Ultra High Frequency Volatility Estimation with Market Microstructure Noise

Yacine Aït-Sahalia
Princeton University

Per A. Mykland
The University of Chicago

Lan Zhang
Carnegie-Mellon University

1. Introduction

1. Introduction

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

$$Y_{\mathcal{T}} = X_{\mathcal{T}} + \varepsilon_{\mathcal{T}}$$

1. Introduction

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

$$Y_\tau = X_\tau + \varepsilon_\tau$$

- ε 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 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,..., $n\Delta = T$.

- Without noise, the realized volatility $\sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2$ estimates the quadratic variation $\int_0^T \sigma_t^2 dt$.

- 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,..., $n\Delta = T$.

- Without noise, the realized volatility $\sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2$ estimates the quadratic variation $\int_0^T \sigma_t^2 dt$.

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

- We show that the situation changes radically in the presence of market microstructure noise:

- We show that the situation changes radically in the presence of market microstructure noise:
  - Computing RV using all the data (say every second) leads to an estimate of the variance of the noise, not the quadratic variation that one seeks to estimate.

- We show that the situation changes radically in the presence of market microstructure noise:
  - Computing RV using all the data (say every second) leads to an estimate of the variance of the noise, not the quadratic variation that one seeks to estimate.
  - In practice, people sample sparsely at some lower frequency (5 mn to 30 mn usually)

- We show that the situation changes radically in the presence of market microstructure noise:
  - Computing RV using all the data (say every second) leads to an estimate of the variance of the noise, not the quadratic variation that one seeks to estimate.
  - In practice, people sample sparsely at some lower frequency (5 mn to 30 mn usually)
  - If one insists upon sampling sparsely, what is the right frequency? We show how to determine the optimal sparse frequency.

- We show that the situation changes radically in the presence of market microstructure noise:
  - Computing RV using all the data (say every second) leads to an estimate of the variance of the noise, not the quadratic variation that one seeks to estimate.
  - In practice, people sample sparsely at some lower frequency (5 mn to 30 mn usually)
  - If one insists upon sampling sparsely, what is the right frequency? We show how to determine the optimal sparse frequency.

- But even if sampling optimally, one is throwing away a large amount of data.

- But even if sampling optimally, one is 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 $n = T/\delta = 23,400$.

- But even if sampling optimally, one is 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 $n = T/\delta = 23,400$.
  - But sampling sparsely even at the highest frequency used by empirical researchers (once every 5 mn) means throwing away 299 out of every 300 observations: the sample size used is only $n_{sparse} = 78$.

- But even if sampling optimally, one is 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 $n = T/\delta = 23,400$.

  - But sampling sparsely even at the highest frequency used by empirical researchers (once every 5 mn) means throwing away 299 out of every 300 observations: the sample size used is only $n_{sparse} = 78$.

- This violates one of the most basic principles of statistics.

- But even if sampling optimally, one is 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 $n = T/\delta = 23,400$.

  - But sampling sparsely even at the highest frequency used by empirical researchers (once every 5 mn) means throwing away 299 out of every 300 observations: the sample size used is only $n_{sparse} = 78$.

- This violates one of the most basic principles of statistics.

- We propose a solution which makes use of the full data sample:

- We propose a solution which makes use of the full data sample:
  - Our estimator (TSRV) is based on subsampling, averaging and bias-correction.

- We propose a solution which makes use of the full data sample:
  - Our estimator (TSRV) is based on subsampling, averaging and bias-correction.
  - This is the only consistent estimator (we know of) for quadratic variation in the presence of market microstructure noise.

- We propose a solution which makes use of the full data sample:
  - Our estimator (TSRV) is based on subsampling, averaging and bias-correction.
  - This is the only consistent estimator (we know of) for quadratic variation in the presence of market microstructure noise.

