Robust Estimation of Realized Correlation

Yiyao Luo

Economics Department
University of North Carolina at Chapel Hill

November 18, 2022

Pearson sample correlation estimator

$$P = \frac{\sum_{i=1}^{n} \Delta X_i \Delta Y_i}{\sqrt{\sum_{i=1}^{n} \Delta X_i^2 \sum_{i=1}^{n} \Delta Y_i^2}}$$

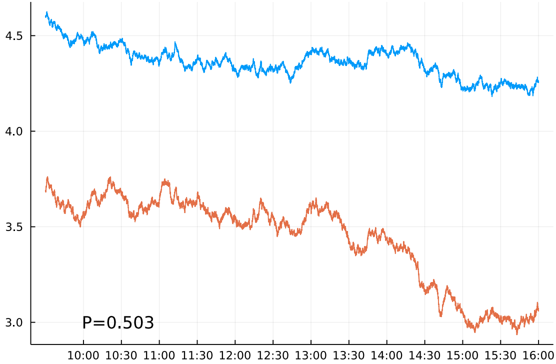


Figure 1: $\mathrm{Corr}(\Delta X_i, \Delta Y_i) = 0.5$

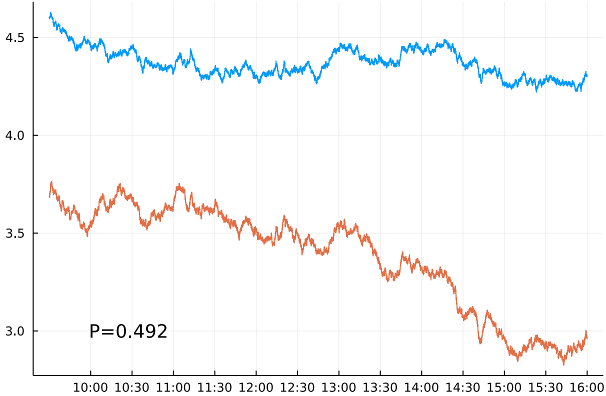
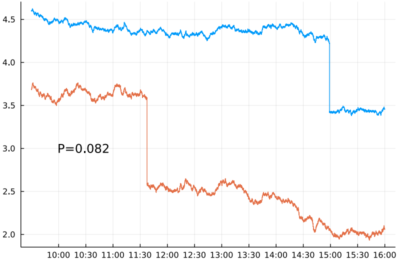
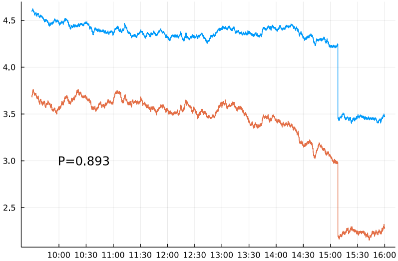


Figure 2: Add i.i.d. noise on the efficient log-prices, $\rho = 0.5$

Jumps, $\rho = 0.5$



(a) Individual Jump



(b) CoJump

Time-varying volatility

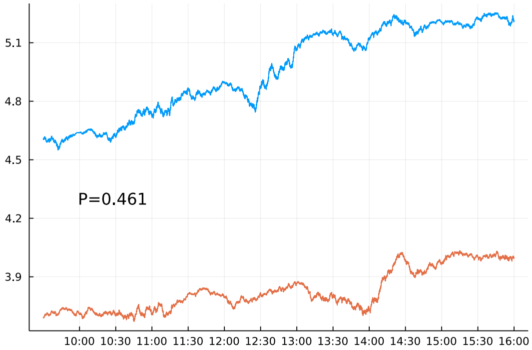
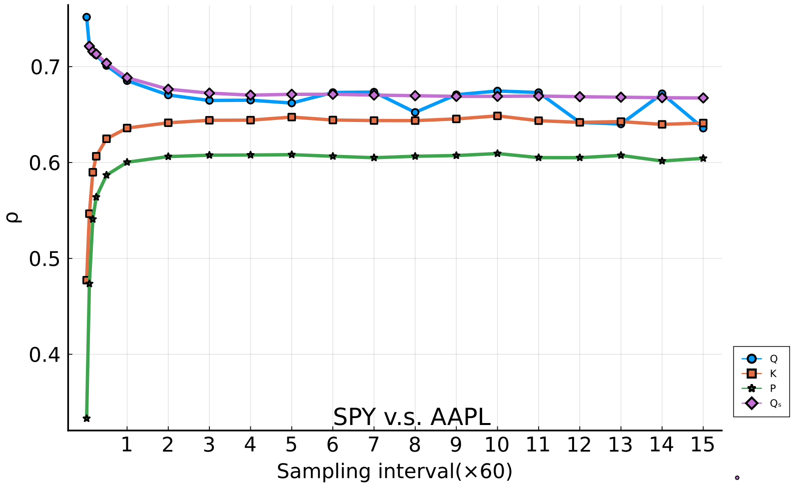


Figure 4: Simulate efficient log-prices in Heston model, $\rho = 0.5$

Epps Effect



Back

Literature

- Epps effect: Renò (2003), Precup and Iori (2007), Münnix et al. (2011); Tóth and Kertész (2007); Chang et al. (2021).
- Correct estimators for various microstructure issues:
  - Noise: Andersen et al. (2001), Bandi and Russell (2008); Podolskij and Vetter (2009), Jacod et al. (2009), Christensen et al. (2010), Christensen et al. (2013).
  - Asynchronous trading: Hayashi and Yoshida (2005), Zhang (2011), Barndorff-Nielsen et al. (2011).
  - Rounding error: Delattre and Jacod (1997), Rosenbaum (2009), Li et al. (2018).
  - Jump: Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen and Shephard (2007), Boudt et al. (2011).

Some robust alternatives

- Sign/rank-based estimators
  - Quadrant, Blomqvist (1950)
  - Kendall, Kendall (1938)

  Apply to high-frequency data, see Vander Elst and Veredas (2016).
- Gaussian rank estimator: Boudt et al. (2012).

Relation between ρ and q (sign-concordance)

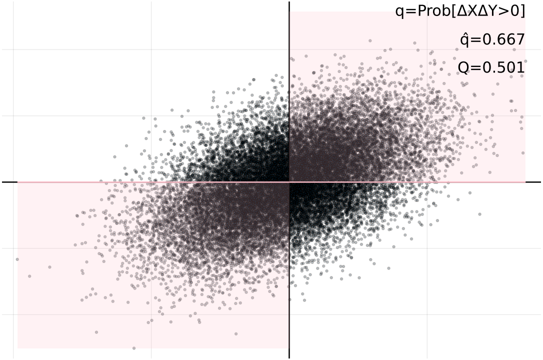


Figure 5: Scatter plot of $(\Delta X_i, \Delta Y_i)$, $\rho = 0.5$

Elliptical distributions (Normal, Student's t, mixed-Normal):

$$\rho = \varrho(q) = \sin(\pi(q - \frac{1}{2}))$$

Noise

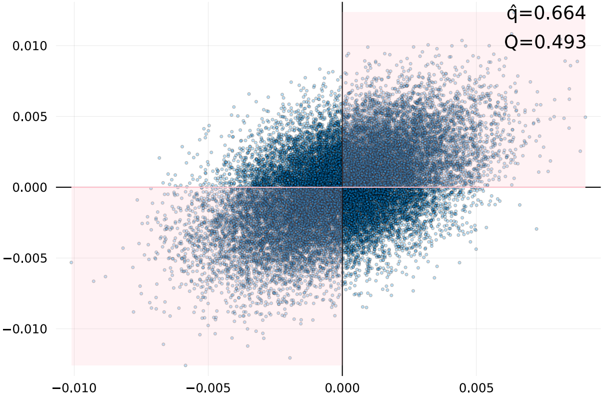
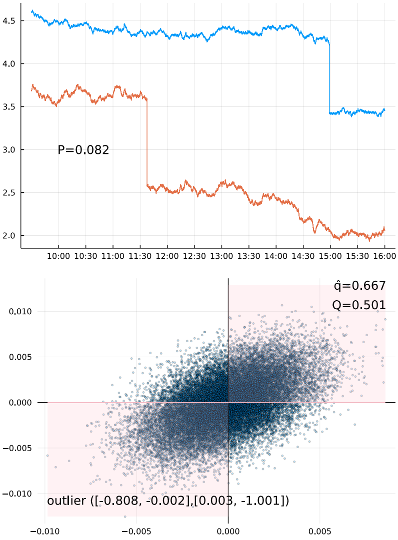
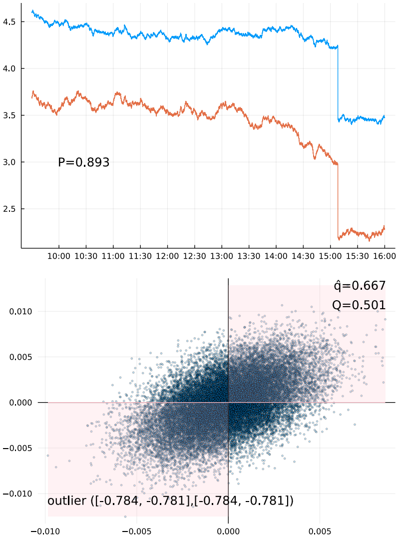


