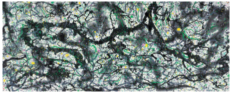


High Frequency Volatility Estimation

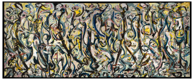
with Dependent Microstructure Noise

Zhou Fang **Nate Hattersley**

Department of Mathematics/Department of Economics
The University of Texas at Austin

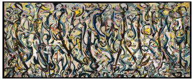
April 11, 2023

Overview



1. **Motivation**
2. **Independent Microstructure Noise**
3. **Dependent Microstructure Noise**
4. **MSRV**
5. **Empirics**

Volatility Estimation



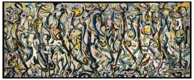
S_t is one asset's price at time t, which satisfies the following dynamics

$$X_t = \log S_t$$
$$dX_t = \mu_t dt + \sigma_t dW_t$$

The quantity to estimate is

$$\int_0^T \sigma_t^2 dt$$

Volatility Estimation



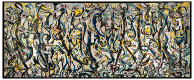
Assume there are n observations of log prices every $\Delta t = \frac{T}{n}$ time, denote as $\{X_{t_i}\}_{i=0}^{n}$,

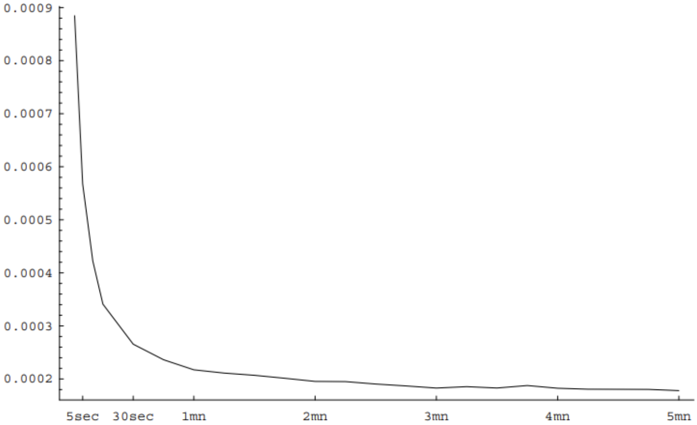
$$\lim_{n\to\infty} \sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2 \xrightarrow{p} \int_0^T \sigma_t^2 dt$$

The real world doesn't work in this way, otherwise, life is too easy.
What we observe is the "true" log prices + noises

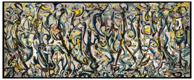
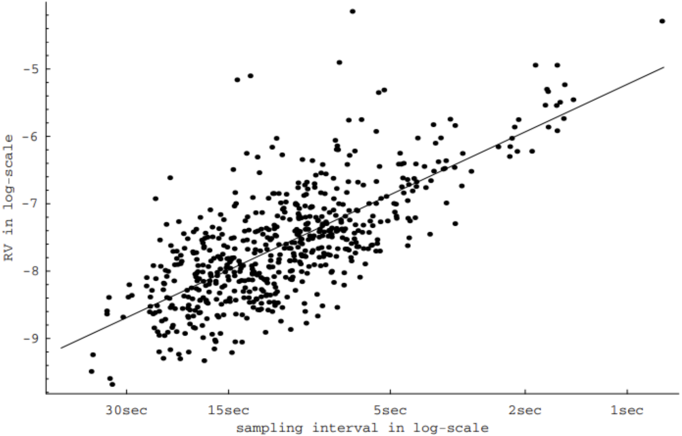
$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}$$

Realized Volatility vs Sampling Interval

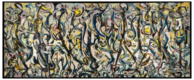




Realized Volatility vs Sampling Interval

Quadratic Estimator



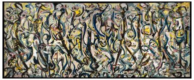
This is under the assumption that for each observation, the noises $\{\epsilon_{t_i}\}_{i=0}^{n}$ are iid.
Define the quadratic estimator as $\langle Y, Y\rangle_T^{(all)} = \sum_{t_i}(Y_{t_{i+1}} - Y_{t_i})^2$

bias of quadratic estimator

$$\frac{1}{\sqrt{n}}\left(\langle Y, Y\rangle_T^{(all)} - 2nE[\epsilon^2]\right) \xrightarrow{\mathcal{L}} 2\sqrt{(E[\epsilon^4])}\mathcal{N}(0,1)$$

Quadratic Estimator is not ideal !!!