2. Outline

2. Outline

- Without market microstructure noise

2. Outline

- Without market microstructure noise
- What happens when noise is present

2. Outline

- Without market microstructure noise

- What happens when noise is present
  - The fifth best approach: use all the data to compute RV

2. Outline

- Without market microstructure noise

- What happens when noise is present
  - The fifth best approach: use all the data to compute RV
  - The fourth best approach: sample sparsely (5, 10 or 15 mn) to compute RV

2. Outline

- Without market microstructure noise

- What happens when noise is present
  - The fifth best approach: use all the data to compute RV
  - The fourth best approach: sample sparsely (5, 10 or 15 mn) to compute RV
  - The third best approach: sample sparsely at an optimally determined frequency to compute RV

2. Outline

- Without market microstructure noise

- What happens when noise is present
  - The fifth best approach: use all the data to compute RV
  - The fourth best approach: sample sparsely (5, 10 or 15 mn) to compute RV
  - The third best approach: sample sparsely at an optimally determined frequency to compute RV

– The second best approach: subsampling and averaging

- The second best approach: subsampling and averaging
- The first best approach, TSRV: use the fifth best to bias-correct the second best

- The second best approach: subsampling and averaging
- The first best approach, TSRV: use the fifth best to bias-correct the second best

- Monte Carlos

- The second best approach: subsampling and averaging
- The first best approach, TSRV: use the fifth best to bias-correct the second best

- Monte Carlos

- Two extensions: time series dependence in the noise, and MSRV

- The second best approach: subsampling and averaging
- The first best approach, TSRV: use the fifth best to bias-correct the second best

- Monte Carlos

- Two extensions: time series dependence in the noise, and MSRV

- Data analysis: INTC, MSFT

- The second best approach: subsampling and averaging
- The first best approach, TSRV: use the fifth best to bias-correct the second best

- Monte Carlos
- Two extensions: time series dependence in the noise, and MSRV
- Data analysis: INTC, MSFT

3. Without Market Microstructure Noise

3. Without Market Microstructure Noise

- Suppose first that σ is constant. Without noise, the log-returns are iid $N(0, \sigma^2 \Delta)$. The MLE for σ^2 coincides with the realized volatility of the process,

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2,$$

3. Without Market Microstructure Noise

- Suppose first that σ is constant. Without noise, the log-returns are iid $N(0, \sigma^2\Delta)$. The MLE for σ^2 coincides with the realized volatility of the process,

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2,$$

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

3. Without Market Microstructure Noise

- Suppose first that σ is constant. Without noise, the log-returns are iid $N(0, \sigma^2\Delta)$. The MLE for σ^2 coincides with the realized volatility of the process,

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2,$$

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

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

- When volatility is stochastic, $dX_t = \sigma_t dW_t$:

- When volatility is stochastic, $dX_t = \sigma_t dW_t$:
  - Realized volatility $\sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2$ estimates the quadratic variation $\int_0^T \sigma_t^2 dt$.

- When volatility is stochastic, $dX_t = \sigma_t dW_t$:
  - Realized volatility $\sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2$ estimates the quadratic variation $\int_0^T \sigma_t^2 dt$.
  - The sum converges to the integral, with a known distribution: Jacod (1994), Jacod and Protter (1998), etc.

- When volatility is stochastic, $dX_t = \sigma_t dW_t$:
  - Realized volatility $\sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2$ estimates the quadratic variation $\int_0^T \sigma_t^2 dt$.
  - The sum converges to the integral, with a known distribution: Jacod (1994), Jacod and Protter (1998), etc.
  - As in the constant σ case, selecting Δ as small as possible ($= n$ as large as possible) is optimal.

4. With Noise

4. With Noise

- When $dX_t = \sigma_t dW_t$, but we observe X with noise, the object of interest remains the quadratic variation of X :

$$\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$$

over a fixed time period $[0, T]$, or possibly several such time periods.

4. With Noise

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

- Asymptotics are in $\Delta \to 0$, with T fixed.

4. With Noise

- When $dX_t = \sigma_t dW_t$, but we observe X with noise, the object of interest remains the quadratic variation of X :

$$\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$$

  over a fixed time period $[0, T]$, or possibly several such time periods.

