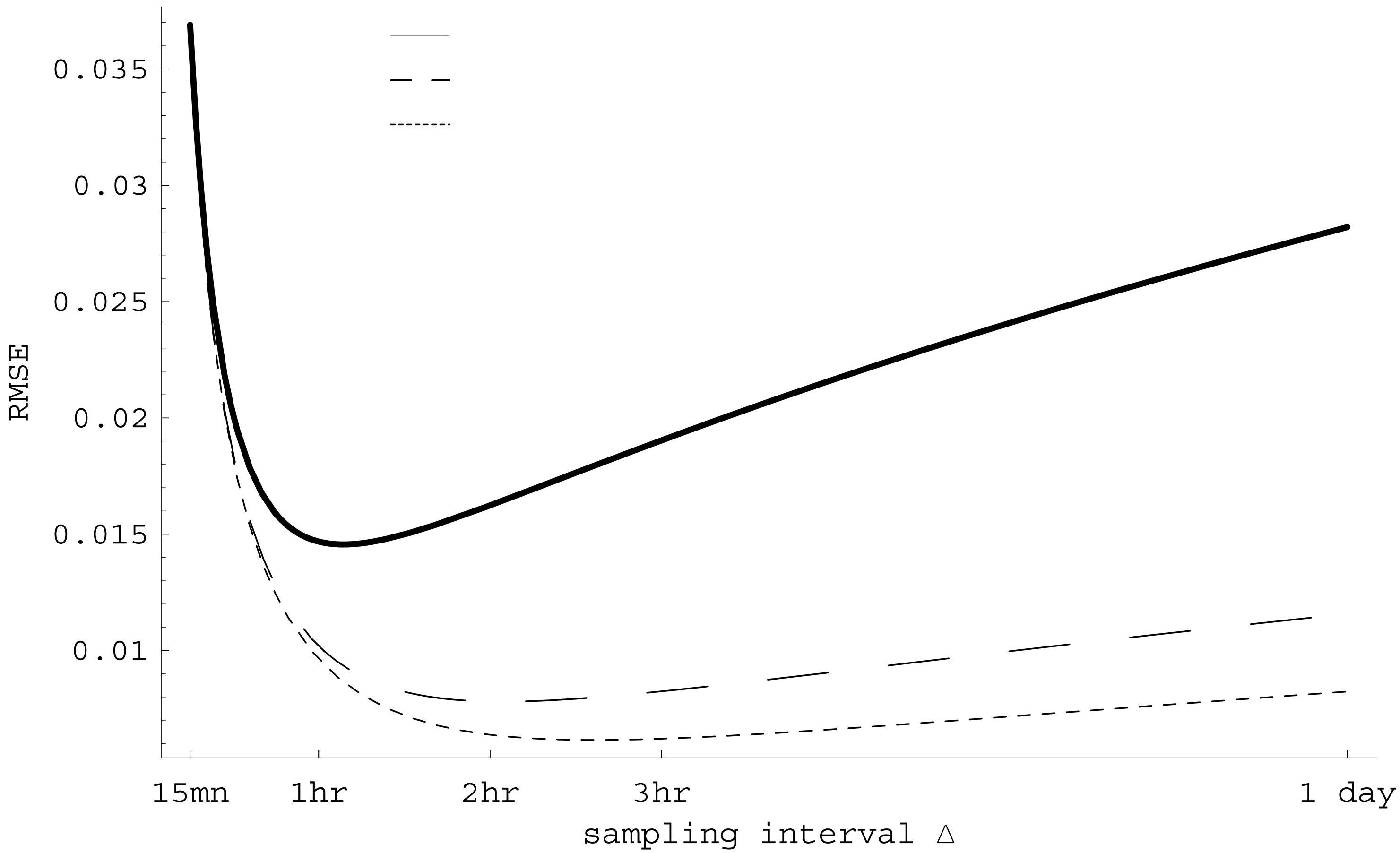
$$Cum_4(U) = E\left[U^4\right] - 3E\left[U^2\right]^2.$$

Its RMSE has a unique minimum in Δ which is reached at the optimal sampling interval:

$$\Delta^* = \left(\frac{2a^4T}{\sigma^4}\right)^{1/3} \left(\left(1 - \left(1 - \frac{2\left(3a^4 + \text{Cum}_4[U]\right)^3}{27\sigma^4 a^8 T^2}\right)^{1/2}\right)^{1/3} + \left(1 + \left(1 - \frac{2\left(3a^4 + \text{Cum}_4[U]\right)^3}{27\sigma^4 a^8 T^2}\right)^{1/2}\right)^{1/3} \right)$$

As T grows, we have

$$\Delta^* = \frac{2^{2/3}a^{4/3}}{\sigma^{4/3}}T^{1/3} + O\left(\frac{1}{T^{1/3}}\right).$$


0.035
0.03
0.025
0.02
0.015
0.01
RMSE
15mn
1hr
2hr
3hr
1 day
sampling interval Δ

Value of a	**T = 1 day**	**T = 1 year**
	Panel A: $\sigma = 30\%$	**Stocks**
0.01%	1 mn	4 mn
0.05%	5 mn	31 mn
0.1%	12 mn	1.3 hr
0.15%	22 mn	2.2 hr
0.2%	32 mn	3.3 hr
0.3%	0.9 hr	5.6 hr
0.4%	1.4 hr	1.3 day
0.5%	2 hr	1.7 day
0.6%	2.6 hr	2.2 days
0.7%	3.3 hr	2.7 days
0.8%	4.1 hr	3.2 days
0.9%	4.9 hr	3.8 days
1.0%	5.9 hr	4.3 days
	Panel B: $\sigma = 10\%$	**Currencies**
0.005%	4 mn	23 mn
0.01%	9 mn	58 mn
0.02%	23 mn	2.4 hr
0.05%	1.3 hr	8.2 hr
0.10%	3.5 hr	20.7 hr

5. Accounting for Microstructure Noise: Likelihood Corrections

- With $U \sim N(0, a^2)$ (an assumption we will relax below), the likelihood function for the $Y's$ is then given by

$$l(\eta, \gamma^2) = -\ln\det(V)/2 - N\ln(2\pi\gamma^2)/2 - (2\gamma^2)^{-1}Y'V^{-1}Y$$

$$V = \left[v_{ij}\right] = \begin{pmatrix} 1+\eta^2 & \eta & \cdots & 0 \\ \eta & 1+\eta^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \eta \\ 0 & \cdots & \eta & 1+\eta^2 \end{pmatrix}$$