Sparse Estimator

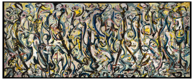


Denote the sparse estimator as $\langle Y, Y\rangle_T^{(sp)}$, and $n_{(sp)}$ is the number of observations taken, $n_{(sp)} \ll n$

bias of sparse estimator

$$\langle Y, Y\rangle_T^{(sp)} \overset{\mathcal{L}}{\approx} \int_0^T \sigma_t^2 dt + 2n_{(sp)}E[\epsilon^2] + \left[4n_{(sp)}E[\epsilon^4] + \frac{2T}{n_{(sp)}} \int_0^T \sigma_t^4 dt\right]^{1/2} \mathcal{N}(0,1)$$

Average Estimator



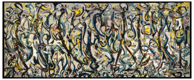
K sparse estimators, and $\bar{n}$ is the average number of observations. Averaging K sparse estimators gives an average estimator $\langle Y, Y \rangle_t^{(avg)}$

bias of average estimator

$$\langle Y, Y \rangle_T^{(avg)} \overset{\mathcal{L}}{\approx} \int_0^T \sigma_t^2 dt + 2\bar{n}E[\epsilon^2] + \left[4\frac{\bar{n}}{K}E[\epsilon^4] + \frac{4T}{3\bar{n}} \int_0^T \sigma_t^4 dt\right]^{1/2} \mathcal{N}(0,1)$$

Still Biased, So annoying

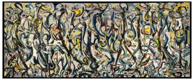
Two Scales Estimator



the legendary two scales realized volatility estimator (TSRV)

$$\widehat{\langle X, X \rangle}_T = \langle Y, Y \rangle_T^{(avg)} - \frac{\bar{n}}{n} \langle Y, Y \rangle_T^{(all)}$$

Implications of independent noises



How to check if the microstructure noises are independent or not? The following equation will give one criterion

$$E[(Y_{t_j} - Y_{t_{j-1}})(Y_{t_i} - Y_{t_{i-1}})] = \begin{cases} \int_{t_{i-1}}^{t_i} \sigma_t^2 dt + 2E[\epsilon^2] & \text{if } j = i \\ -E[\epsilon^2] & \text{if } j = i+1 \\ 0 & \text{if } j > i+1 \end{cases}$$

Autocorrelogram

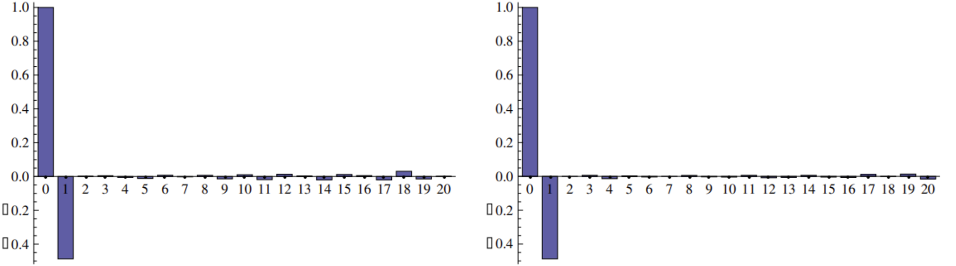
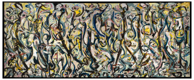


Fig. 4. Log-return autocorrelogram from transactions on the stocks of American International Group, Inc. (trading symbol: AIG) and 3M Co. (trading symbol: MMM), last ten trading days in April 2004.

Autocorrelogram



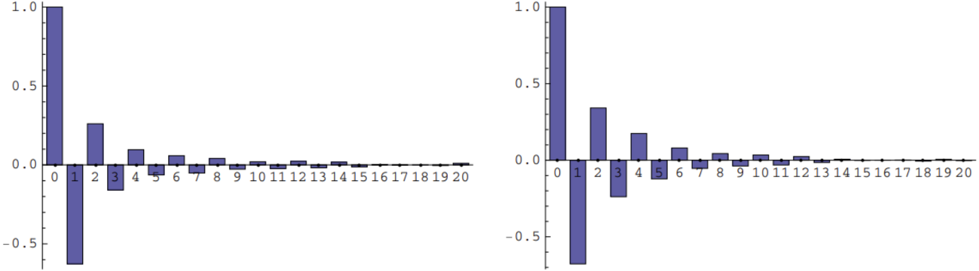
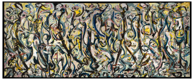


Fig. 5. Log-return autocorrelogram from transactions for Intel (trading symbol: INTC) and Microsoft (trading symbol: MSFT), last ten trading days in April 2004.