- Asymptotics are in $\Delta \longrightarrow 0$, with T fixed.

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

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

4.1. Completely Ignoring the Noise: Fifth Best

4.1. Completely Ignoring the Noise: Fifth Best

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

4.1. Completely Ignoring the Noise: Fifth Best

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

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

conditionally on the X process.

4.1. Completely Ignoring the Noise: Fifth Best

- 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.

4.2. Sampling Sparsely at an Arbitrary Frequency: Fourth Best

4.2. Sampling Sparsely at an Arbitrary Frequency: Fourth Best

- Of course, sampling as prescribed by $[Y,Y]_T^{(all)}$ is not what people do in practice

4.2. Sampling Sparsely at an Arbitrary Frequency: Fourth Best

- Of course, sampling as prescribed by $[Y, Y]_T^{(all)}$ is not what people do in practice

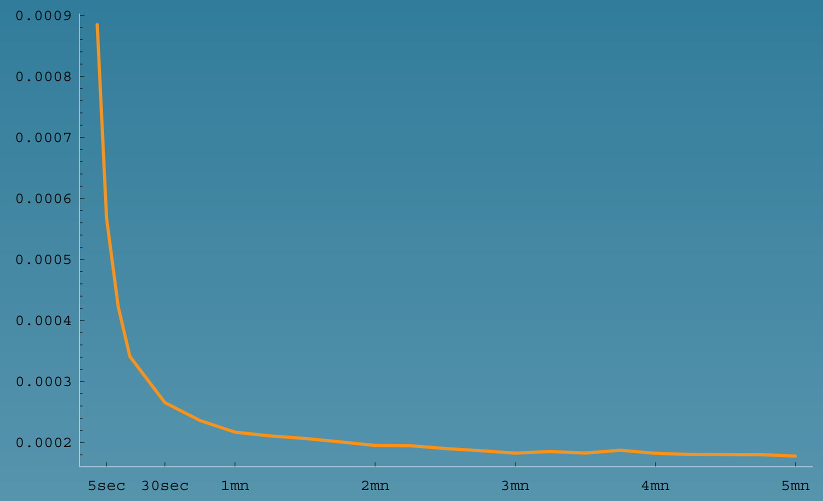
- Instead, they use the estimator $[Y, Y]_T^{(sparse)}$ constructed by summing squared log-returns at some lower frequency: 5 mn, or 10, 15, 30 mn, typically.

4.2. Sampling Sparsely at an Arbitrary Frequency: Fourth Best

- Of course, sampling as prescribed by $[Y, Y]_T^{(all)}$ is not what people do in practice

- Instead, they use the estimator $[Y, Y]_T^{(sparse)}$ constructed by summing squared log-returns at some lower frequency: 5 mn, or 10, 15, 30 mn, typically.

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



- As $\Delta = T/n \longrightarrow 0$, 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.

4.3. Sampling Sparsely at an Optimal Frequency: Third Best

4.3. Sampling Sparsely at an Optimal Frequency: Third Best

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

4.3. Sampling Sparsely at an Optimal Frequency: Third Best

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

- To determine optimally the sparse sampling frequency, we show that:

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

4.3. Sampling Sparsely at an Optimal Frequency: Third Best

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

- To determine optimally the sparse sampling frequency, we show that:

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

- This gives rise to the third best estimator we define as $[Y, Y]_T^{(sparse,opt)}$.

4.4. Correcting for Microstructure Noise: Second Best

4.4. Correcting for Microstructure Noise: Second Best

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

4.4. Correcting for Microstructure Noise: Second Best

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

- Indeed, the fourth and third best estimators do better than the fifth best.

4.4. Correcting for Microstructure Noise: Second Best

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

- Indeed, the fourth and third best estimators do better than the fifth best.

- But this entails discarding a very substantial fraction of the data sample.