Proposition 1: The MLE $(\hat{\sigma}^2, \hat{a}^2)$ is consistent and its asymptotic variance is given by

$$\mathsf{AVAR}_{\mathsf{normal}}(\hat{\sigma}^2, \hat{a}^2)$$

$$= \begin{pmatrix} 4\left(\sigma^6 \Delta \left(4a^2 + \sigma^2 \Delta\right)\right)^{1/2} + 2\sigma^4 \Delta & -\sigma^2 \Delta h \\ \bullet & \frac{\Delta}{2}\left(2a^2 + \sigma^2 \Delta\right) h \end{pmatrix}$$

with

$$h \equiv 2a^2 + \left(\sigma^2 \Delta \left(4a^2 + \sigma^2 \Delta\right)\right)^{1/2} + \sigma^2 \Delta.$$

- Since $\mathsf{AVAR}_{\mathsf{normal}}(\hat{\sigma}^2)$ is increasing in Δ, it is optimal to sample as often as possible.

- Further,

$$\text{AVAR}_{\text{normal}}(\hat{\sigma}^2) = 8\sigma^3 a \Delta^{1/2} + 2\sigma^4 \Delta + o(\Delta),$$

- Thus the loss of efficiency relative to the case where no market microstructure noise is present (and $\text{AVAR}(\hat{\sigma}^2) = 2\sigma^4\Delta$) is at order $\Delta^{1/2}$.

6. The Effect of Misspecifying the Distribution of the Microstructure Noise

- We now study the situation where one includes $U's$ into the analysis, but with a misspecified model:
  - Specifically, we consider the case where the $U's$ are assumed to be normally distributed when they are not.
  - We still suppose that the $U's$ are iid with mean zero and variance a^2.
- Since the econometrician assumes $U \sim N$, inference is still done with the Gaussian log-likelihood $l(\sigma^2, a^2)$, using the scores $\dot{l}_{\sigma^2}$ and $\dot{l}_{a^2}$ as moment functions.

- Since the expected values of $\dot{l}_{\sigma^2}$ and $\dot{l}_{a^2}$ only depend on the second order moment structure of the log-returns Y, which is unchanged by the absence of normality, the moment functions are unbiased:

$$E_{\text{true}}[\dot{l}_{\sigma^2}] = E_{\text{true}}[\dot{l}_{a^2}] = 0$$

  where "true" denotes the true distribution of the $Y's$.

- Hence the estimator $(\hat{\sigma}^2, \hat{a}^2)$ based on these moment functions remains consistent.

- The effect of misspecification therefore lies in the AVAR.

- By using the cumulants of the distribution of U, we express the AVAR in terms of deviations from normality.

- We obtain that:

Theorem 2: The estimator $(\hat{\sigma}^2, \hat{a}^2)$ is consistent and its asymptotic variance is given by

$$\mathrm{AVAR}_{\mathrm{true}}(\hat{\sigma}^2, \hat{a}^2) = \mathrm{AVAR}_{\mathrm{normal}}(\hat{\sigma}^2, \hat{a}^2) + Cum_4(U) \begin{pmatrix} 0 & 0 \\ 0 & \Delta \end{pmatrix}$$

where $\mathrm{AVAR}_{\mathrm{normal}}(\hat{\sigma}^2, \hat{a}^2)$ is the asymptotic variance in the case where the distribution of U is Normal.

- Robustness to Misspecification of the Noise Distribution

- We have shown that $\text{AVAR}_{\text{normal}}(\hat{\sigma}^2, \hat{a}^2)$ coincides with $\text{AVAR}_{\text{true}}(\hat{\sigma}^2, \hat{a}^2)$ for all but the (a^2, a^2) term.

- We show in the paper how to interpret this in terms of the profile likelihood and the second Bartlett identity.

7. Randomly Spaced Sampling Intervals

- We now relax the assumption that Δ is constant.

- Indeed, one essential feature of transaction data in finance is that the time that separates successive observations is random, or at least time-varying.

- So, we consider the case where the $\Delta_i's$ are random (for simplicity iid, independent of the W process).

- We Taylor-expand around $\Delta_0 = E[\Delta]$:

$$\Delta_i = \Delta_0 (1 + \epsilon \xi_i)$$

  - ϵ and Δ_0 are nonrandom
  - the $\xi_i' s$ are iid random variables with mean zero

- We will Taylor-expand around $\epsilon = 0$

Theorem 3: The MLE $(\hat{\sigma}^2, \hat{a}^2)$ is again consistent, this time with asymptotic variance

$$\text{AVAR}(\hat{\sigma}^2, \hat{a}^2) = A^{(0)} + \epsilon^2 A^{(2)} + O(\epsilon^3)$$

where

- $A^{(0)}$ is the asymptotic variance matrix already present in the deterministic sampling case except that it is evaluated at $\Delta_0 = E[\Delta]$.

- The second order correction term $A^{(2)}$ is proportional to $Var[\xi]$ (recall $\Delta_i = \Delta_0 (1 + \epsilon \xi_i)$) and is therefore zero in the absence of sampling randomness.

8. Extensions

8.1. Presence of a Drift Coefficient

- The presence of a drift does not alter our earlier conclusions. Suppose that

$$X_t = \mu t + \sigma W_t.$$

- We show that

$$\text{AVAR}(\hat{\sigma}^2, \hat{a}^2) = E[\Delta] D^{-1}_{\sigma^2,a^2} S_{\sigma^2,a^2} D^{-1}_{\sigma^2,a^2}.$$

is thus the same as if μ were known, in other words, as if $\mu = 0$, which is the case we focused on.

8.2. Serially Correlated Noise

- Suppose that, instead of being iid, the market microstructure noise follows

$$dU_t = -bU_t dt + c dZ_t$$

- This type of noise could capture the gradual adjustment of prices in response to a shock such as a large trade, or other information effects, while iid noise captures the frictions (random bid-ask bounces caused by noise traders, rounding errors, etc.)

- Now the variance contributed by the noise is of order $O(\Delta)$, that is of the same order as the variance of the efficient price process $\sigma^2\Delta$, instead of being of order $O(1)$ when the noise is iid.