Dependent Noises



Assumption:
noise process ϵ_t is independent of X_t, stationary, and strong mixing with the mixing coefficients decaying exponentially. Together with some $\kappa > 0$, $E[\epsilon^{4+\kappa}] < \infty$

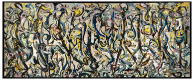
No one cares!

Assumption's Implication

There is a constant $\rho < 1$ so that for all i,

$$|\text{Cov}(\epsilon_{t_i}, \epsilon_{t_{i+l}})| \leq \rho^l \text{Var}(\epsilon)$$

TSRV



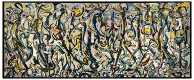
Define lag sparse estimator as

$$\langle Y, Y\rangle_T^{(J, r)} = \sum_{0 \le j(i-1) \le n-r-j} \left(Y_{t_{ji+r}} - Y_{t_{j(i-1)+r}}\right)^2$$

Define average lag sparse estimator as

$$\begin{aligned}\langle Y, Y\rangle_T^{(J)} &= \frac{1}{J}\sum_{r=0}^{J-1} \langle Y, Y\rangle_T^{(J, r)} \\ &= \frac{1}{J}\sum_{i=0}^{n-J} (Y_{t_{i+J}} - Y_{t_i})^2\end{aligned}$$

TSRV



Define a generalized version of TSRV

$$\widehat{\langle X, X\rangle}_T^{(tsrv)} = \langle Y, Y\rangle_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}\langle Y, Y\rangle_T^{(J)}$$

where $\bar{n}_K = \frac{n-K+1}{K}$, and $\bar{n}_J = \frac{n-J+1}{J}$

very important lemma

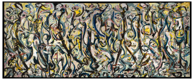
Under the noise dependence assumption, and $n \to \infty$

$$\sum_{i=0}^{n-J}(X_{t_{i+J}} - X_{t_i})(\epsilon_{t_{i+J}} - \epsilon_{t_i}) = \mathcal{O}(\sqrt{J})$$

which results in the following decomposition

$$\langle Y, Y\rangle_T^{(J)} = \langle X, X\rangle_T^{(J)} + \langle \epsilon, \epsilon\rangle_T^{(J)} + \mathcal{O}(\frac{1}{\sqrt{J}})$$

Decomposition of TSRV estimator



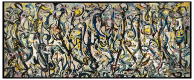
TSRV decomposition

According to the above lemma, for $1 \leq J \leq K$, and $K = o(n)$

$$\widehat{\langle X, X \rangle}_T^{(tsrv)} = \left[\langle X, X \rangle_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} \langle X, X \rangle_T^{(J)} \right] + \left[\langle \epsilon, \epsilon \rangle_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} \langle \epsilon, \epsilon \rangle_T^{(J)} \right] + \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$$

The first term is the signal term, and the second term is the noise term

Decomposing the Error Term



Consider the noise term:

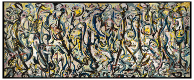
$$\left[\langle \epsilon, \epsilon \rangle_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} \langle \epsilon, \epsilon \rangle_T^{(J)}\right]$$

Under the iid noise, this is equal to zero and the estimator

$$\widehat{\langle X, X \rangle}_T^{(tsrv)} = \langle Y, Y \rangle_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J} \langle Y, Y \rangle_T^{(J)}$$

is consistent. What about the more general case?

Limiting Distribution of Noise Term



Assumption 1

Sequences $\{J_n\}_{n=1}^{\infty}$, $\{K_n\}_{n=1}^{\infty}$ satisfy $\limsup_{n\to\infty} \frac{J_n}{K_n} < 1$. This is not a restrictive assumption, and is satisfied when $1 \leq J < K,\ K = o(n)$.