4.4. Correcting for Microstructure Noise: Second Best

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

- Indeed, the fourth and third best estimators do better than the fifth best.

- But this entails discarding a very substantial fraction of the data sample.

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

4.4. Correcting for Microstructure Noise: Second Best

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

- Indeed, the fourth and third best estimators do better than the fifth best.

- But this entails discarding a very substantial fraction of the data sample.

- 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 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 $n/K \longrightarrow \infty$ as $n \longrightarrow \infty$.

- 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 $n/K \longrightarrow \infty$ as $n \longrightarrow \infty$.
  - 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.

- 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 $n/K \longrightarrow \infty$ as $n \longrightarrow \infty$.
  - 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.

- 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 $n/K \longrightarrow \infty$ as $n \longrightarrow \infty$.
  - 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.

- 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 $n/K \longrightarrow \infty$ as $n \longrightarrow \infty$.
  - 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 estimators $[Y,Y]_T^{(k)}$ obtained on K grids of average size $\bar{n}$.

- 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 estimators $[Y,Y]_T^{(k)}$ obtained on K grids of average size $\bar{n}$.

- We show that:

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

- 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 estimators $[Y,Y]_T^{(k)}$ obtained on K grids of average size $\bar{n}$.

- We show that:

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

4.5. Two Scales Realized Volatility: First Best

4.5. Two Scales Realized Volatility: First Best

- While a better estimator than $[Y,Y]_T^{(all)}$, $[Y,Y]_T^{(avg)}$ remains biased.

4.5. Two Scales Realized Volatility: First Best

- While a better estimator than $[Y,Y]_T^{(all)}$, $[Y,Y]_T^{(avg)}$ remains biased.

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

4.5. Two Scales Realized Volatility: First Best

- While a better estimator than $[Y,Y]_T^{(all)}$, $[Y,Y]_T^{(avg)}$ remains biased.

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

- But recall that $E[\varepsilon^2]$ can be consistently approximated by the fifth best estimator:

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

4.5. Two Scales Realized Volatility: First Best

- While a better estimator than $[Y,Y]_T^{(all)}$, $[Y,Y]_T^{(avg)}$ remains biased.

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

- But recall that $E[\varepsilon^2]$ can be consistently approximated by the fifth best estimator:

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

- 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

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

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

- 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 call this estimator Two Scales Realized Volatility.

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

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

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

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

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

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

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

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

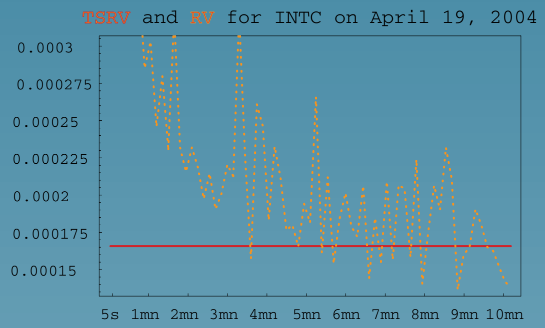
- To the best of our knowledge, this is the only consistent estimator for $\langle X, X \rangle_T$ in the presence of market microstructure noise.