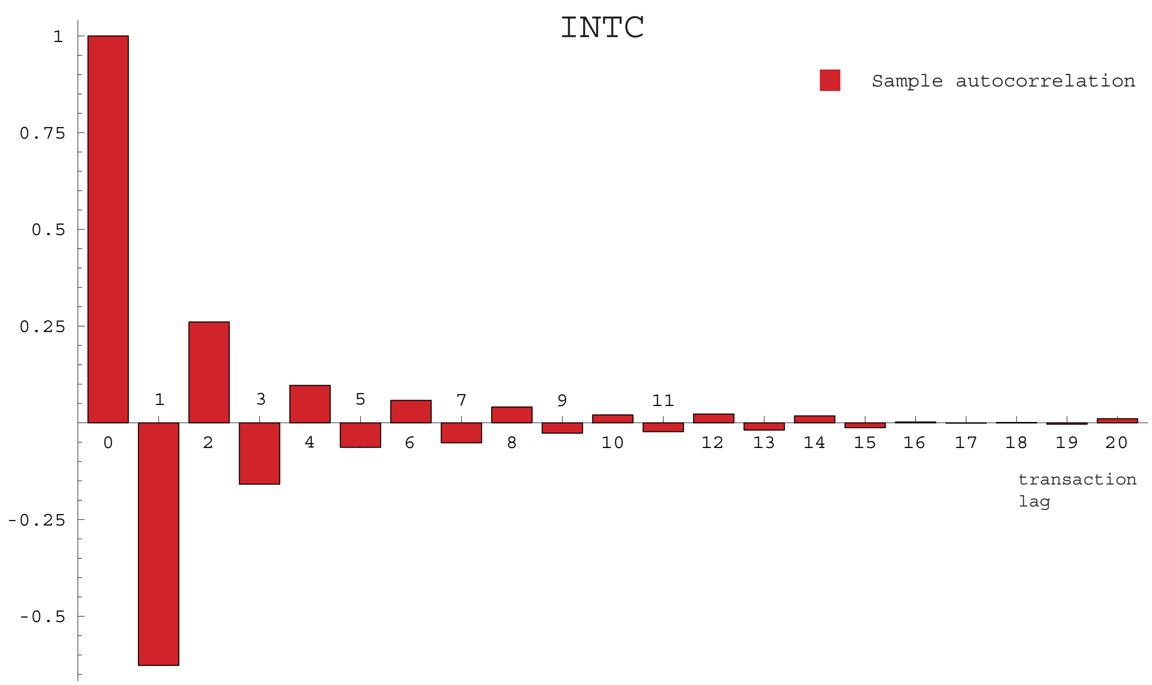
- If one ignores the presence of this type of serially correlated noise when estimating σ^2, then:

$$RMSE\left[\hat{\sigma}^2\right] = c^2 - \frac{bc^2}{2}\Delta + \frac{\left(\sigma^2 + c^2\right)^2 \Delta}{c^2 T} + O(\Delta^2) + O\left(\frac{1}{T^2}\right)$$

- So that for large T increasing Δ first reduces $RMSE\left[\hat{\sigma}^2\right]$.

- Hence the optimal sampling frequency is again finite. But this type of noise is not nearly as bad as iid noise for the purpose of inferring σ^2.

- In the data, both types of noise are typically present together: for example, here is the autocorrelation structure for INTC, last 10 trading days in April 2004:



8.3. Noise Correlated with the Price Process

- We have assumed so far that the U process was uncorrelated with the W process.

- But microstructure noise attributable to informational effects is likely to be correlated with the efficient price process, since it is generated by the response of market participants to information signals.

- The form of the variance matrix of the observed log-returns Y must be altered, replacing $\gamma^2 v_{ij}$ with

$$\begin{aligned} cov(Y_i, Y_j) &= cov(\sigma\left(W_{\tau_i} - W_{\tau_{i-1}}\right) + U_{\tau_i} - U_{\tau_{i-1}}, \sigma\left(W_{\tau_j} - W_{\tau_{j-1}}\right) + U_{\tau_j} - U_{\tau_{j-1}}) \\ &= \sigma^2 \Delta \delta_{ij} + cov(\sigma\left(W_{\tau_i} - W_{\tau_{i-1}}\right), U_{\tau_j} - U_{\tau_{j-1}}) \\ &\quad + cov(\sigma\left(W_{\tau_j} - W_{\tau_{j-1}}\right), U_{\tau_i} - U_{\tau_{i-1}}) + cov(U_{\tau_i} - U_{\tau_{i-1}}, U_{\tau_j} - U_{\tau_{j-1}}) \end{aligned}$$

9. The Nonparametric Case: Stochastic Volatility

- When $dX_t = \sigma_t dW_t$, the object of interest is now the quadratic variation
$$\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$$
over a fixed time period $[0, T]$, or possibly several such time periods.

- The estimation is based on observations $0 = t_0 < t_1 < ... < t_n = T$, and asymptotic results are obtained when $\max \Delta t_i \to 0$.

- The usual estimator of $\langle X, X \rangle_T$ is the realized volatility

$$[X, X]_T = \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2.$$

- The sum converges to the integral, with a known distribution: Jacod (1994).

- In the context of stochastic volatility, however, ignoring market microstructure noise leads to an even more dangerous situation than when σ is constant and $T \longrightarrow \infty$.
  - After suitable scaling, the realized volatility is a consistent and asymptotically normal estimator – but of the quantity $2nE[U^2]$.
  - This quantity has nothing to do with the object of interest, $\langle X, X \rangle_T$.
- Said differently, market microstructure noise totally swamps the variance of the price signal at the level of the realized volatility.

9.1. The Fifth Best Approach

- Ignoring Market Microstructure Noise when Volatility is Stochastic

- We show that, if one uses all the data (say sampled every second),

$$[Y,Y]_T^{(all)} \overset{\mathcal{L}}{\approx} \underbrace{\langle X,X\rangle_T}_{\text{object of interest}} + \underbrace{2nE[\varepsilon^2]}_{\text{bias due to noise}} + \underbrace{[\underbrace{4nE[\varepsilon^4]}_{\text{due to noise}} + \underbrace{\frac{2T}{n}\int_0^T \sigma_t^4 dt}_{\text{due to discretization}}}_{\text{total variance}}]^{1/2} Z_{\text{total}}.$$

  conditionally on the X process.

- $[Y,Y]_T^{(all)}$ has a positive bias whose magnitude increases linearly with the sample size n.

- The magnitude of the bias is $O(n)$, that of the object of interest $\langle X, X \rangle_T$ is $O(1)$.

- So the bias dwarfs everything else.