Figure 6: Add i.i.d. noise on the efficient log-prices, $\rho = 0.5$

Jumps, $\rho = 0.5$



(a) Individual Jump



(b) CoJump

Time-varying volatility

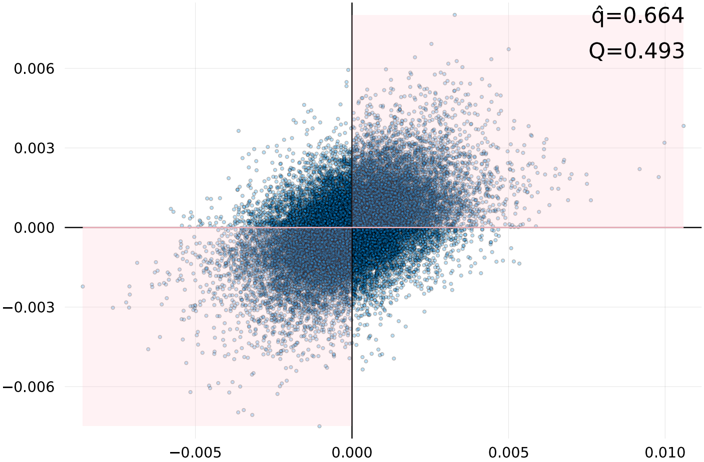


Figure 8: Scatter plot of $(\Delta X_i, \Delta Y_i)$ generated in Heston model, $\rho = 0.5$

Motivation

- Pearson is the best in "ideal" case but fragile
  - Fragility is measured by influence function (Rousseeuw et al. (2011))
- Quadrant is robust when applied to high-frequency financial data
  - Low stability on finite samples
- This paper: Improve efficiency of Quadrant

Main results

- Introduce subsampled Quadrant (Q_s) correlation estimator as a natural extension of Quadrant estimator.
- Subsampled Quadrant correlation estimator is more accurate than Pearson estimator while facing microstructure issues.
- Quadrant and Subsampled Quadrant estimators are consistent in models with time-varying volatility.

Notation

- Observe 2-dimensional log-price processes $(X_t, Y_t)_{t=0}^{T}$, then returns sampled at every second (transaction) are defined as

$$\Delta X_i = X_i - X_{i-1}$$

for $i = 1, \ldots, T$.

- Sparse sampling: sample at every S-th second (transaction)

$$\Delta_S X_{iS} = X_{iS} - X_{(i-1)S}$$

for $i = 1, \ldots, n$ with $n = \lfloor \frac{T}{S} \rfloor$.

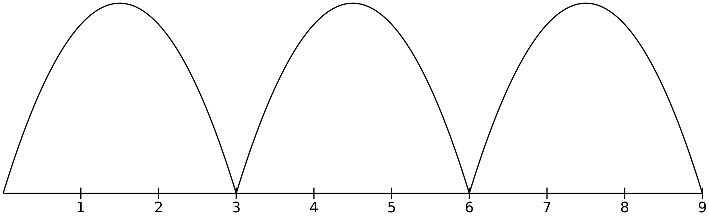
- Subsampling (Bartlett sampling):

$$\Delta_S X_i = X_i - X_{i-S}$$

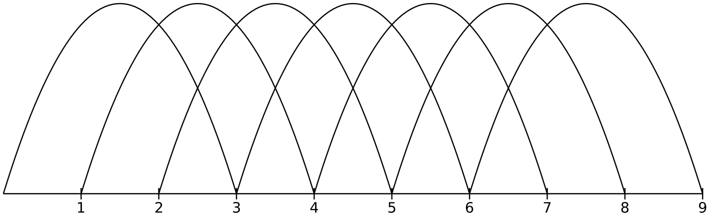
for $i = S, \ldots, T$.

  - Econometric: Newey West estimator
  - High-frequency financial data: Zhang et al. (2005) and Zhang (2011)

Sparse sampling versus subsampling, $S = 3$



(a) Sparse sampling: $\Delta_3 X_i,\ i = 3, 6, 9$



(b) Subsampling: $\Delta_3 X_i,\ i = 3, \ldots, 9$

Quadrant, Kendall, and Pearson

$(X_t, Y_t)_{t=0}^{T}$ are observed prices: with $n = \lfloor \frac{T}{S} \rfloor$

- Quadrant: Figure

$$Q = \varrho(\hat{q}_Q) \text{ with } \hat{q}_Q = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{\Delta_S X_{iS} \Delta_S Y_{iS} > 0\}},$$

- Kendall:

$$K = \varrho(\hat{q}_K) \text{ with } \hat{q}_K = \frac{2}{n(n-1)} \sum_{i<j} \mathbf{1}_{\{(\Delta_S X_{iS} - \Delta_S X_{jS})(\Delta_S Y_{iS} - \Delta_S Y_{jS}) > 0\}}$$

- Pearson:

$$P = \frac{\sum_{i=1}^{n} \Delta_S X_{iS} \Delta_S Y_{iS}}{\sqrt{\sum_{i=1}^{n} \Delta_S X_{iS}^2 \sum_{i=1}^{n} \Delta_S Y_{iS}^2}}$$

Subsampled Quadrant

Subsampled Quadrant estimator (Q_s) on returns within a window of width s moving at lag 1, or all consecutive S-second (S-transaction) returns:

$$Q_S = \varrho(\hat{q}_{Q_S}) \text{ with } \hat{q}_{Q_S} = \frac{1}{N} \sum_{i=S}^{T} \mathbf{1}_{\{\Delta_S X_i \Delta_S Y_i > 0\}},$$

where $N = T - S$.

Appendix

Properties

Asymptotics

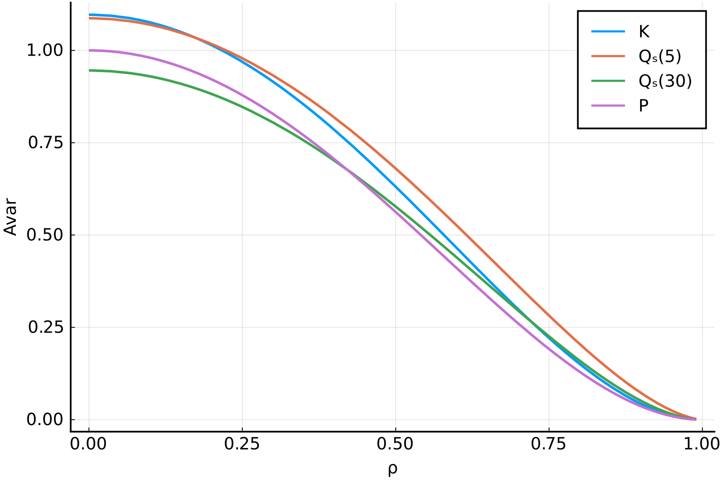
For fixed S, $n = \lfloor \frac{T}{s} \rfloor$, then as $T \to \infty$,

$$\sqrt{n}(\bullet - \rho) \xrightarrow{d} N(0, V_\bullet), \quad \bullet = P, Q, K, Q_S$$

where

- $V_P = (1-\rho^2)^2$
- $V_Q = (\frac{\pi^2}{4} - \arcsin^2 \rho)(1-\rho^2)$
- $V_K = (\frac{\pi^2}{9} - 4\arcsin^2(\frac{\rho}{2}))(1-\rho^2)$
- $V_{Q_S} = \frac{1-\rho^2}{S}[\frac{\pi^2}{4} - \arcsin^2 \rho + 2\sum_{h=1}^{S-1} \arcsin^2(\frac{S-h}{S}) - \arcsin^2(\rho\frac{S-h}{S})]$

Asymptotics



Influence function

The influence function of a statistical functional R at distribution H is defined as

$$\mathrm{IF}((x_0, y_0), R, H) = \lim_{\varepsilon \to 0} \frac{R((1-\varepsilon)H + \varepsilon\Delta_{(x_0,y_0)}) - R(H)}{\varepsilon}$$

where $\Delta_{(x_0,y_0)}$ is a Dirac measure of mass at (x_0, y_0).