5. Monte Carlo Simulations

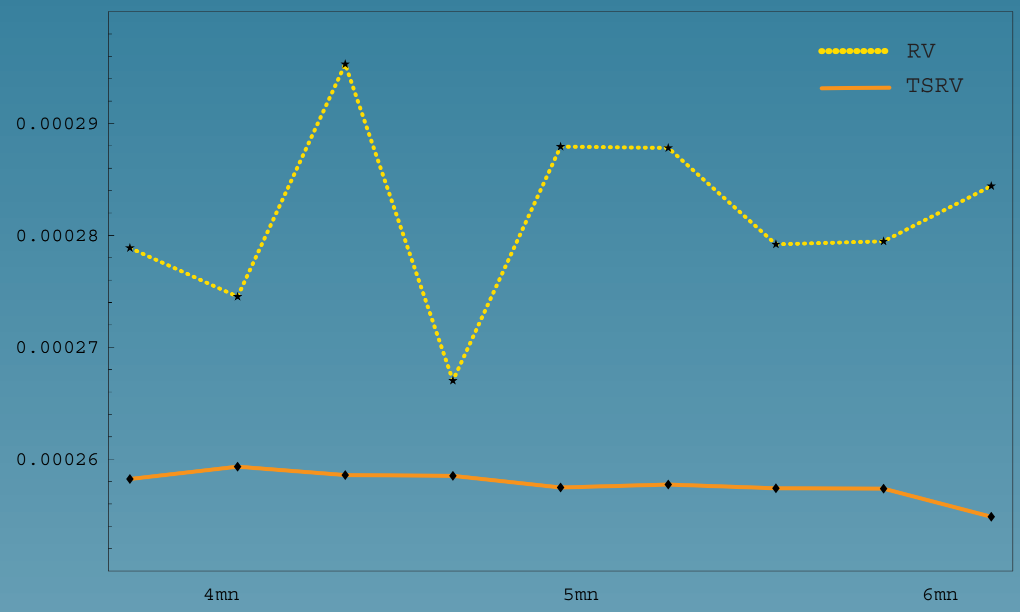
	Fifth Best $[Y,Y]_T^{(all)}$	RV Fourth Best $[Y,Y]_T^{(sparse)}$	Third Best $[Y,Y]_T^{(sparse,opt)}$	Second Best $[Y,Y]_T^{(avg)}$	TSRV First Best $\widehat{\langle X,X\rangle}_T^{(adj)}$
Small Sample Bias	$1.1699\ 10^{-2}$	$3.89\ 10^{-5}$	$2.18\ 10^{-5}$	$1.926\ 10^{-5}$	$2\ 10^{-8}$
Asymptotic Bias	$1.1700\ 10^{-2}$	$3.90\ 10^{-5}$	$2.20\ 10^{-5}$	$1.927\ 10^{-5}$	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

6. 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:



7. Dependent Market Microstructure Noise

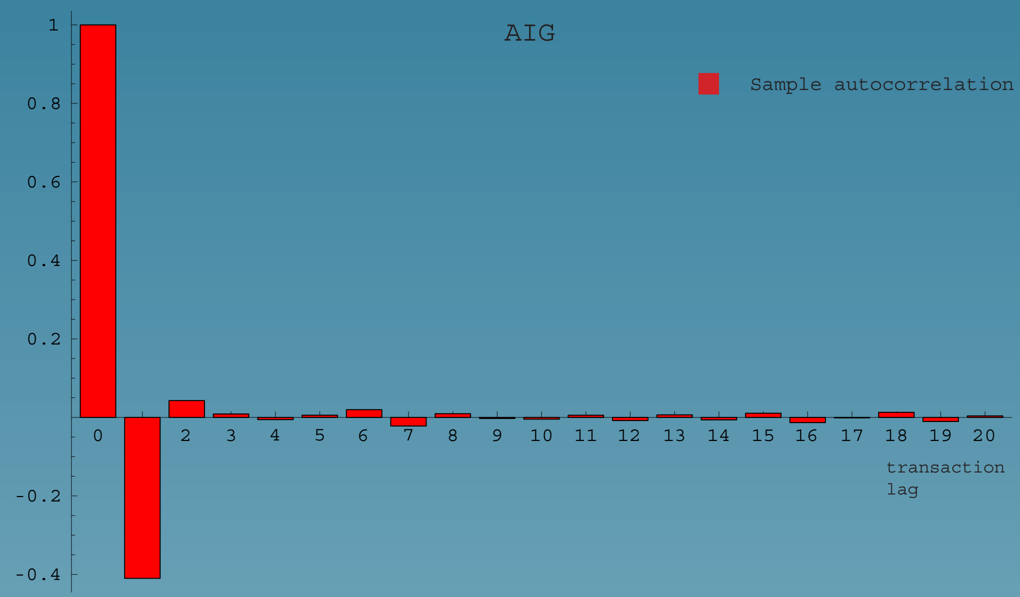
7. Dependent Market Microstructure Noise