9.2. The Fourth Best Approach

- Sampling Sparsely at an Arbitrary Frequency

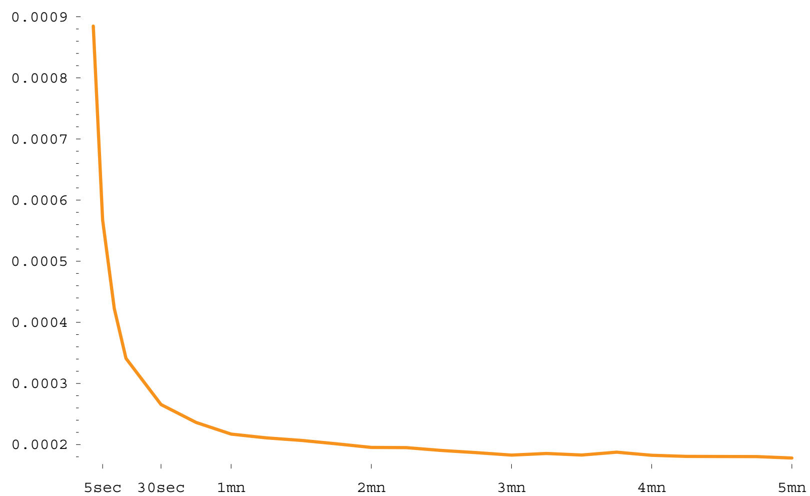
- Of course, completely ignoring the noise and sampling as prescribed by $[Y, Y]_T^{(all)}$ is not what empirical researchers do in practice.

- They use the estimator $[Y, Y]_T^{(sparse)}$ constructed from 5 mn returns, say.

- For example, if $T = 1$ day and transactions occur every $\delta = 1$ second, then the original sample size is $n = T/\delta = 23,400$.

- But sampling sparsely once every 5 mn means throwing out 299 out of every 300 observations, and the sample size used is only $n_{sparse} = 78$.

- Here is the fourth best estimator for different values of Δ, averaged for the 30 DJIA stocks, last 10 trading days in April 2004:



- As $\Delta = T/n \to 0$, the graph shows that the estimator diverges as predicted by our result ($2nE[\varepsilon^2]$) instead of converging to the object of interest $\langle X, X \rangle_T$ as predicted by standard asymptotic theory.

9.3. The Third Best Approach

- Sampling Sparsely at an Optimal Frequency

- If one insists upon sampling sparsely, what is the right answer? Is it 5 mn, 10 mn, 15 mn?

- As in the parametric case, if one insists upon sampling sparsely, we show how to determine optimally the sparse sampling frequency:

$$n_{sparse}^{*} = \left(\frac{T}{4 \ E[U^2]^2} \int_0^T \sigma_t^4 dt \right)^{1/3}.$$

- Using every $K^* \approx n/n_{sparse}^*$ observation gives rise to the estimator we define as $[Y,Y]_T^{(sparse,opt)}$.

9.4. The Second Best Approach

- Correcting for Microstructure Noise when Volatility is Stochastic

- We have just argued that one could benefit from using infrequently sampled data.

- And yet, one of the most basic lessons of statistics is that one should not do this.

- We present a method to tackle the problem:
  - We 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.
  - To the extent that there is a benefit to subsampling, this benefit can now be retained, while the variation of the estimator can be lessened by the averaging.

- This gives rise to the estimator

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

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

- We show that:

$$[Y,Y]_T^{(avg)} \overset{\mathcal{L}}{\approx} \underbrace{\langle X,X\rangle_T}_{\text{object of interest}} + \underbrace{2\bar{n}E[U^2]}_{\text{bias due to noise}} + \underbrace{[\underbrace{4\frac{\bar{n}}{K}E[U^4]}_{\text{due to noise}} + \underbrace{\frac{4T}{3\bar{n}}\int_0^T \sigma_t^4\,dt}_{\text{due to discretization}}}_{\text{total variance}}]^{1/2} Z_{\text{total}}$$

- So, $[Y,Y]_T^{(avg)}$ remains a biased estimator of the quadratic variation $\langle X,X\rangle_T$ of the true return process.

- But the bias $2\bar{n}E[U^2]$ now increases with the average size of the subsamples, and $\bar{n} \leq n$.

- Thus, $[Y,Y]_T^{(avg)}$ is a better estimator than $[Y,Y]_T^{(all)}$.

- The optimal trade-off between the bias and variance for the estimator $[Y,Y]_T^{(avg)}$ consists in setting $K^* \approx n/\bar{n}^*$ subsamples with

$$\bar{n}^* = \left(\frac{T}{6E[U^2]^2}\int_0^T \sigma_t^4 dt\right)^{1/3}.$$

9.5. The First Best Approach

- Two Scales Realized Volatility

- The bias of $[Y,Y]_T^{(avg)}$ is $2\bar{n}E[U^2]$.

- Recall that $E[U^2]$ can be consistently approximated using the fifth best estimator

$$\widehat{E[U^2]} = \frac{1}{2n}[Y,Y]_T^{(all)}$$

- Hence the bias of $[Y,Y]^{(avg)}$ can be consistently estimated by $\frac{\bar{n}}{n}[Y,Y]_T^{(all)}$.