Influence function

- Pearson:

$$\mathrm{IF}((x_0, y_0), R_P, \Phi_\rho) = x_0 y_0 - (\frac{x_0^2 + y_0^2}{2})\rho$$

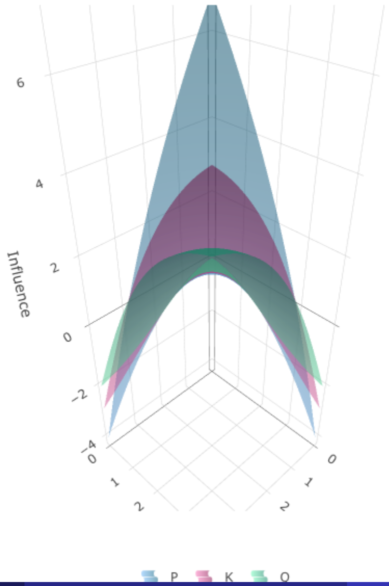
- Kendall:

$$\mathrm{IF}((x_0, y_0), R_K, \Phi_\rho) = 2\pi s\sqrt{1-\rho^2}\times$$
$$[2\Phi_\rho(\tfrac{x_0}{\sqrt{2S-1}}, \tfrac{y_0}{\sqrt{2S-1}}) - \Phi(\tfrac{x_0}{\sqrt{2S-1}}) - \Phi(\tfrac{y_0}{\sqrt{2S-1}}) + 1 - q]$$

- Quadrant and Subsampled Quadrant:

$$\mathrm{IF}((x_0, y_0), R_Q, \Phi_\rho) = \pi S\sqrt{1-\rho^2}\times$$
$$[2\Phi_\rho(\tfrac{x_0}{\sqrt{S-1}}, \tfrac{y_0}{\sqrt{S-1}}) - \Phi(\tfrac{x_0}{\sqrt{S-1}}) - \Phi(\tfrac{y_0}{\sqrt{S-1}}) + 1 - q]$$

$\rho = 0.5$ and $S = 5$



Consistency

$(X_t, Y_t)_{t\in[0,T]}$ be a bivariate Itô semimartingale process characterized by

$$\begin{pmatrix} X_t \\ y_t \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \int_0^t a_u du + \int_0^t \sigma_u dW_u$$

where $a = (a_x, a_y)$ is a locally bounded predictable drift function and

$$\sigma = \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{pmatrix}$$

is a cadlag volatility process. W denotes a 2-dimensional Wiener process and $\mathrm{Corr}(dW_x, dW_y) = \rho$.

Consistency

Assumption

The cadlag process σ satisfies, for any $\varepsilon > 0$, there is $\delta > 0$ such that for any $t_1, t_2 \in [0, T]$ and $0 < t_2 - t_1 < \delta$

$$\left| \frac{\int_{t_1}^{t_2} \sigma_{x,u}\sigma_{y,u} du}{\sqrt{\int_{t_1}^{t_2} \sigma_{x,u}^2 du \int_{t_1}^{t_2} \sigma_{y,u}^2 du}} - 1 \right| < \varepsilon.$$

Theorem

Under the above representation of (X, Y) and Assumption, Q and Q_S are consistent to ρ.

Inconsistency

- Pearson:

$$P \xrightarrow{p} \rho \times \frac{\int_0^T \sigma_{x,u}\sigma_{y,u}du}{\sqrt{\int_0^T \sigma_{x,u}^2 du \int_0^T \sigma_{y,u}^2 du}}.$$

- Kendall:

$$\begin{aligned}&\mathrm{Prob}[(\Delta_S X_{iS} - \Delta_S X_{jS})(\Delta_S Y_{iS} - \Delta_S Y_{jS}) > 0]\\ =&\frac{1}{\pi}\arcsin(\rho G(\sigma, i, j)) + \frac{1}{2}\end{aligned}$$

$$G(\sigma, i, j) = \frac{\int_{(i-1)S}^{iS} \sigma_{x,u}\sigma_{y,u}du + \int_{(j-1)S}^{jS} \sigma_{x,u}\sigma_{y,u}du}{\sqrt{(\int_{(i-1)S}^{iS} \sigma_{x,u}^2 du + \int_{(j-1)S}^{jS} \sigma_{x,u}^2 du)(\int_{(i-1)S}^{iS} \sigma_{y,u}^2 du + \int_{(j-1)S}^{jS} \sigma_{y,u}^2 du)}}$$

Noise

We observe efficient prices, (X_t^*, Y_t^*),contaminated by noise

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} X_t^* \\ Y_t^* \end{pmatrix} + \epsilon_t$$

where $(\epsilon_t)_{t\in[0,T]}$ is 2-dimensional i.i.d. process with $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{E}[\epsilon_t \epsilon_t'] = \Psi$.

Observed increment over $[(i-1)S, iS)$:

$$\begin{pmatrix} \Delta_S X_{iS} \\ \Delta_S Y_{iS} \end{pmatrix} = \begin{pmatrix} \Delta_S X_{iS}^* \\ \Delta_S Y_{iS}^* \end{pmatrix} + \epsilon_{iS} - \epsilon_{(i-1)S}$$

for $i = 1, \ldots, n$.

The noise term $\epsilon_{iS} - \epsilon_{(i-1)S}$:

- flips the signs of $\Delta_S X_{iS}$ and $\Delta_S Y_{iS}$;
- drives the estimated correlation to zero.

Observing stale prices

- Rounding error: observed prices are rounded to one cent: for efficient prices X_t we have
$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \alpha\lfloor X_t^*/\alpha \rfloor \\ \alpha\lfloor Y_t^*/\alpha \rfloor \end{pmatrix}$$
with $\alpha = 0.01$.
  - Large bias in volatility estimation for Pearson estimator
  - Zero returns have no information about sign concordance for Quadrant estimators
  - Zero returns flip the rank between two returns on one asset. Example
- Asynchronous trading: prices are not updated simultaneously.
  - Need to synchronize
  - Induce bias (Epps effect).

Simulation

Heston model

We simulate log-prices as, for $\bullet = X, Y$

$$
\begin{aligned}
d\bullet_t^* &= \mu dt + \sigma_t dW_{\bullet,t} \\
d\sigma^2_{\bullet,t} &= \kappa(\bar{\sigma}^2_\bullet - \sigma^2_{\bullet,t})dt + s\sigma_{\bullet,t} dB_{\bullet,t} \\
\sigma^2_{\bullet,0} &\sim \Gamma(2\kappa_\bullet \bar{\sigma}^2_\bullet / s^2, s^2_\bullet / 2\kappa_\bullet)
\end{aligned}
$$

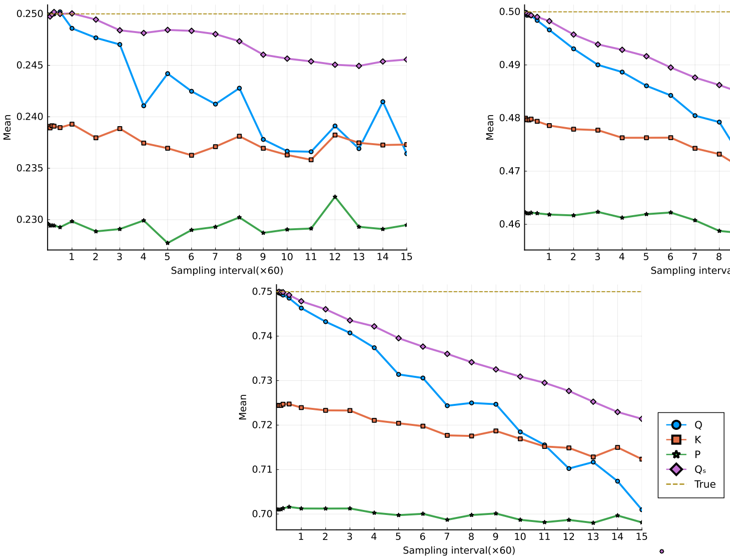
where $\mathrm{corr}(dW_{\bullet,t}, dB_{\bullet,t}) = \varrho$ and $\mathrm{corr}(dW_{\bullet,t}, dW_{\bullet,t}) = \rho$.
The model is calibrated as in Aït-Sahalia et al. (2010):

Table 1: Parameters calibration

	μ	$\bar{\sigma}^2$	κ	s	ϱ
X	0.05	0.16	3	0.8	-0.6
Y	0.03	0.09	2	0.5	-0.75

In a trading day, we let dt be one second and generate (X_t, Y_t) for $t = 0, \ldots, T$ with $T = 23400$.

Correlation signature plot (Heston)



Microstructure issues

- Generate latent log-prices:

$$dX_t^* = \sigma_x dW_{x,t} \text{ and } dY_t^* = \sigma_y dW_{y,t}$$

  where $\mathrm{corr}(dW_{x,t}, dW_{y,t}) = \rho$ and $(\sigma_x, \sigma_y) = (0.15, 0.45)$.