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

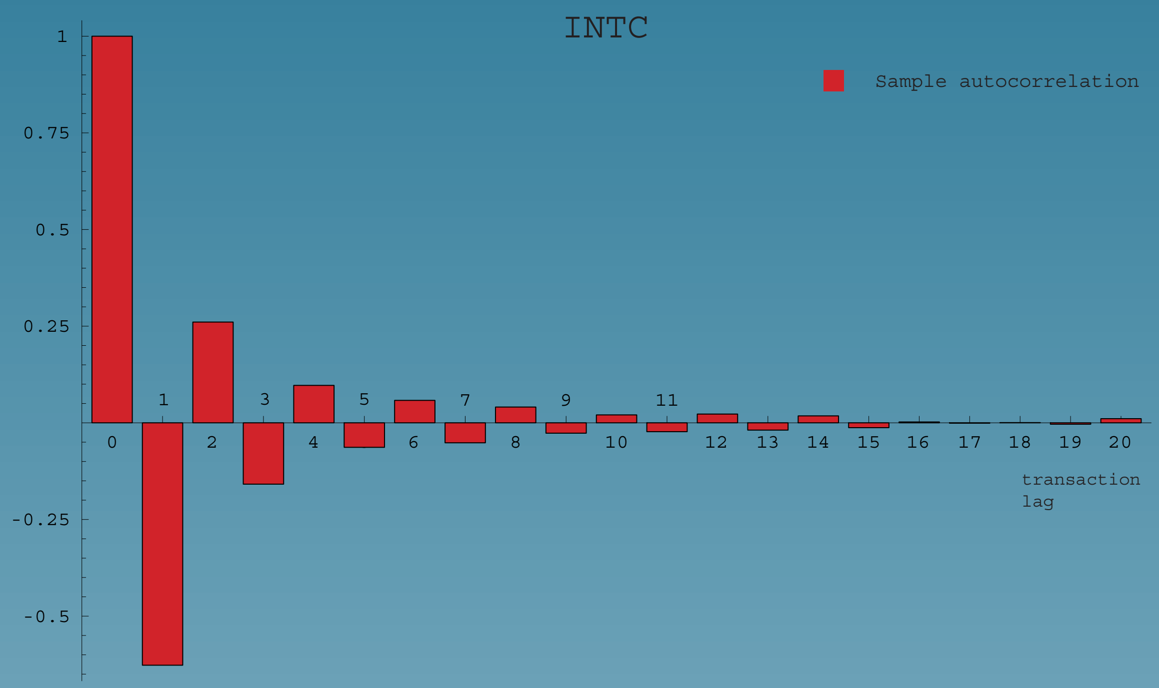
- In that case, log-returns are MA(1):

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

- 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

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

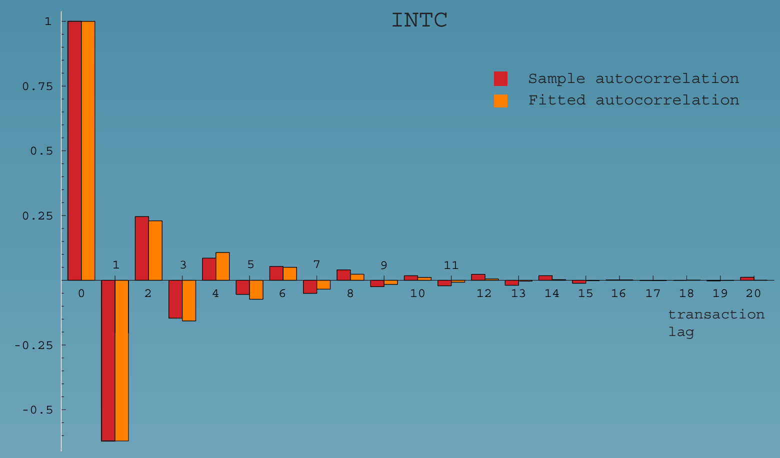
where U is iid, V is $AR(1)$ and $U \perp V$.