A bias-adjusted estimator for $\langle X, X \rangle$ can thus be constructed as

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

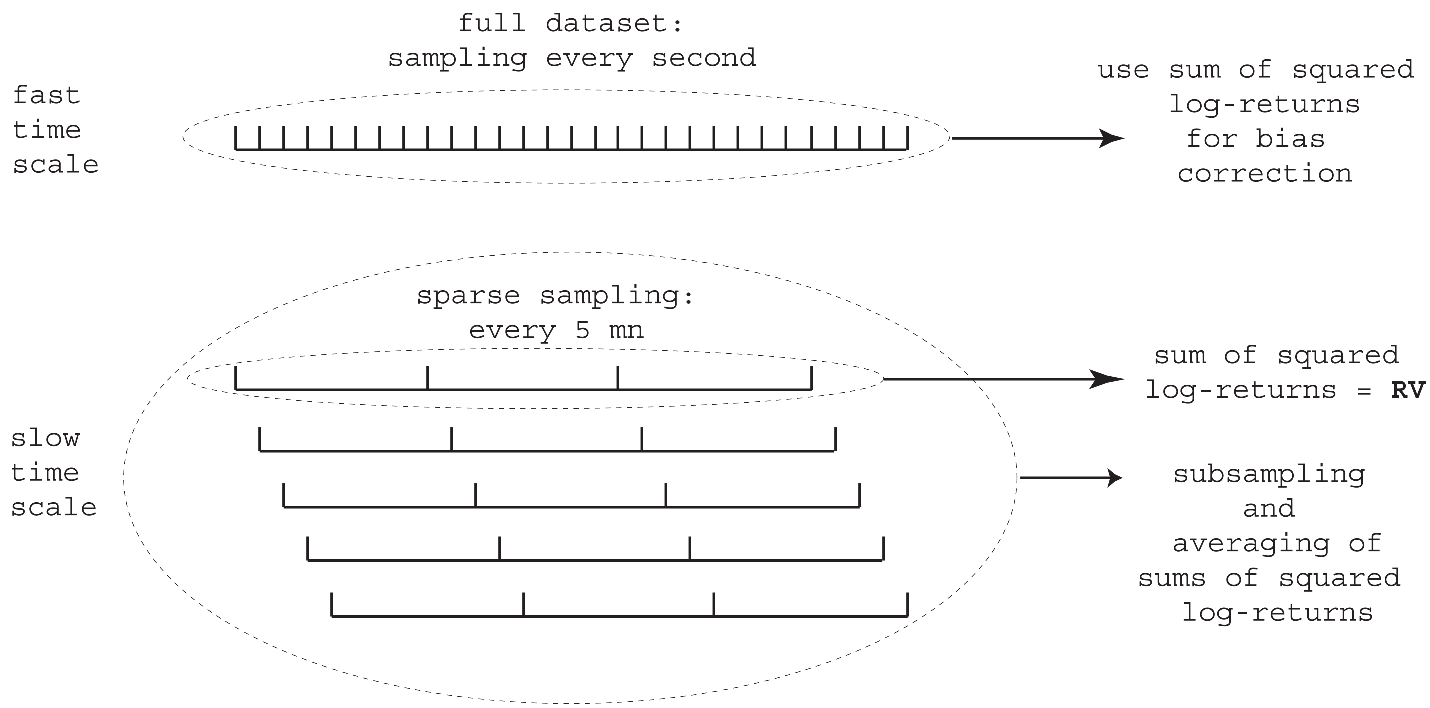
- We show that if the number of subsamples is selected as $K = cn^{2/3}$, then

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

- Unlike all the previously considered ones, this estimator is now correctly centered

- It only converges at rate $n^{-1/6}$ but it's better than being (badly) biased, and now there is no limit to how often one should sample (every second? so n can be quite large, e.g., $n = 23,400$ vs. $n_{sparse} = 78$).

9.6. Summary: TSRV Construction

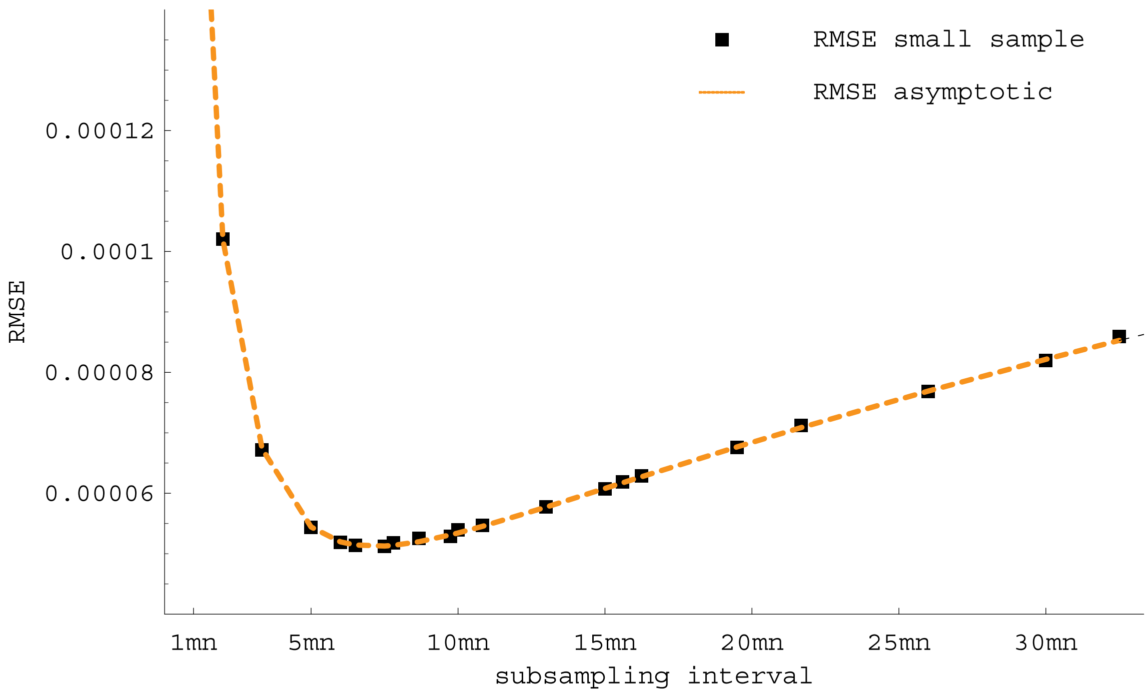


TSRV = subsampling and averaging, then
bias-correcting using ultra high frequency data