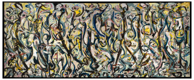
Proposition 1

Under Assumption 1,

$$\frac{K}{n^{1/2}} \left(\text{noise} - E[\text{noise}]\right) \xrightarrow{\mathcal{L}} \xi Z$$

With $E\left|\text{noise}\right|$ bounded above by a $o\left(\frac{n}{K}(\rho^K + \rho^J)\right)$ term. Explicit expression for ξ given in the paper

Limiting Distribution of Signal Term



Proposition 2

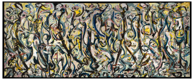
Assuming $1 \leq J \leq K$ and $K = o(n)$,

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

Where η has a limiting distribution independent of Z that is given in another paper of theirs.

Notice how convergence rate changes with J fixed vs $J \to \infty$. Pay a price for accounting for *too much* serial dependence.

The Estimator

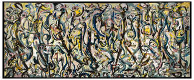


To account for this additional bias from serial dependence, the authors will suggest adjusting the original TSRV estimator in one of two ways.

First, an adjustment to the TSRV estimator that was a SSC in the original paper but here provides consistency under dependence:

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

The Estimator



To account for this additional bias from serial dependence, the authors will suggest adjusting the original TSRV estimator in one of two ways.

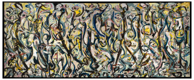
First, an adjustment to the TSRV estimator that was a SSC in the original paper but here provides consistency under dependence:

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

They suggest that optimal choice is $K = \mathcal{O}(n^{2/3})$ and to pick J such that $\text{Cov}(\epsilon_{t_0}, \epsilon_{t_J}) = o(n^{-1/2})$. In this case,

$$\widehat{\langle X, X \rangle}_T^{(tsrv,adj)} = \langle X, X \rangle_T + \left(2E\epsilon^2 \frac{n}{K} \text{Cov}(\epsilon_{t_0}, \epsilon_{t_J}) + \frac{n^{1/2}}{K} \xi Z_1 + \left(\frac{K}{n}\right)^{1/2} \eta \sqrt{T} Z_2\right)(1 + o_p(1))$$

Adjusting for Bias

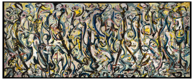


When K is large, this may be a slight underestimate. The second adjustment is an "area-adjusted" bias correction:

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

Proposition 4: these two estimators have the same asymptotic behavior

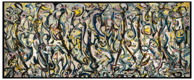
Which to Prefer?



Authors recommend the second,

> *especially for moderate sample size. It should be emphasized, however, that the bias-calculation is based on an assumption of a constant σ and on borrowing information from the middle of the interval $[0, T]$.*

The More the Merrier, Right?

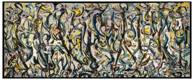


Consider a weighted sum of estimators at $K_1, K_2, \ldots, K_M$ different time scales:

$$\widehat{\langle X, X\rangle}_T^{(msrv)} = \sum_{i=1}^{M} a_i \langle Y, Y\rangle_T^{K_i} + 2\widehat{E\epsilon^2}$$

Previous work showed this converged in the iid case at $n^{-1/4}$ under suitable assumptions on a_i.

Another Decomposition



We can write the MSRV estimator in the dependent case as

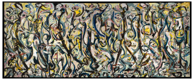
$$\widehat{\langle X, X\rangle}_T^{(msrv)} = \underbrace{\sum_{i=1}^{M} a_i \langle X, X\rangle_T^{(K_i)}}_{\text{signal}} + \underbrace{\sum_{i=1}^{M} a_i U_{n,K_i}}_{\text{noise}}$$

$$+2 \times \underbrace{\sum_{i=1}^{M} a_i \langle X, \epsilon\rangle_T^{(K_i)}}_{\text{signal-noise interaction}} + \underbrace{\sum_{i=1}^{M} a_i E_{n,K_i} + 2E\epsilon^2}_{\text{end points of noise}} + O_p(n^{-1/2})$$

where

$$U_{n,K_i} = -\frac{2}{K_i} \sum_{j=K_i}^{n} \epsilon_{t_j} \epsilon_{t_j - K_i}, \quad E_{n,K_i} = -\frac{1}{K_i} \sum_{j=0}^{K_i - 1} \epsilon_{t_j}^2 - \frac{1}{K_i} \sum_{j=n-K_i+1}^{n} \epsilon_{t_j}^2$$

Now, Some Sensible Weights



Consider a class of weights

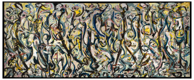
$$a_i = \frac{i}{M^2} h\left(\frac{i}{M}\right) - \frac{1}{2M^2}\left(\frac{i}{M}\right) h'\left(\frac{i}{M}\right)$$

Where $h \in \mathcal{C}^1$ and

$$\int_0^1 xh(x)dx = 1, \int_0^1 h(x)dx. = 0$$

Within this class of weights, the signal term is asymptotically unbiased.

