# Realized Principal Component Analysis of Noisy High-Frequency Data

Bezirgen Veliyev (Aarhus University, Denmark)
joint with F. Benvenuti (Aalborg) and K. Christensen (Aarhus)

WU Vienna, December 3, 2025.

# Outline

# PCA in High-Frequency Settings

## PCA

- Given a high-dimensional dataset, principal components analysis (PCA) is a tool to identify a small number of common factors that summarize a large portion of the variation in the dataset.
- PCA has been well researched in i.i.d. and stationary time series settings.
- Ait-Sahalia and Xiu (2019, JASA) develop a high-frequency version of PCA and its asymptotic theory under very general conditions.
- The dimension of assets $d$ is large but fixed.
- An absence of microstructure noise is assumed.

# General Microstructure Noise Setting

- High-frequency data are contaminated by microstructure noise (bid–ask bounce, discrete pricing, order-book frictions).
- Plain realized measures typically use sparse sampling (e.g. every 5 or 1 minutes), effectively discarding most tick data.
- To exploit tick-by-tick high-frequency data, noise-robust estimators are necessary; initial wave of noise robust estimators were proposed during 2005–2010.
- Jacod, Li, Zheng (2017, ECTA) propose a general noise framework that accommodates serial dependence, heteroskedasticity in noise, and random, endogenous observation times.
- Jacod, Li, Zheng (2019) and Li and Linton (2024) develop pre-averaging estimators for integrated volatility and spot volatility, respectively.

# Our contributions

**Multivariate Spot Covariance**

- Under the general microstructure noise setting (excluding random observation times), we propose a multivariate spot covariance estimator.
- We extend Li and Linton (2024) from the univariate to the multivariate context.
- To the best of our knowledge, this is the first noise-robust spot covariance estimator in the literature under such generality.

**Realized PCA**

- Using this spot covariance estimator, we extend realized principal component analysis of Ait-Sahalia and Xiu (2019).
- In this regard, we develop a noisy version of a Jacod and Rosenbaum (2013)-type estimator for volatility functionals.

# Methodological Challenges

## Eigenvalue Separation

- We have to impose spatial localization condition to ensure that eigenvalues and eigenvectors of the spot covariance matrix are locally well-defined and unique.

## Theoretical Challenge and Solution

- Pre-averaging CLT theory relies on big blocks and small blocks to ensure conditional independence: big blocks drive the limiting distribution, while small blocks are negligible.
- For bias correction of eigenvalue functionals, we must aggregate information from $p = p_n$ nearby blocks, and still show that several bias-correction terms are negligible as $n \to \infty$.

# Related literature

**Earlier Noise Robust Estimators of IV/SV**

- Univariate IV: Zhang et al. (2005), Barnorff-Nielsen et al. (2008), Jacod et al. (2009), Xiu (2010), Hautsch and Podolskij (2013), Ikeda (2015), Varneskov (2016), Li et al. (2020), ...
- Multivariate IV: Zhang (2006), Christensen et al. (2010), Barnorff-Nielsen et al. (2011),...
- SV: Zu and Boswijk (2014), Bibinger et al. (2019), Todorov (2019), Figueroa-Lopez and Wu (2024),...

**Recent Estimators under General Noise Setting (Univariate)**

- Jacod, Li, Zheng (2017, 2019), Da, Xiu (2021), Li, Linton (2022, 2024),...

**Principal Component Analysis under High Frequency Data**

- Ait-Sahalia, Xiu (2019), Dai, Lu, Xiu (2019), Chen, Mykland, Zhang (2020), ...

**Generalized Volatility Functionals**

- Mykland, Zhang (2009), Jacod, Rosenbaum (2013), Li, Xiu (2016), Li, Todorov, Tauchen (2016, 2017), Chen (2019), ...

# Recommended References for Students

**High-Frequency Econometrics**

- **Ait-Sahalia & Jacod (2014)**
  *High-Frequency Financial Econometrics*.
  The standard reference on noise-robust estimators, pre-averaging, jumps, asymptotic theory, and empirical implementation.

**Semimartingale Theory and Discretization**

- **Jacod & Protter (2012)**
  *Discretization of Processes*.
  A fundamental text on limit theorems for semimartingales under high-frequency observation. Essential background for modern HF econometrics.