- 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

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

- 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.

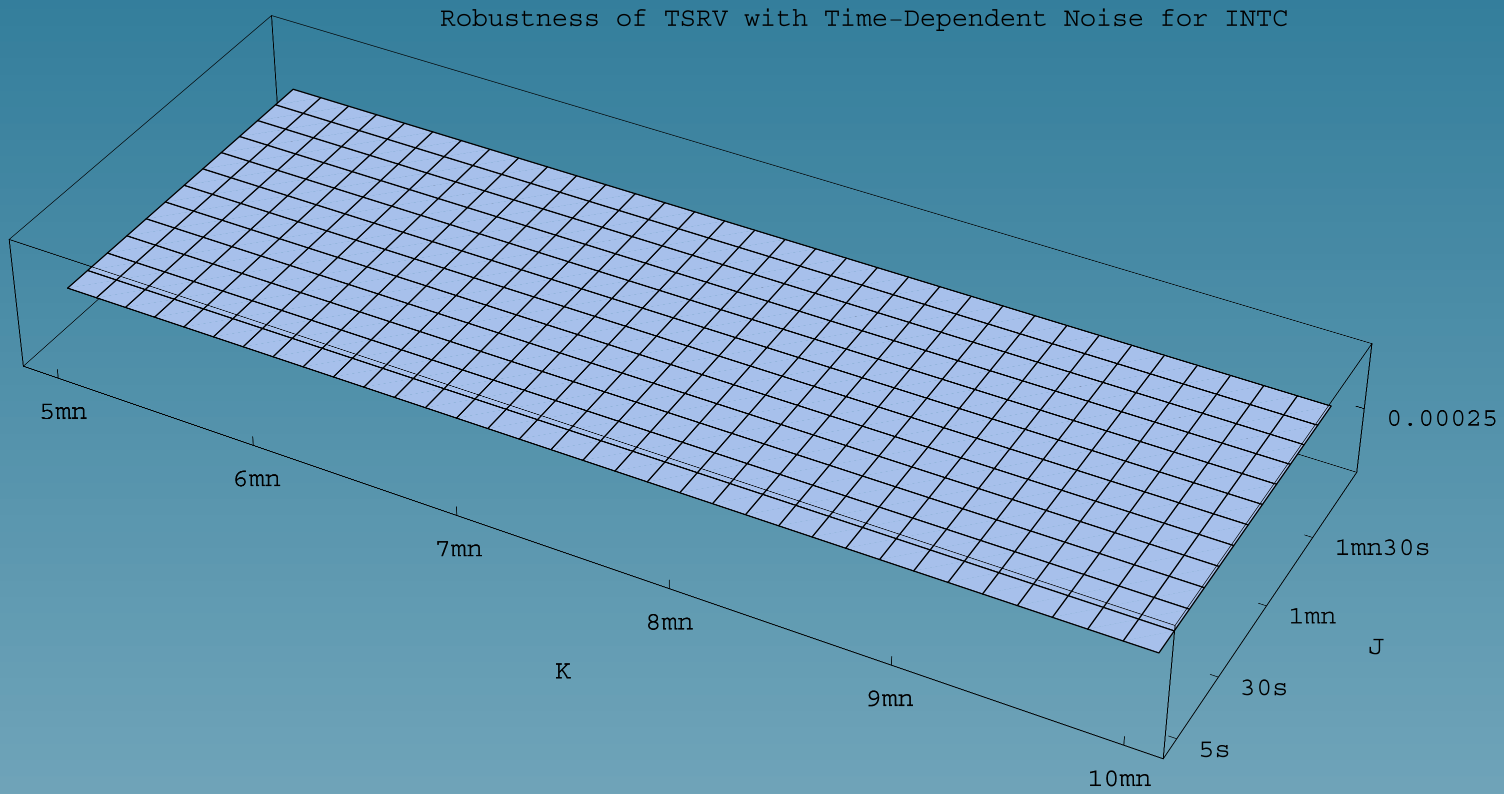
- The TSRV Estimator with (J, K) Time Scales