Adding Dependence to MSRV

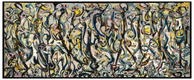


The only extra bias to account for due to dependence is $\sum_{i=1}^{M} a_i U_{n,K_i}$. Authors show that with our class of weights, we an bound it such that

$$\left| E\left[\sum_{i=1}^{M} a_i U_{n,K_i} \right] \right| \leq O(M^{-1})$$

Hence, they claim that if the rest of the estimator is $o_p(M^{-1/2}) = o_p(n^{-1/4})$, then this bias doesn't asymptotically matter. They calculate the limiting distribution of $\widehat{\langle X, X \rangle}_T^{(msrv)}$, it's $n^{-1/4}$-consistent with a very complicated variance.

Reminder Why TSRV is Better than RV



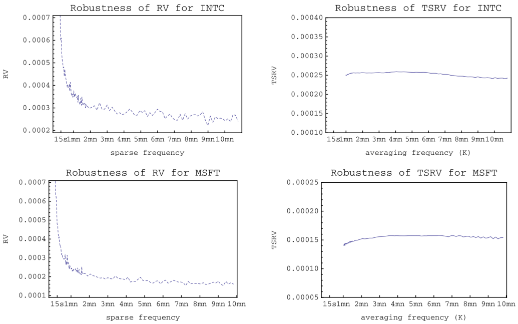


Figure: It’s way more stable than RV

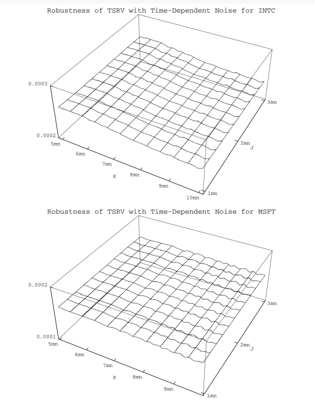
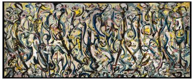
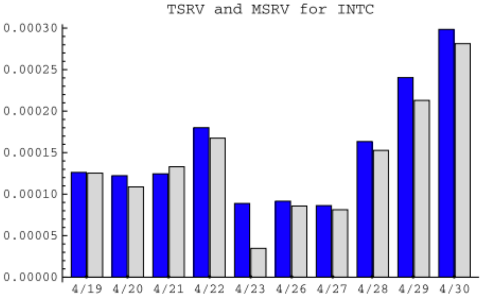


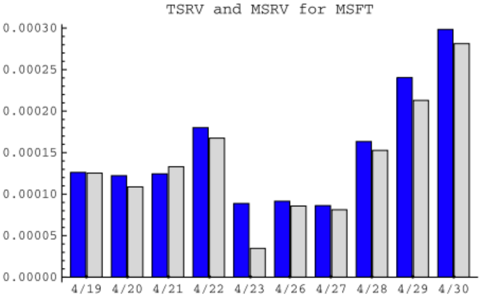
Figure: Robust to choice of J and K

TSRV and MSRV are Similar

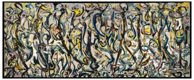


Authors suggest it is a trade off of computational complexity to go from $n^{-1/6}$ convergence to $n^{-1/4}$. Qualitatively, MSRV looks only slightly different:





References



Aït-Sahalia, Yacine and Mykland, Per A and Zhang, Lan (2008)
Ultra high frequency volatility estimation with dependent microstructure noise
Journal of Econometrics 160(1), 160–175.

Zhang, Lan and Mykland, Per A and Aït-Sahalia, Yacine (2005)
A tale of two time scales: Determining integrated volatility with noisy high-frequency data
Journal of the American Statistical Association 100(472), 1394–1411.

Thank You