# Efficient Price Process

We model the latent efficient log-price $X$ as a $d$-dimensional Itô semimartingale on $[0, T]$:

$$X_t = X_0 + \int_0^t b_s \, ds + \int_0^t \sigma_s \, dW_s + \int_{[0,t]\times\mathbb{R}^d} \delta(s,z)\, \mu(ds, dz).$$

- $b_s$ : drift process (optional, $\mathbb{R}^d$-valued)
- $\sigma_s$ : spot volatility matrix ($d \times d$)
- $W$ : $d$-dimensional Brownian motion
- $\mu$ : Poisson random measure with intensity $\nu(ds, dz) = ds \otimes \bar{\nu}(dz)$
- $\delta(s, z)$ : jump size of $X$ at time $s$ with mark $z$

**Interpretation:** The model allows for:

- Stochastic drift and volatility
- Finite or infinite activity jumps
- Arbitrary jump sizes and arrival patterns

# Assumptions on Drift, Volatility, and Jumps

## Assumption (V)

*The drift and volatility processes admit continuous Itô semimartingale:*

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s \, ds + \int_0^t \tilde{\sigma}_s \, d\tilde{W}_s,$$

$$b_t = b_0 + \int_0^t \check{b}_s \, ds + \int_0^t \check{\sigma}_s \, d\check{W}_s,$$

*where $\tilde{W}$ and $\check{W}$ are Brownian motions of dimensions $d^2$ and $d$, respectively.*

*There exists $r \in [0,1)$ and a localization sequence $\tau_n \uparrow \infty$ such that:*

$$\|\delta(t,z)\|^r \wedge 1 \leq J_n(z), \qquad \|U_t\| \leq n, \quad U \in \{\tilde{b}, \tilde{\sigma}, \check{b}, \check{\sigma}\}, \ t \leq \tau_n.$$

**Meaning:**

- Volatility and drift evolve smoothly (continuous semimartingales)
- Jumps of $X$ satisfy a summability condition ($r < 1$)
- Sufficient regularity for stable central limit theorems

# Objects of Interest

Our analysis is built around $d \times d$ dimensional **spot covariance process**

$$c_t = \sigma_t \sigma_t^\top.$$

**Eigenvalues/eigenvectors of $c_s$**

- $\lambda_1(c_s) > \lambda_2(c_s) > \cdots > \lambda_d(c_s)$ : ordered eigenvalues,
- $\gamma_r(c_s)$ : eigenvector associated with $\lambda_r(c_s)$.

**Targets**

$$\begin{aligned} \text{Integrated eigenvalue:} \qquad & V(\lambda_r)_t = \int_0^t \lambda_r(c_s)\, ds, \\ \text{Integrated eigenvector:} \qquad & R(\lambda_r)_t = \int_0^t \gamma_r(c_s)\, ds, \\ \text{Principal component process:} \qquad & PC(\lambda_r)_t = \int_0^t \gamma_r(c_s)^\top dX_s. \end{aligned}$$

**This talk focuses on estimating $V(\lambda_r)_t$.** Other two objects follow analogously from the same theory.

# High-Frequency Asymptotics

**Infill asymptotics:**

- Process $X_t$ defined on a *fixed* time interval $[0, T]$,
- Discrete observations $X_{i\Delta_n}$ with sampling $\Delta_n \to 0$ (frequency $\to \infty$).

**Continuous semimartingale**

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s.$$

**Realized volatility (RV):**

$$RV_t^n = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} (X_{i\Delta_n} - X_{(i-1)\Delta_n})^2.$$

**Key facts**

- RV converges to integrated volatility:

$$RV_t^n \xrightarrow{\mathbb{P}} \int_0^t \sigma_s^2 \, ds.$$

- Central limit theorem:

$$\Delta_n^{-1/2}(RV_t^n - \int_0^t \sigma_s^2 ds) \xrightarrow{ds} MN\left(0,\, 2\int_0^t \sigma_s^4 ds\right).$$

# Idea of Jacod and Rosenbaum (2013)

**Goal:** Estimate nonlinear functionals of spot volatility,

$$V(g)_t = \int_0^t g(\sigma_s^2)\,ds, \qquad g \in C^3.$$

**Key idea:** Construct a *local* estimator of $\sigma_s^2$ by averaging over a short window.

$$\widehat{c}_i^n = \frac{1}{k_n\Delta_n}\sum_{j=0}^{k_n-1}(X_{(i+j)\Delta_n} - X_{(i+j-1)\Delta_n})^2, \qquad k_n \to \infty, \quad k_n\Delta_n \to 0.$$

**Plug-in estimator:**

$$\widetilde{V}(g)_t^n = \Delta_n \sum_{i=0}^{\lfloor t/\Delta_n\rfloor - k_n + 1} g(\widehat{c}_i^n),$$

consistent for $V(g)_t$, but with a second-order bias.