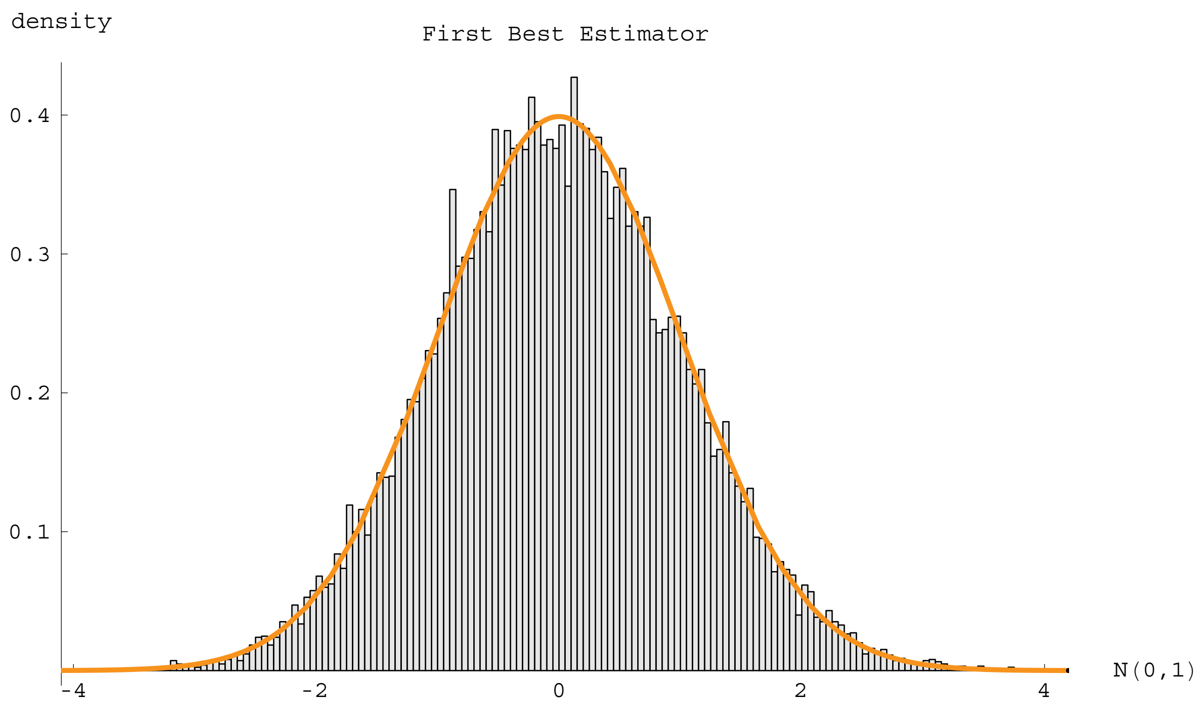
9.7. Monte Carlo Simulations

	Fifth Best $[Y,Y]_T^{(all)}$	RV Fourth Best $[Y,Y]_T^{(sparse)}$	Third Best $[Y,Y]_T^{(sparse,opt)}$	Second Best $[Y,Y]_T^{(avg)}$	TSRV First Best $\widehat{\langle X,X\rangle}_T^{(adj)}$
Small Sample Bias	$1.1699\ 10^{-2}$	$3.89\ 10^{-5}$	$2.18\ 10^{-5}$	$1.926\ 10^{-5}$	$2\ 10^{-8}$
Asymptotic Bias	$1.1700\ 10^{-2}$	$3.90\ 10^{-5}$	$2.20\ 10^{-5}$	$1.927\ 10^{-5}$	0
Small Sample Variance	$1.791\ 10^{-8}$	$1.4414\ 10^{-9}$	$1.59\ 10^{-9}$	$9.41\ 10^{-10}$	$9\ 10^{-11}$
Asymptotic Variance	$1.788\ 10^{-8}$	$1.4409\ 10^{-9}$	$1.58\ 10^{-9}$	$9.37\ 10^{-10}$	$8\ 10^{-11}$
Small Sample RMSE	$1.1699\ 10^{-2}$	$5.437\ 10^{-5}$	$4.543\ 10^{-5}$	$3.622\ 10^{-5}$	$9.4\ 10^{-6}$
Asymptotic RMSE	$1.1700\ 10^{-2}$	$5.442\ 10^{-5}$	$4.546\ 10^{-5}$	$3.618\ 10^{-5}$	$8.9\ 10^{-6}$
Small Sample Relative Bias	182	0.61	0.18	0.15	−0.00045
Small Sample Relative Variance	82502	1.15	0.11	0.053	0.0043
Small Sample Relative RMSE	340	1.24	0.37	0.28	0.065

- RMSE determination of the third best estimator

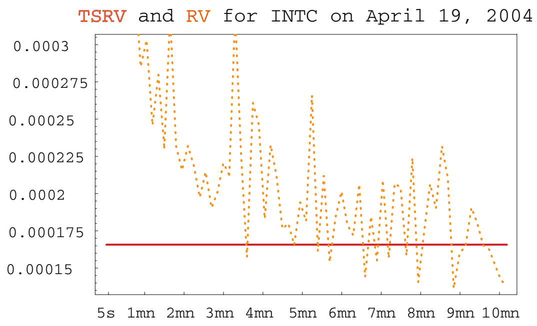


- Standardized distribution of the TSRV (first best) estimator: simulations vs. theory

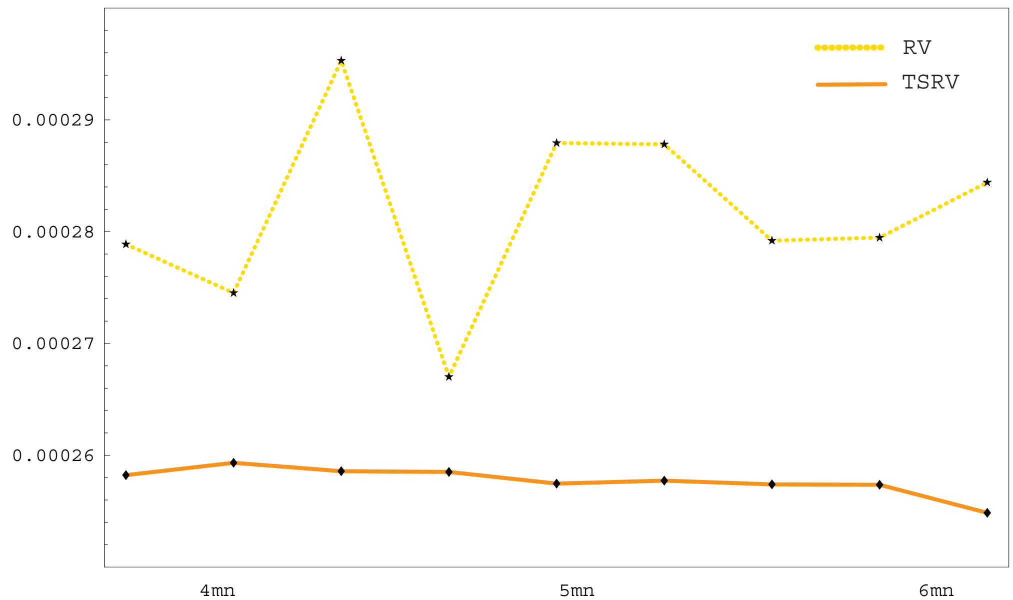


10. Data Analysis

- Here is a comparison of RV to TSRV for INTC, last 10 trading days in April 2004:



- Zooming around the 5 minutes sampling frequency:



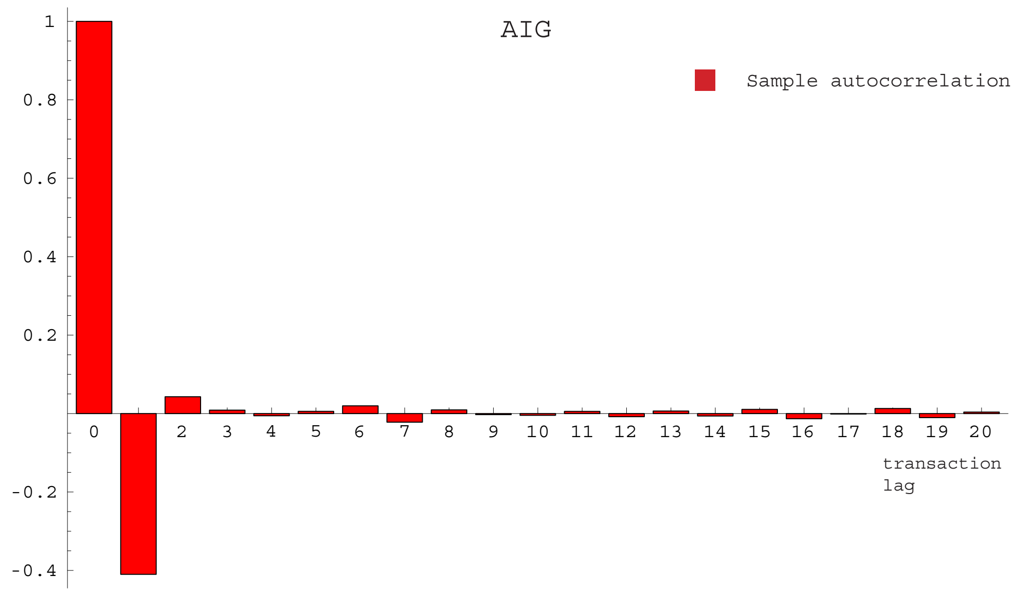
11. Dependent Market Microstructure Noise

- So far, we have assumed that the noise ε was iid.

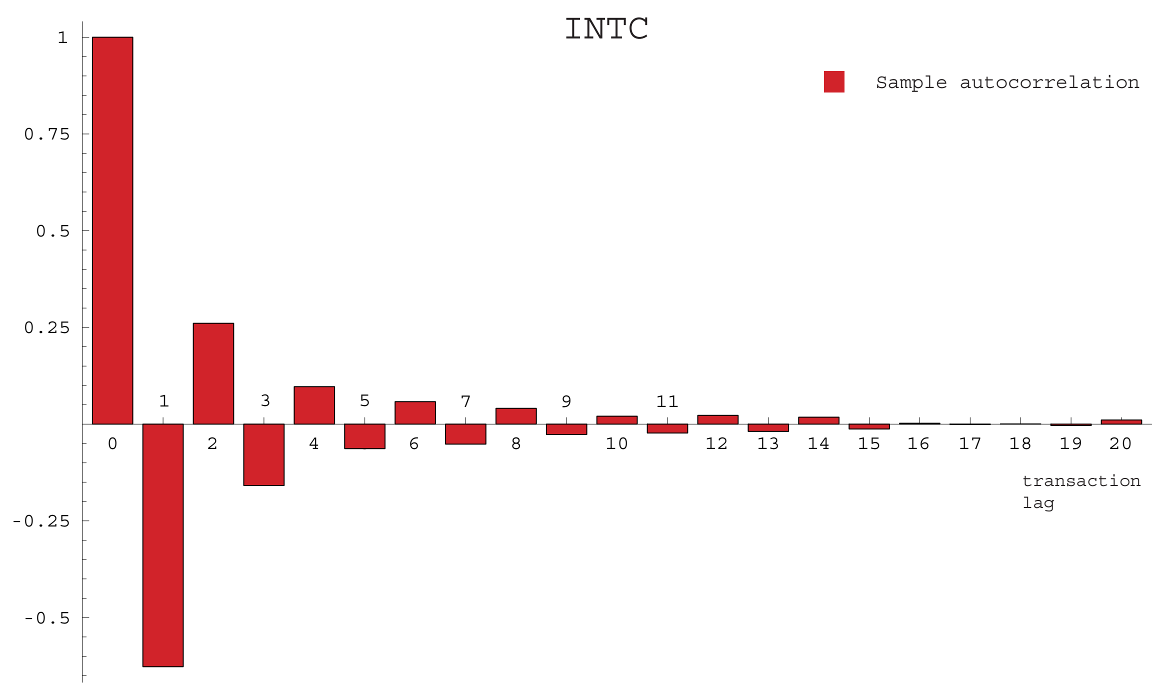
- In that case, log-returns are MA(1):

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

- For example, here is the autocorrelogram for AIG transactions, last 10 trading days in April 2004:



- But here is the autocorrelogram for INTC transactions, same last 10 trading days in April 2004:

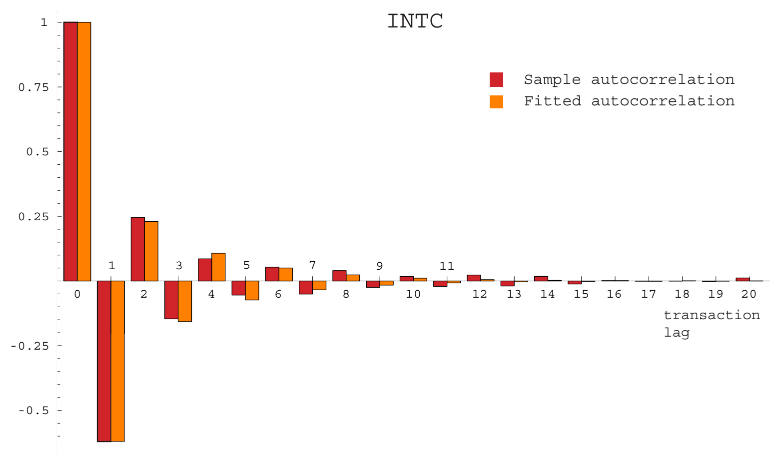


- A simple model to capture this higher order dependence is

$$\varepsilon_{t_i} = U_{t_i} + V_{t_i}$$

where U is iid, V is $AR(1)$ and $U \perp V$.