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

- 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:



8. Further Refinement: Multiple Scales Realized Volatility

8. Further Refinement: Multiple Scales Realized Volatility

- We have seen that TSRV provides:

8. Further Refinement: Multiple Scales Realized Volatility

- We have seen that TSRV provides:
  - the first consistent and asymptotic (mixed) normal estimator of the quadratic variation $\langle X, X \rangle_T$;

8. Further Refinement: Multiple Scales Realized Volatility

- 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;

8. Further Refinement: Multiple Scales Realized Volatility

- 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 $n^{-1/6}$.

8. Further Refinement: Multiple Scales Realized Volatility

- 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 2004).

- 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 2004).

- The resulting estimator, MSRV has the form of

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

- 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 2004).

- 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 $n^{-1/4}$.

- 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 $n^{-1/4}$.

- Optimal weights are given in closed-form.

- For suitably selected weights a_i and $M = O(n^{1/2})$, $\widehat{\langle X, X \rangle}_T^{(\mathrm{msrv})}$ converges to the $\langle X, X \rangle_T$ at rate $n^{-1/4}$.

- Optimal weights are given in closed-form.

- We also provide an analysis of this estimator under dependence of the noise.

9. Conclusions

9. Conclusions

- Two Scales Realized Volatility

9. Conclusions

- Two Scales Realized Volatility
  - In the limit where all the data is used, realized volatility converges to the variance of the noise, not the quadratic variation of the log-returns

9. Conclusions

- Two Scales Realized Volatility
  - In the limit where all the data is used, realized volatility converges to the variance of the noise, not the quadratic variation of the log-returns
  - The practical response so far has been to use sparse sampling: once every 5 or 10 minutes.

9. Conclusions

- Two Scales Realized Volatility
  - In the limit where all the data is used, realized volatility converges to the variance of the noise, not the quadratic variation of the log-returns
  - The practical response so far has been to use sparse sampling: once every 5 or 10 minutes.
  - But it is possible instead to correct for the noise by subsampling, averaging and bias-correcting and obtain a well behaved estimator that makes use of all the data: TSRV

9. Conclusions

- Two Scales Realized Volatility
  - In the limit where all the data is used, realized volatility converges to the variance of the noise, not the quadratic variation of the log-returns
  - The practical response so far has been to use sparse sampling: once every 5 or 10 minutes.
  - But it is possible instead to correct for the noise by subsampling, averaging and bias-correcting and obtain a well behaved estimator that makes use of all the data: TSRV

- The difference matters:

- The difference matters:
  - In Monte Carlo simulations, the RMSE of TSRV is orders of magnitude smaller than that of RV

- The difference matters:
  - In Monte Carlo simulations, the RMSE of TSRV is orders of magnitude smaller than that of RV
  - In empirical examples, the difference is also meaningful: for INTC, TSRV $=$ 0.0025 vs. RV in the range (0.0029, 0.0035).

- The difference matters:
  - In Monte Carlo simulations, the RMSE of TSRV is orders of magnitude smaller than that of RV
  - In empirical examples, the difference is also meaningful: for INTC, TSRV = 0.0025 vs. RV in the range (0.0029, 0.0035).

- And one final important message:

- And one final important message:
  - Any time one has an impulse to discard data, one can usually do better: using likelihood corrections in the parametric volatility case or subsampling and averaging in the stochastic volatility case.

- And one final important message:
  - Any time one has an impulse to discard data, one can usually do better: using likelihood corrections in the parametric volatility case or subsampling and averaging in the stochastic volatility case.
  - No matter what the model is, no matter what quantity is being estimated.