**Bias-corrected estimator:**

$$\widehat{V}(g)_t^n = \Delta_n \sum_{i}^{\lfloor t/\Delta_n\rfloor - k_n + 1} \left[ g(\widehat{c}_i^n) - \frac{1}{k_n} g''(\widehat{c}_i^n)(\widehat{c}_i^n)^2 \right].$$

**CLT:** If $k_n^3\Delta_n \to \infty$, $k_n^2\Delta_n \to 0$, then $\Delta_n^{-1/2}\left(\widehat{V}(g)_t^n - V(g)_t\right) \Longrightarrow MN(0,\ \Sigma_t)$.

# Univariate Pre-Averaging for Integrated Volatility

We observe noisy prices

$$Y_{i\Delta_n} = X_{i\Delta_n} + \epsilon_i, \qquad \mathbb{E}[\epsilon_i] = 0,\ \mathbb{E}[\epsilon_i^2] = \omega^2,\ (\epsilon_i) \text{ i.i.d. and independent of } X.$$

**Target:**

$$IV = \int_0^1 \sigma_t^2 \, dt = \text{integrated volatility of the efficient price.}$$

**Pre-averaged returns**

$$\bar{Y}_i^n = \sum_{j=1}^{k_n - 1} \varphi(j/k_n)(Y_{(i+j)\Delta_n} - Y_{(i+j-1)\Delta_n}), \qquad k_n \sim \theta/\sqrt{\Delta_n},\ \varphi \text{ is a kernel function}$$

**Pre-averaging estimator (Jacod et al. 2009):**

$$\widehat{IV}^{PA} = \frac{1}{\theta\psi_2\sqrt{\Delta_n}} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor - k_n} (\bar{Y}_i^n)^2 - \frac{\psi_1}{\theta^2\psi_2}\widehat{\omega}^2, \quad \widehat{\omega}^2 = \frac{1}{2\lfloor 1/\Delta_n \rfloor} \sum_{i=1}^{\lfloor 1/\Delta_n \rfloor} \left(Y_{i\Delta_n} - Y_{(i-1)\Delta_n}\right)^2.$$

**Intuition:**

- Pre-averaging smooths high-frequency noise: $\bar{\epsilon}_i^n = O_p(\Delta_n^{1/4})$.
- Squared pre-averaged returns estimate $c_1 IV + c_2\omega^2$.
- A bias correction removes the effect of microstructure noise.

# Pre-averaging: notation and constants

Pre-averaging relies on a kernel function

$$\varphi : [0,1] \to \mathbb{R}, \qquad \varphi(0) = \varphi(1) = 0,$$

which smooths high-frequency returns and controls the noise–signal tradeoff.

The choice of $\varphi$ determines several constants that appear in the bias correction and the asymptotic variance of the estimator:

$$\phi_1(t) = \int_t^1 \varphi'(u)\varphi'(u-t)\mathrm{d}u, \quad \phi_2(t) = \int_t^1 \varphi(u)\varphi(u-t)\mathrm{d}u,$$

$$\psi_1 = \phi_1(0), \quad \psi_2 = \phi_2(0),$$

$$\Phi_{11} = \int_0^1 \phi_1^2(t)\mathrm{d}t, \quad \Phi_{12} = \int_0^1 \phi_1(t)\phi_2(t)\mathrm{d}t, \quad \Phi_{22} = \int_0^1 \phi_2^2(t)\mathrm{d}t.$$

**Explicit constants for triangular kernel** $\varphi(x) = \min(x, 1-x)$,:

$$\psi_1 = 1, \qquad \psi_2 = \frac{1}{12},$$

$$\Phi_{11} = \frac{1}{6}, \qquad \Phi_{12} = \frac{1}{96}, \qquad \Phi_{22} = \frac{151}{80640}.$$