- Fitted autocorrelogram for INTC:



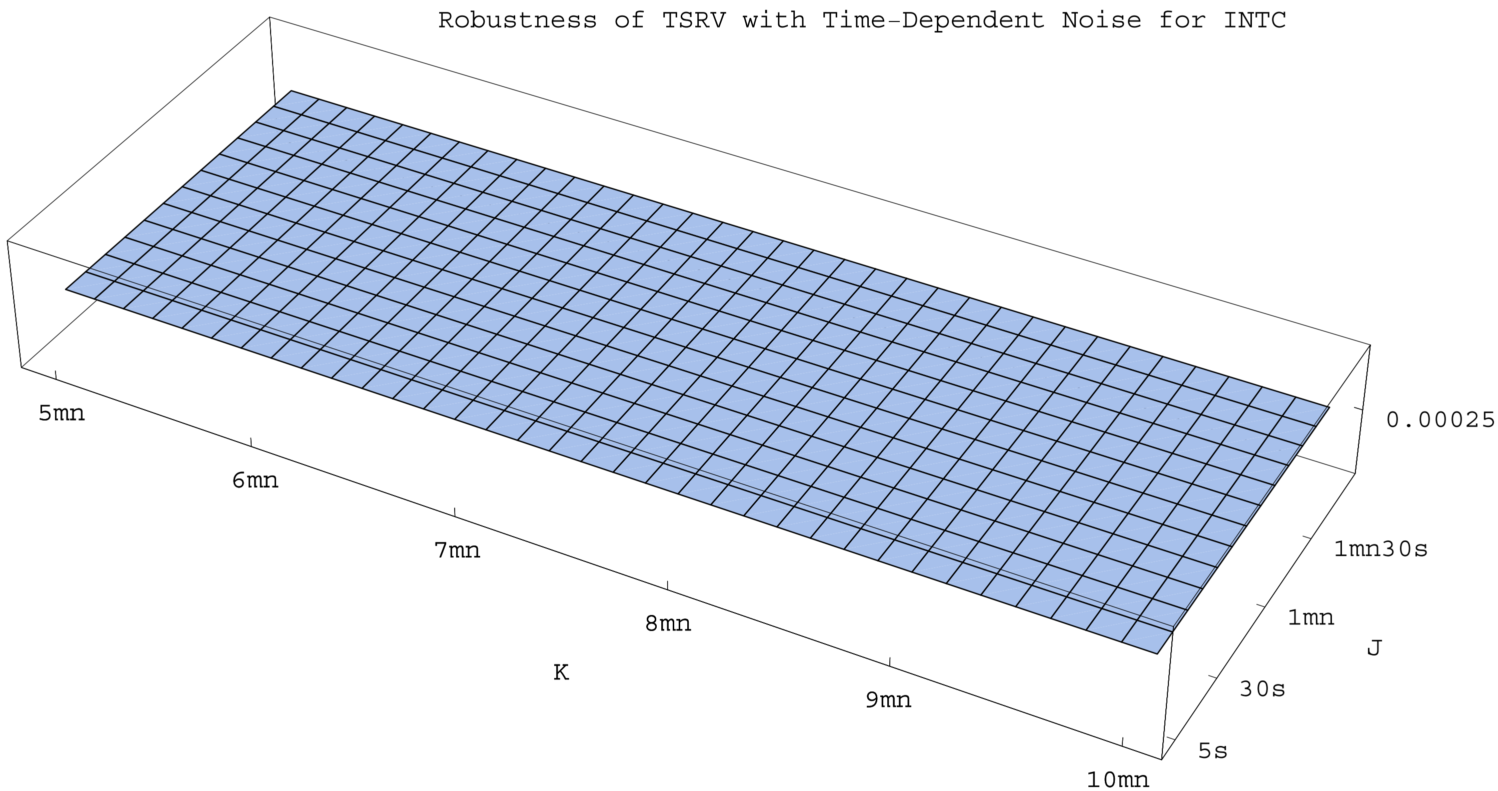
- The TSRV Estimator with (J, K) Time Scales

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

- We show that if we select $J/K \to 0$ when $n \to \infty$, then this estimator is robust to (essentially) arbitrary time series dependence in microstructure noise.

- Specifically, we let the noise process ε_{t_i} be stationary and strong mixing with exponential decay. We also suppose that $E\left[\varepsilon^{4+\kappa}\right] < \infty$ for some $\kappa > 0$.

- Robustness to the selection of the slow (K) and fast (J) time scales, INTC again:



12. Further Refinement: MSRV

- 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 robust to arbitrary serial dependence in market microstructure noise;
  - 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 (Zhang, 2006).

- The resulting estimator, MSRV has the form of

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

- TSRV corresponds to the special case where $M = 1$, i.e., where one uses a single slow time scale in conjunction with the fast time scale to bias-correct it.

- For suitably selected weights a_i and $M = O(n^{1/2})$, $\widehat{\langle X, X \rangle}_T^{(\text{msrv})}$ converges to the $\langle X, X \rangle_T$ at rate $n^{-1/4}$.
  - Weights are of the form $a_i = \frac{i}{M^2} h\left(\frac{i}{M}\right) - \frac{i}{2M^3} h'\left(\frac{i}{M}\right)$, where h is a continuously differentiable real-valued function.
  - The optimal choice of h is

$$h^*(x) = 12\left(x - \frac{1}{2}\right)$$

- When computed the optimal weights MSRV estimator has the following distribution:

$$\widehat{\langle X, X \rangle}_T^{(\text{msrv})} \overset{\mathcal{L}}{\approx} \langle X, X \rangle_T + \frac{1}{n^{1/4}} \Big[\underbrace{\Upsilon}_{\text{total variance}}\Big]^{1/2} Z_{\text{total}}$$

13. Conclusions

- The Parametric Case: Constant Volatility
  - In the presence of market microstructure noise that is unaccounted for, it is optimal to sample less often than would otherwise be the case: we derive the optimal sampling frequency.
  - A better solution, however, is to model the noise term explicitly, for example by likelihood methods, which restores the first order statistical effect that sampling as often as possible is optimal.
  - But, more surprisingly, we also demonstrate that the likelihood correction is robust to misspecification of the assumed distribution of the noise term.

- The Nonparametric Case: Stochastic Volatility

  - Matters are much worse: realized volatility estimates pure noise

  - But it is possible to correct for the noise by subsampling and averaging and obtain well behaved estimators that make use of all the data

- These results suggest that attempts to incorporate market microstructure noise when estimating continuous-time models based on high frequency data should have beneficial effects.

- And one final important message of the two papers:

- Any time one has an impulse to sample sparsely, one can always do better: for example, using likelihood corrections in the parametric case or subsampling and averaging in the nonparametric case.

- No matter what the model is, no matter what quantity is being estimated.