- Independent noise:

$$X_t = X_t^* + \epsilon_{x,t} \text{ and } Y_t = Y_t^* + \epsilon_{y,t}$$

  where $\epsilon_{\bullet,t} \sim N(0, \omega^2)$ and $\omega^2 = \xi^2 \sqrt{T^{-1} \sum_{i=1}^{T} \sigma_{\bullet,i}^4}$ with $\xi^2 = 0.001$, for $\bullet = x, y$.

- Rounding errors: round off the observed prices with tick size one cent.
- Asynchronous trading: for $\bullet = X, Y$

$$\tilde{\bullet}_t = \begin{cases} \lfloor \exp(\bullet_t)/0.01 \rfloor \times 0.01 & \text{with probability } p_\bullet^{\mathrm{tr}} \\ \tilde{\bullet}_{t-1} & \text{otherwise} \end{cases}$$

  where $(p_x^{\mathrm{tr}}, p_y^{\mathrm{tr}}) = (0.8, 0.5)$.

Synchronizing observation

- Time sampling scheme:
  - Tick time sampling (TTS): $\{\tau_{x,k}\}_{k=0}^{N_x}$ are the timestamps at which the asset X is observed. Use $\{\tau_i\}_{i=0}^{N} = \{\tau_{x,k}\}_{k=0}^{N_x} \bigcup \{\tau_{y,k}\}_{k=0}^{N_y}$ as sampling grid. $[\tau_{i-1}, \tau_i]$ refers to a unit sampling interval, and log-returns over it are:
  $$\Delta\tilde{X}_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}}$$
- Synchronization: i-th observation of asset j
  - Previous tick (PT):
  $$\tilde{X}_{\tau_i} = \tilde{X}_{\tau_{x,k}}$$
  where $k = \max(s|\tau_{x,s} \le \tau_i)$.

Correlation signature plot (constant volatility): $\rho = 0.5$

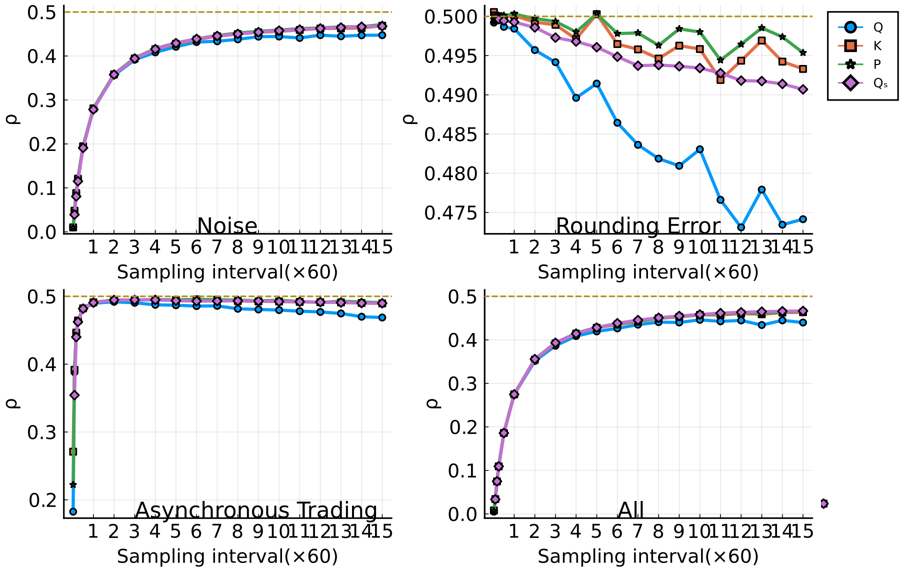


Table 2: Mean squared errors of correlation estimators on returns contaminated by noise, rounding error, and asynchronous trading ($\times 10^{-2}$).

S	1	5	10	15	30	60	120	180	300	600
					$\rho = 0.25$					
Q	6.13	5.49	4.65	4.01	2.85	1.92	1.71	2.13	3.16	5.50
K	6.05	5.45	4.59	3.93	2.65	1.58	1.06	1.12	1.51	2.77
Q_s		5.45	4.55	3.84	2.51	1.35	0.71	0.64	0.85	1.74
P	6.07	5.45	4.57	3.92	2.62	1.54	1.01	1.02	1.33	2.39
					$\rho = 0.5$					
Q	24.50	21.82	18.22	15.43	10.19	5.67	3.29	2.80	2.94	4.53
K	24.22	21.74	18.15	15.35	10.01	5.34	2.58	1.81	1.51	2.11
Q_s		21.79	18.09	15.29	9.90	5.14	2.24	1.40	1.04	1.35
P	24.29	21.72	18.13	15.33	9.95	5.28	2.50	1.70	1.36	1.80
					$\rho = 0.75$					
Q	55.13	49.07	40.97	34.50	22.50	11.99	5.58	3.88	3.04	3.22
K	54.50	48.92	40.80	34.39	22.24	11.68	5.04	3.01	1.75	1.37
Q_s		48.99	40.76	34.37	22.22	11.48	4.77	2.70	1.37	0.92
P	54.64	48.89	40.76	34.34	22.18	11.62	4.96	2.91	1.61	1.12

$T = 23400$ in 5000 replications.

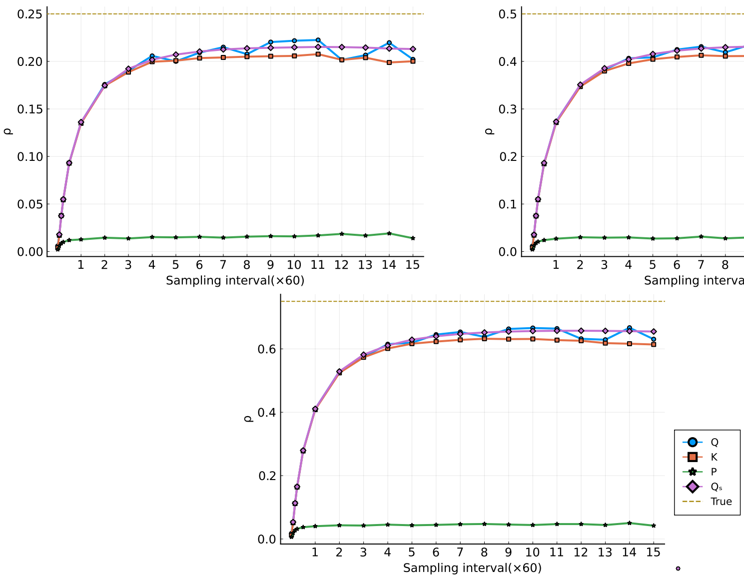
Jumps

We add jumps on the log-prices after noise

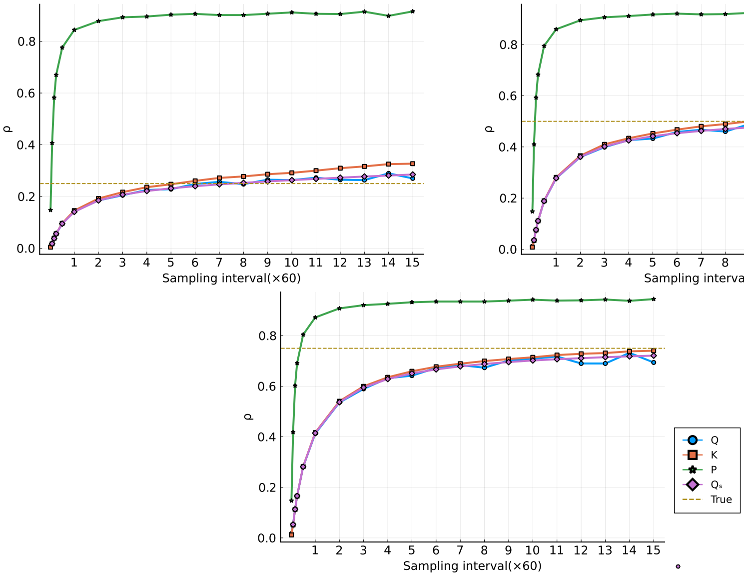
$$\begin{pmatrix} X_t \\ y_t \end{pmatrix} = \begin{pmatrix} X_t^* \\ y_t^* \end{pmatrix} + \epsilon_t + \sum_{s \le t} J_s$$

J_s includes one individual jump on each asset or a cojump. Jump size is uniformly drawn from $[-2, -1] \bigcup [1, 2]/\sqrt{2}$.

Bias: microstructure issues and individual jumps



Bias: microstructure issues and cojump



Microstructure issues and individual jump

Table 3: Mean squared errors of correlation estimators on returns generated in Levy model with microstructure issues and jumps ($\times 10^{-2}$).