# General Microstructure Noise Model

Observed log-prices are contaminated by noise:

$$Y_{i\Delta_n} = X_{i\Delta_n} + \epsilon_{i\Delta_n}.$$

**Assumption (N)**

*For each asset $j = 1, \ldots, d$,*

$$\epsilon^j_{i\Delta_n} = \omega^j_{i\Delta_n} \chi^j_i,$$

*where:*

- $\omega^j_t$ *is a continuous Itô semimartingale (time-varying noise level),*
- $\chi^j_i$ *is a stationary, mean-zero, unit-variance sequence,*
- $(\chi^j_i)$ *is v-polynomially* $\rho$*-mixing with* $v > 4$ *(serial dependence),*
- $\chi$ *has finite moments of all orders, and* $\chi$ *is independent of* $(X, \omega)$*.*

Noise autocovariance is $r(\ell) = \mathbb{E}[\chi_0 \chi_\ell^\top]$, and its long-run covariance matrix is

$$\mathcal{R} = \sum_{\ell=-\infty}^{\infty} r(\ell).$$

# Asymptotic Variance

**Without noise:**

- Asymptotic variance involves quarticity $2\sigma_t^4$.
- In the multivariate case: $c_t^{jl} c_t^{km} + c_t^{jm} c_t^{kl}$.

**With general microstructure noise:**

- The limit variance involves both the price covariance $c_t$ and the noise covariance $\omega_t \mathcal{R} \omega_t$,
- plus *mixed* terms capturing signal–noise interactions.

Precisely, the limiting variance kernel is

$$\Gamma(c_t,\ \omega_t \mathcal{R} \omega_t)^{jk,lm},$$

where for arbitrary $x, y \in \mathbb{R}^{d\times d}$,

$$\Gamma(x,y)^{jk,lm} = \frac{2}{\psi_2^2}\left[\theta \Phi_{22} \Lambda(x)^{jk,lm} + \frac{\Phi_{12}}{\theta} \Xi(x,y)^{jk,lm} + \frac{\Phi_{11}}{\theta^3} \Lambda(y)^{jk,lm}\right],$$

$$\Xi(x,y)^{jk,lm} = x^{jl} y^{km} + x^{jm} y^{kl} + x^{km} y^{jl} + x^{kl} y^{jm},$$

$$\Lambda(x)^{jk,lm} = x^{jl} x^{km} + x^{jm} x^{kl}.$$

# Spot Covariance Estimator

To estimate the instantaneous covariance $c_t$, we smooth pre-averaged returns over a local window of length $h_n$.

Let $N_t^n = \lfloor t/\Delta_n \rfloor$. The spot covariance estimator is

$$\widehat{c}_t^n = \frac{1}{\theta\psi_2\, h_n\sqrt{\Delta_n}} \sum_{u=N_t^n}^{N_t^n+h_n-1} \bar{Y}_u^n(\bar{Y}_u^n)^\top \mathbf{1}\{\|\bar{Y}_u^n\| \leq u_n\} - \frac{1}{\theta^2\psi_2}\widehat{\Psi}_t^n.$$

The long-run noise covariance is estimated by

$$\widehat{\Psi}_t^n = \sum_{\ell=-\ell_n}^{\ell_n} \psi_1(|\ell|)\, U(\ell)_t^n,$$

with

$$U(\ell)_t^n = \frac{1}{2h_n} \sum_{u=N_t^n+m_n}^{N_t^n+m_n+h_n-1} \Big[(Y_u^n - Y_{u-m_n}^n)(Y_{u+\ell}^n - Y_{u+\ell+m_n}^n)^\top + \mathrm{sym.}\Big].$$

This generalizes the univariate estimator of Li & Linton (2024).

# Tuning Parameters and Rate Notation

The spot covariance estimator involves several smoothing and truncation parameters. We express their asymptotic growth using exponents:

$$h_n \asymp \Delta_n^{-h}, \qquad \ell_n \asymp \Delta_n^{-\ell}, \qquad m_n \asymp \Delta_n^{-m}, \qquad u_n \asymp (k_n\Delta_n)^{\varpi}.$$

- $h_n$: local window size for spot smoothing; larger $h_n$ averages out noise.
- $\ell_n$: range of autocovariances used to estimate long-run noise variance.
- $m_n$: lag used to decorrelate the noise sequence.
- $u_n$: truncation level in the pre-averaging tail control (depends on $\varpi$).