S	1	5	10	15	30	60	120	180	300	480
					$\rho = 0.25$					
Q	5.99	5.46	4.63	4.00	2.80	1.94	1.77	2.19	3.22	4.90
K	6.06	5.45	4.56	3.91	2.63	1.62	1.13	1.21	1.60	2.39
Q_s		5.41	4.53	3.82	2.49	1.36	0.76	0.70	0.93	1.48
P	6.15	5.97	5.86	5.82	5.77	5.88	6.05	6.29	6.67	7.65
					$\rho = 0.5$					
Q	23.89	21.61	18.19	15.41	10.25	5.70	3.32	2.89	3.26	4.53
K	24.23	21.70	18.14	15.37	10.11	5.49	2.81	2.09	1.97	2.52
Q_s		21.61	18.07	15.22	9.87	5.20	2.37	1.59	1.25	1.46
P	24.57	23.79	23.27	23.01	22.78	22.63	22.59	22.87	23.48	24.28
					$\rho = 0.75$					
Q	53.67	48.64	40.82	34.44	22.53	12.06	5.71	4.12	3.29	3.63
K	54.47	48.83	40.78	34.44	22.46	11.91	5.43	3.53	2.38	2.31
Q_s		48.56	40.59	34.25	22.18	11.57	5.00	2.99	1.74	1.43
P	55.29	53.50	52.27	51.68	50.95	50.62	50.49	50.81	51.19	51.49

Microstructure issues and cojump

Table 4: Mean squared errors of correlation estimators on returns generated in Levy model with microstructure issues and jumps ($\times 10^{-2}$).

S	1	5	10	15	30	60	120	180	300	480
					$\rho = 0.25$					
Q	5.98	5.48	4.59	3.94	2.73	1.78	1.65	1.98	2.92	5.76
K	6.05	5.44	4.51	3.84	2.51	1.39	0.86	0.88	1.29	2.60
Q_s		5.41	4.51	3.78	2.43	1.26	0.62	0.54	0.73	1.62
P	3.90	7.58	15.85	21.59	30.35	36.75	40.48	41.97	43.14	44.14
					$\rho = 0.5$					
Q	23.87	21.63	18.18	15.36	10.05	5.55	2.98	2.47	2.78	4.43
K	24.20	21.70	18.10	15.24	9.81	5.07	2.26	1.42	1.20	1.77
Q_s		21.58	18.01	15.17	9.74	4.99	2.08	1.23	0.85	1.20
P	15.25	6.00	5.57	7.17	10.91	14.17	16.32	17.16	17.87	18.46
					$\rho = 0.75$					
Q	53.73	48.66	40.69	34.41	22.30	11.89	5.43	3.61	2.67	2.57
K	54.50	48.84	40.65	34.28	22.03	11.35	4.71	2.65	1.33	0.95
Q_s		48.61	40.58	34.15	21.98	11.29	4.64	2.55	1.23	0.73
P	39.12	16.08	6.82	4.26	2.64	2.74	3.25	3.46	3.65	3.79

Correlation signature plot (Heston with microstructure issues)

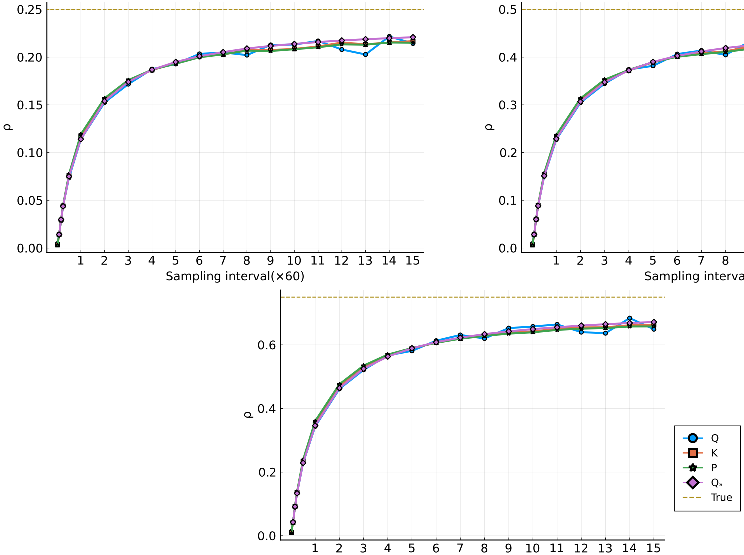


Table 5: Mean squared errors of correlation estimators on returns contaminated by rounding error and asynchronous trading ($\times 10^{-2}$).

S	1	5	10	15	30	60	120	240	420	840
					$\rho = 0.25$					
Q	6.05	5.61	5.01	4.41	3.45	2.53	2.21	2.91	4.60	8.68
K	6.09	5.60	4.92	4.32	3.20	2.09	1.47	1.53	2.23	4.15
Q_s		5.58	4.87	4.27	3.09	1.91	1.10	0.85	1.16	2.43
P	6.11	5.59	4.90	4.28	3.14	2.00	1.38	1.40	1.95	3.52
					$\rho = 0.5$					
Q	24.12	22.28	19.50	17.09	12.58	8.10	5.03	3.73	4.34	6.79
K	24.37	22.33	19.39	16.94	12.18	7.50	4.20	2.59	2.40	3.37
Q_s		22.27	19.37	16.97	12.21	7.44	3.91	2.01	1.50	2.02
P	24.42	22.30	19.30	16.82	11.92	7.23	3.98	2.46	2.25	2.97
					$\rho = 0.75$					
Q	54.27	50.07	43.57	38.11	27.54	17.12	9.35	4.99	3.66	4.39
K	54.84	50.17	43.38	37.81	26.88	16.12	8.34	4.03	2.59	2.39
Q_s		50.05	43.44	38.04	27.22	16.44	8.36	3.77	2.02	1.45
P	54.94	50.10	43.22	37.50	26.30	15.43	7.90	3.84	2.52	2.16

$T = 23400$ in 5000 replications.

Empirical

Data Description

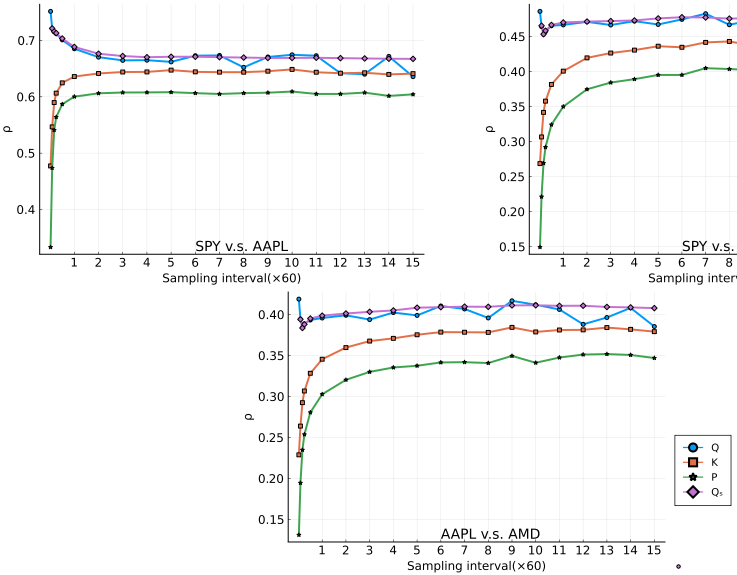
22 stocks:

- S&P 500 constituents
- 11 industry sectors
- Sampled from Jan 1, 2015 till Dec 31, 2021
- Transaction data from NYSE Trade and Quote (TAQ) database

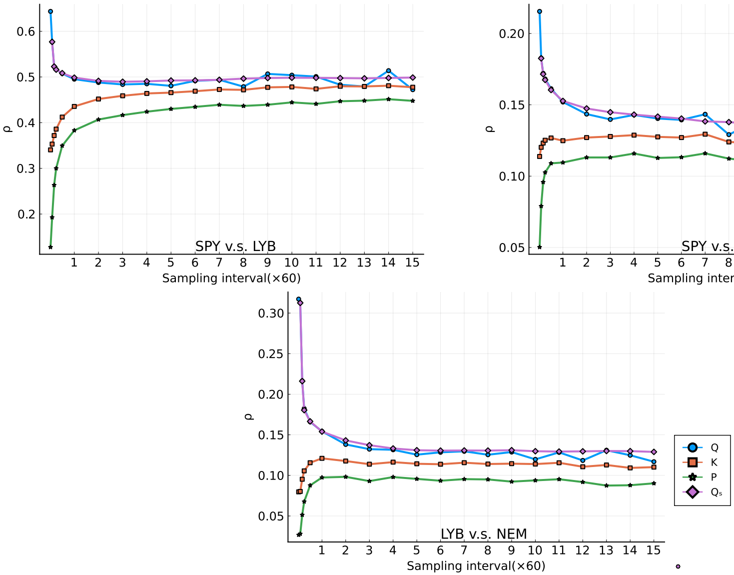
Table 6: Descriptive statistics for the selected S&P 500 stocks and S&P 500 ETF.

Sector	Ticker	N	$E[\tau_i - \tau_{i-1}]$	$\#Zero$	$\hat{\sigma}_\epsilon^2$	$\hat{IV}$
	SPY	12056	2.275	3488	1.71×10^{-9}	6.57×10^{-5}
Utilities	D	3137	8.037	1114	1.92×10^{-8}	1.37×10^{-4}
	DUK	3329	7.467	1132	1.32×10^{-4}	1.35×10^{-4}
Real estate	AMT	2492	9.754	476	3.92×10^{-6}	4.9×10^{-4}
	PLD	2606	9.771	879	3.16×10^{-8}	1.69×10^{-4}
Materials	LYB	2804	9.226	664	5.28×10^{-8}	3.33×10^{-4}
	NEM	4056	6.670	1659	3.55×10^{-8}	3.77×10^{-4}
Information Technology	AAPL	11706	2.271	3067	6.19×10^{-9}	1.86×10^{-4}
	AMD	6139	14.865	2134	2.70×10^{-7}	8×10^{-4}
Industrials	AAL	4088	6.829	1692	5.43×10^{-8}	7.42×10^{-4}
	UNP	3715	6.862	868	2.40×10^{-8}	1.85×10^{-4}
Health Care	JNJ	5523	4.571	1721	8.58×10^{-9}	1.08×10^{-4}
	MRK	5119	5.054	2090	1.11×10^{-8}	1.37×10^{-4}
Financials	JPM	7977	3.217	2887	8.76×10^{-9}	1.78×10^{-4}
	WFC	5772	4.495	2689	1.29×10^{-8}	2.28×10^{-4}
Energy	HAL	4591	5.564	2074	4.79×10^{-8}	6.08×10^{-4}
	XOM	6794	3.831	2753	9.94×10^{-9}	2×10^{-4}
Consumer Staples	PG	5271	4.857	1962	8.44×10^{-9}	1.09×10^{-4}
	WMT	5654	4.461	2116	8.47×10^{-9}	1.18×10^{-4}
Consumer Discretionary	TSLA	7485	5.121	557	5.17×10^{-8}	7×10^{-4}
	AMZN	5869	5.028	393	2×10^{-8}	2.11×10^{-4}
Communication Services	DIS	6273	4.193	1819	1.01×10^{-8}	1.71×10^{-4}
	FB	9225	2.805	1891	1.01×10^{-8}	2.38×10^{-4}

Information Technology: correlation estimates



Materials: correlation estimates



Market β

Recall Capital Asset Pricing Model (CAPM):

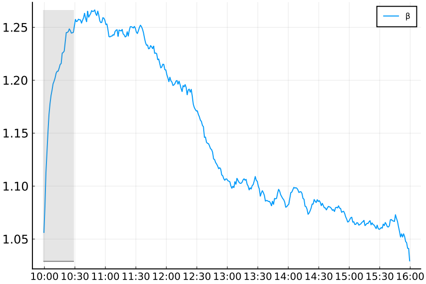
$$\mathbb{E}r_i = r_f + \beta(\mathbb{E}R - r_f)$$

where

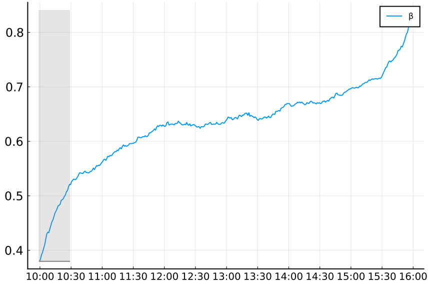
- $\mathbb{E}r_i$: expected return on asset i
- r_f: risk-free rate
- $\mathbb{E}R$: expected market return

Andersen, Thyrsgaard, and Todorov (2021) (ATT) demonstrate variation in β within a trading day.

Intraday time-varying β



(a) CAT v.s. SPY



(b) JNJ v.s. SPY

Figure 16: Average intraday β's for 1 minute returns of stocks, Caterpillar and Johnson&Johnson, over years 2015 - 2021. β's are estimated using 30-minute window.

Intraday time-varying β

Let Y_t and X_t be prices of asset of interest and market respectively. ATT estimates intraday β_τ's using window over lS seconds, for $\tau = l, \dots, n$

$$\beta_\tau = \frac{\sum_{i=\tau-l+1}^{\tau} \Delta_S X_{iS} \Delta_S Y_{iS} \mathbf{1}_{\{|\Delta_S X_{iS}| \leq v_x, |\Delta_S Y_{iS}| \leq v_y\}}}{\sum_{i=\tau-l+1}^{\tau} \Delta_S X_{iS}^2 \mathbf{1}_{\{|\Delta_S Y_{iS}| \leq v_y, |\Delta_S X_{iS}| \leq v_x\}}}$$

where

- $\Delta_S X_{iS} = X_{iS} - X_{(i-1)S}$ and $\Delta_S Y_{iS} = Y_{iS} - Y_{(i-1)S}$
- v_x and v_y are thresholds to avoid jumps.

Intraday correlation

Note

$$\beta_\tau = \rho_\tau \times \frac{\sigma_{y,\tau}}{\sigma_{x,\tau}}$$

- ρ_τ is the correlation between asset Y and market X, estimated with Quadrant, Kendall, subsampled Quadrant, and Pearson.
  - Subsampled Quadrant uses all possible returns over S sampling intervals

    $$\Delta_S X_i = X_i - X_{i-S} \text{ and } \Delta_S Y_i = Y_i - Y_{i-S}$$

    for $i = (\tau - l + 1)S, \ldots, \tau S$.
  - Quadrant, Kendall, and Pearson use non-overlapping returns over S sampling intervals.

Intraday relative volatility

- $\frac{\sigma_{y,\tau}}{\sigma_{x,\tau}}$ is the relative volatility between asset Y and market X.
  - Realized variance:

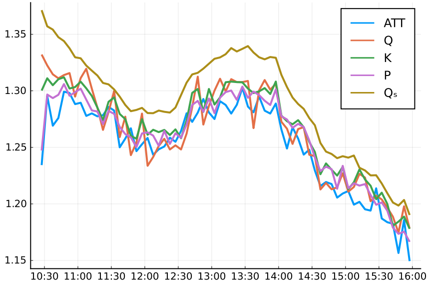
$$\sqrt{\frac{\sum_{i=\tau-l+1}^{\tau} \Delta_S Y_{iS}^2 \mathbf{1}_{\{|\Delta_S X_{iS}| \leq v_x, |\Delta_S Y_{iS}| \leq v_y\}}}{\sum_{i=\tau-l+1}^{\tau} \Delta_S X_{iS}^2 \mathbf{1}_{\{|\Delta_S X_{iS}| \leq v_x, |\Delta_S Y_{iS}| \leq v_y\}}}}$$

  - Local relative volatility:

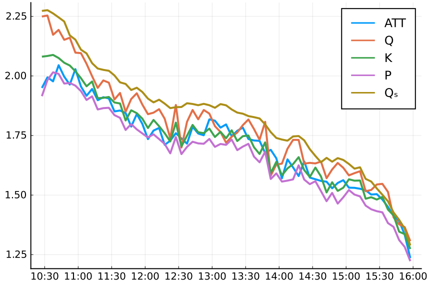
$$\frac{\sum_{i=(\tau-l+1)S}^{\tau S} |\Delta_S Y_i| \mathbf{1}_{\{|\Delta_S Y_i| \leq v_y, |\Delta_S X_i| \leq v_x\}}}{\sum_{i=(\tau-l+1)S}^{\tau S} |\Delta_S X_i| \mathbf{1}_{\{|\Delta_S Y_i| \leq v_y, |\Delta_S X_i| \leq v_x\}}}$$

Figure

Intraday time-varying β



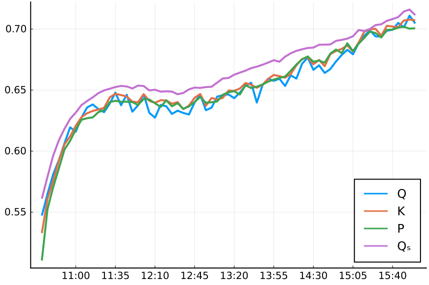
(a) AAPL v.s. SPY



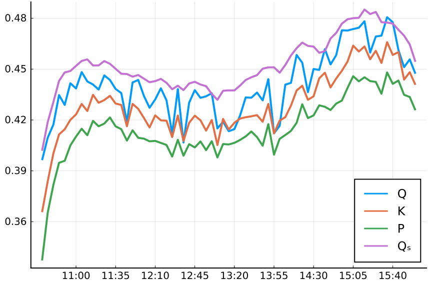
(b) AMD v.s. SPY

Figure 17: Average β estimates by time of day using returns in the previous one-hour window with $S = 180$.

Intraday time-varying correlation



(a) AAPL v.s. SPY

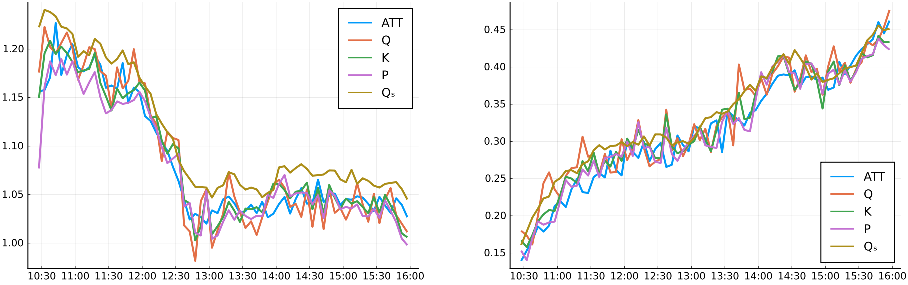


(b) AMD v.s. SPY

Figure 18: Average β estimates by time of day using returns in the previous one-hour window with $S = 180$.

Relative volatility

Intraday time-varying β

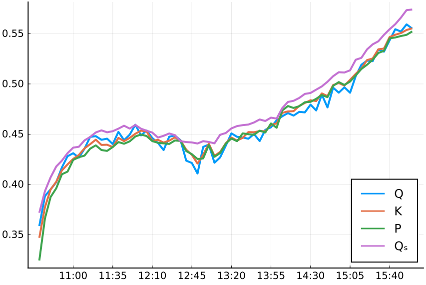


(a) LYB v.s. SPY

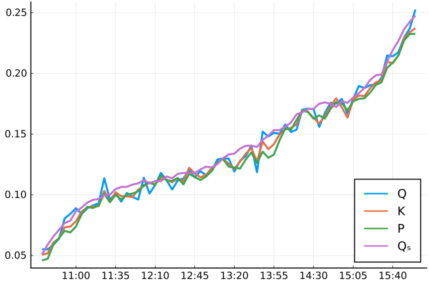
(b) NEM v.s. SPY

Figure 19: Average β estimates by time of day using returns in the previous one-hour window with $S = 180$.

Intraday time-varying correlation



(a) LYB v.s. SPY



(b) NEM v.s. SPY

Figure 20: Average β estimates by time of day using returns in the previous one-hour window with $S = 180$.

Relative volatility

Positive Definiteness Correction

High dimensional case ($d > 2$)

- Estimate correlation matrix by plugging bivariate correlations with Quadrant, Kendall, Subsampled Quadrant estimators
- Cannot guarantee positive definiteness
- Compare four correction methods

Correction for positive definiteness

- Higham's method algorithm (Higham (2002)): find the nearest valid correlation matrix in terms of Frobenius norm by projection.
- Quadratically convergent Newton method: see Qi and Sun (2006).
- Maximization of composite quasi-likelihood:
  - Composite quasi-likelihood of a correlation matrix C based on matrix estimate $\hat{R}$ is defined as follows

$$\mathrm{QLike}^{\mathrm{comp}}(\mathrm{C}) = -\frac{1}{2}\sum_{\mathrm{i}<\mathrm{j}} \log(1 - \mathrm{C}_{\mathrm{ij}}^2) - \sum_{\mathrm{i}<\mathrm{j}} \frac{1 - \hat{\mathrm{R}}_{\mathrm{ij}}\mathrm{C}_{\mathrm{ij}}}{1 - \mathrm{C}_{\mathrm{ij}}^2}.$$

  - We find the nearest correlation matrix to $\hat{R}$ by

$$\max_{\gamma \in \mathbb{R}^{d(d-1)/2}} \mathrm{QLike}^{\mathrm{comp}}(C(\gamma)), \text{ where } \gamma = \mathrm{vech}(\log(C))$$

  C is recovered from a given γ using the result in Archakov and Hansen (2021).

Appendix

Conclusion

- Introduce robust correlation estimator, subsampled Quadrant estimator.
- Subsampled Quadrant estimator is most accurate while facing microstructure issues and is consistent as Quadrant in time-varying volatility models.
- Quadrant, Kendall, and subsampled Quadrant estimators are more robust to the presence of jumps.
- Empirically, find different converging patterns between Quadrant like correlation estimates and others as lowering sampling frequency and compare the estimated intraday time-varying β.

Example

$$X_1 = 101.005, \quad X_2 = 101.001, \quad X_3 = 100.998$$

then

$$\log(X_1) - \log(X_2) > \log(X_2) - \log(X_3)$$

After rounding

$$X_1^* = 101, \quad X_2^* = 101, \quad X_3^* = 100.99$$

and

$$\log(X_1^*) - \log(X_2^*) = 0 < \log(X_2)^* - \log(X_3^*)$$

back

Linear statistics defined on subseries

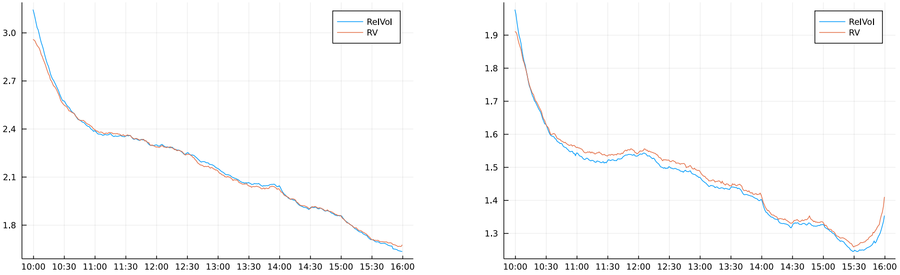
Consider subseries $Z_i = (X_i, \ldots, X_{i+m-1})$ and suppose that $T_i = t(Z_i)$ is an estimate for some parameter in the distribution of X_i's.

To approximate the distribution of $\bar{T}_N = \sum_{i=1}^{N} T_i/N$, we may apply bootstrap resampling on $\{T_i\}_{i=1}^{N}$ as applied in the moving blocks sampling scheme by Künsch (1989), Liu and Singh (1992), and Politis and Romano (1993).

Suppose that $T_i(\omega)$ is the periodogram evaluated at ω based on data Z_i, then $\bar{T}_N(\omega)$ is approximately equal to Bartlett's Kernel estimate of $f(\omega)$.

back

Intraday relative volatility

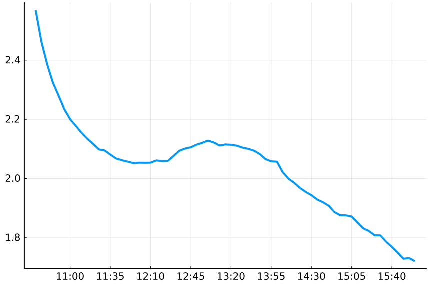


(a) CAT v.s. SPY (b) JNJ v.s. SPY

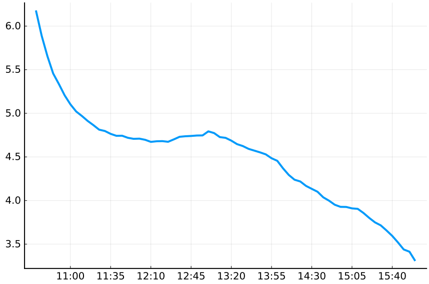
Figure 21: Plot the average intraday relative volatility for 1 minute returns of stocks, Caterpillar, Johnson&Johnson, and market, S&P 500, over year 2015 till 2021. Relative volatility is estimated using 30-minute window.

back

Intraday relative volatility



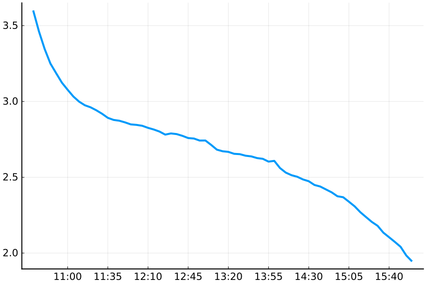
(a) AAPL v.s. SPY



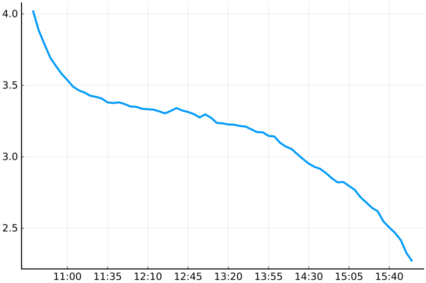
(b) AMD v.s. SPY

back

Intraday relative volatility



(a) LYB v.s. SPY



(b) NEM v.s. SPY

back

Correction for positive definiteness

- **Higham (2002)'s algorithm**:
  Define sets

$$S = \{A = A' \in \mathbb{R}^{d \times d} : A \text{ is p.s.d.}\},$$
$$U = \{A = A' \in \mathbb{R}^{d \times d} : a_{ii} = 1\}.$$

Project a symmetric matrix A onto U with respect to W-norm:

$$P_U(A) = A - W^{-1}\mathrm{diag}(\theta_i)W^{-1}$$

where $\theta = (W^{-1} \circ W^{-1})^{-1}\mathrm{diag}(A - I)$.
Project A onto S with respect to W-norm:

$$P_S(A) = W^{-1/2}[(W^{1/2}AW^{1/2})_+]W^{-1/2}$$

where $A_+ = Q\mathrm{diag}(\max(\lambda_i, 0))Q'$.
To find the nearest matrix to A at the intersection of S and U, we repeat the operation

$$A \leftarrow P_U(P_S(A)).$$

Correction for positive definiteness

- Quadratically convergent Newton method:
  Using this method to solve problem:

$$\min \frac{1}{2}||\hat{R} - C||^2$$
$$\text{s.t. } C \in \mathcal{C} = S \wedge U$$

  where $||\cdot||$ denotes the Frobenius norm, see Qi and Sun (2006).
- Using the similar method to solve the problem of H-weighted type:

$$\min \frac{1}{2}||H \circ (\hat{R} - X)||^2$$
$$\text{s.t. } C \in \mathcal{C}$$

  where $\circ$ denotes the Hadarmard product, and H is symmetric with nonnegative entries.

Correction for positive definiteness

- Example: Consider $d = 5$,

$$R_0 = \begin{pmatrix} 1 & & & & \\ 0.098 & 1 & & & \\ 0.288 & 0.729 & 1 & & \\ 0.593 & -0.504 & -0.031 & 1 & \\ 0.314 & -0.244 & -0.543 & -0.157 & 1 \end{pmatrix}, \text{ eig} = \begin{pmatrix} 0.0005 \\ 0.121 \\ 1.132 \\ 1.651 \\ 2.097 \end{pmatrix}$$

$$\hat{R}_Q = \begin{pmatrix} 1 & & & & \\ 0.044 & 1 & & & \\ 0.309 & 0.675 & 1 & & \\ 0.657 & -0.525 & 0.044 & 1 & \\ 0.249 & -0.200 & -0.593 & -0.181 & 1 \end{pmatrix}, \text{ eig} = \begin{pmatrix} -0.002 \\ 0.101 \\ 1.087 \\ 1.775 \\ 2.038 \end{pmatrix}$$

Higham's algorithm:

$$\hat{R}_{Higham} = \begin{pmatrix} 1 & & & & \\ 0.019 & 1 & & & \\ 0.306 & 0.675 & 1 & & \\ 0.653 & -0.525 & 0.044 & 1 & \\ 0.246 & -0.200 & -0.593 & -0.181 & 1 \end{pmatrix}, \text{ eig} = \begin{pmatrix} 0.010 \\ 0.100 \\ 1.076 \\ 1.774 \\ 2.040 \end{pmatrix}$$

Correction for positive definiteness

- Example (cont'):

$$R_0 = \begin{pmatrix} 1 & & & & \\ 0.098 & 1 & & & \\ 0.288 & 0.729 & 1 & & \\ 0.593 & -0.504 & -0.031 & 1 & \\ 0.314 & -0.244 & -0.543 & -0.157 & 1 \end{pmatrix}, \ \mathrm{eig} = \begin{pmatrix} 0.0005 \\ 0.121 \\ 1.132 \\ 1.651 \\ 2.097 \end{pmatrix}$$

$$\hat{R}_Q = \begin{pmatrix} 1 & & & & \\ 0.044 & 1 & & & \\ 0.309 & 0.675 & 1 & & \\ 0.657 & -0.525 & 0.044 & 1 & \\ 0.249 & -0.200 & -0.593 & -0.181 & 1 \end{pmatrix}, \ \mathrm{eig} = \begin{pmatrix} -0.002 \\ 0.101 \\ 1.087 \\ 1.775 \\ 2.038 \end{pmatrix}$$

Maximizing composite QLike:

$$\hat{R}_{comp.} = \begin{pmatrix} 1 & & & & \\ 0.044 & 1 & & & \\ 0.309 & 0.675 & 1 & & \\ 0.656 & -0.525 & 0.045 & 1 & \\ 0.248 & -0.199 & -0.593 & -0.180 & 1 \end{pmatrix}, \ \mathrm{eig} = \begin{pmatrix} 0.0001 \\ 0.101 \\ 1.086 \\ 1.775 \\ 2.038 \end{pmatrix}$$

Correction for positive definiteness

Table 7: Comparison the accuracy and computation costs of corrected Quadrant and Kendall estimators via four methods ($n = 5$)

Loss/Time	Higham	Comp. QLike	Newton	H-Newton
		Quadrant		
Frobenius	0.0856	0.0841	0.0839	0.0840
Stein's	4.6339	4.1876	4.2304	4.2322
Time (sec)	0.0569	6.7361	0.0007	0.0026
		Kendall		
Frobenius	0.0573	0.0533	0.0533	0.0533
Stein's	4.2331	0.8381	0.9210	0.9210
Time (sec)	2.1019	6.9426	0.0005	0.0018
		Pearson		
Frobenius		0.0504		
Stein's		0.1648		

Frobenius norm: $||A||_F = \sqrt{\sum_{i,j} a_{i,j}^2}$.

Stein's loss function: $L(\hat{R}) = \mathrm{tr}(R_0^{-1}\hat{R}) - log(|R_0^{-1}\hat{R}|) - d$.

Reference I

Y. Aït-Sahalia, J. Fan, and D. Xiu. High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517, 2010.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55, 2001.

T. G. Andersen, M. Thyrsgaard, and V. Todorov. Recalcitrant betas: Intraday variation in the cross-sectional dispersion of systematic risk. *Quantitative Economics*, 12(2):647–682, 2021.

F. M. Bandi and J. R. Russell. Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2):339–369, 2008.

O. E. Barndorff-Nielsen and N. Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37, 2004.

O. E. Barndorff-Nielsen and N. Shephard. *Variation, Jumps, and High-Frequency Data in Financial Econometrics*, volume 3 of *Econometric Society Monographs*, pages 328–372. Cambridge University Press, 2007.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169, 2011.

N. Blomqvist. On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, pages 593–600, 1950.

K. Boudt, C. Croux, and S. Laurent. Outlyingness weighted covariation. *Journal of Financial Econometrics*, 9(4):657–684, 2011.

K. Boudt, J. Cornelissen, and C. Croux. The gaussian rank correlation estimator: robustness properties. *Statistics and Computing*, 22(2):471–483, 2012.

P. Chang, E. Pienaar, and T. Gebbie. The epps effect under alternative sampling schemes. *Physica A: Statistical Mechanics and its Applications*, 583:126329, 2021.

K. Christensen, S. Kinnebrock, and M. Podolskij. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of econometrics*, 159(1):116–133, 2010.

Reference II

K. Christensen, M. Podolskij, and M. Vetter. On covariation estimation for multivariate continuous itô semimartingales with noise in non-synchronous observation schemes. *Journal of Multivariate Analysis*, 120:59–84, 2013.

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

T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2): 359–379, 2005.

N. J. Higham. Computing the nearest correlation matrix-a problem from finance. *IMA journal of Numerical Analysis*, 22(3): 329–343, 2002.

J. Jacod, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276, 2009.

M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

Y. Li, Z. Zhang, and Y. Li. A unified approach to volatility estimation in the presence of both rounding and random market microstructure noise. *Journal of Econometrics*, 203(2):187–222, 2018.

M. C. Münnix, R. Schäfer, and T. Guhr. Statistical causes for the epps effect in microstructure noise. *International Journal of Theoretical and Applied Finance*, 14(08):1231–1246, 2011.

M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658, 2009.

O. V. Precup and G. Iori. Cross-correlation measures in the high-frequency domain. *European Journal of Finance*, 13(4): 319–331, 2007.

H. Qi and D. Sun. A quadratically convergent newton method for computing the nearest correlation matrix. *SIAM journal on matrix analysis and applications*, 28(2):360–385, 2006.

R. Renò. A closer look at the epps effect. *International Journal of theoretical and applied finance*, 6(01):87–102, 2003.

M. Rosenbaum. Integrated volatility and round-off error. *Bernoulli*, 15(3):687–720, 2009.

P. J. Rousseeuw, F. R. Hampel, E. M. Ronchetti, and W. A. Stahel. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

Reference III

B. Tóth and J. Kertész. Modeling the epps effect of cross correlations in asset prices. In *Noise and Stochastics in Complex Systems and Finance*, volume 6601, pages 89–97. SPIE, 2007.

H. Vander Elst and D. Veredas. Smoothing it out: Empirical and simulation results for disentangled realized covariances. *Journal of Financial Econometrics*, 15(1):106–138, 2016.

L. Zhang. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47, 2011.

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