These parameters must grow at compatible rates for the LLN and CLT to hold.

# Spatial Localization

To handle eigenvalues, we must avoid points where eigenvalues collide. When the $r$th eigenvalue is simple, the map $c \mapsto \lambda_r(c)$ is smooth.

**Assumption (SL)**

*There exists an open and convex set $\mathcal{C} \subseteq \mathcal{M}(r)$ such that*

$$c_s \in \mathcal{C}, \qquad 0 \leq s \leq t,$$

*where $\mathcal{M}(r)$ is the set of positive definite covariance matrices with a simple $r$th eigenvalue.*

**Comment:**

This ensures a uniform spectral gap around $\lambda_r(c_s)$, so $\lambda_r$ is differentiable along the path of $c_s$. Therefore Taylor expansions used in bias correction and the CLT are valid.

# Spot Covariance CLT

**Theorem**

*Suppose that the following conditions hold for the rates:*

$$h \in \left(\tfrac{1}{2}, \tfrac{3}{4}\right),\ \ell \in \left(\tfrac{1}{8(v-1)}, \tfrac{1}{6}\right),\ m \in \left(\ell, \tfrac{1}{6}\right),\ \varpi \in \left(\tfrac{1}{4-2r}, \tfrac{2\lfloor v \rfloor - 3}{4\lfloor v \rfloor - 4}\right),\ r < \tfrac{2\lfloor v \rfloor - 4}{2\lfloor v \rfloor - 3}.$$

*Under Assumptions (V) and (N), for any $t \in (0, T]$, as $\Delta_n \to 0$,*

$$h_n^{1/2} \Delta_n^{1/4} \left[\widehat{c}_t^{\,n} - c_t\right] \xrightarrow{d_s} MN(0, Q_t),$$

*where the components of $d^2 \times d^2$-dimensional conditional variance $Q_t$ are given by*

$$Q_t^{jk,lm} = \Gamma\left(c_t, \omega_t \mathcal{R} \omega_t\right)^{jk,lm}.$$

**Comment:**

- This generalizes Christensen, Kinnebrock, Podolskij (2010) to the spot and more general noise setting, and Li and Linton (2024) to multivariate setting.

# Bias Correction

An obvious estimator of integrated eigenvalue $V(\lambda_r)$ is

$$V(\lambda_r)_t^n = \Delta_n \sum_{i=0}^{N_t^n-1} \lambda_r\left(\widehat{c}_i^n\right).$$

This has a second order asymptotic bias in CLT.

**Theorem**

*Suppose that the following rate conditions hold:*

$$h \in \left(\tfrac{1}{2}, \tfrac{3}{4}\right),\ \ell \in \left(\tfrac{1}{4(v-1)}, \tfrac{1}{12}\right),\ m \in \left(\ell, \tfrac{1}{12}\right),\ \varpi \in \left(\tfrac{1}{4-2r}, \tfrac{2\lfloor v\rfloor-3}{4\lfloor v\rfloor-4}\right),\ r < \tfrac{2\lfloor v\rfloor-4}{2\lfloor v\rfloor-3}.$$

*Then, under Assumptions (V), (N) and (SL), for any $t \in (0, T]$, as $\Delta_n \to 0$, it holds that*

$$\Delta_n^{1/2} h_n \left[V(\lambda_r)_t^n - V(\lambda_r)_t\right] \xrightarrow{\mathbb{P}} B(\lambda_r)_t, \tag{1}$$

*where*

$$B(\lambda_r)_t = \frac{1}{2} \sum_{j,k,l,m=1}^{d} \int_0^t \partial^2_{jk,lm} \lambda_r\left(c_s\right) \Gamma\left(c_s, \omega_s \mathcal{R} \omega_s\right)^{jk,lm} \mathrm{d}s.$$

# Bias Corrected Estimator

Its bias-corrected version of the $r$th realized eigenvalue:

$$\widehat{V}(\lambda_r)_t^n = \Delta_n \sum_{i=0}^{N_t^n - 1} \left[ \lambda_r \left( \widehat{c}_i^n \right) - \widehat{B}(\lambda_r)_i^n \right],$$

where

$$\widehat{B}(\lambda_r)_i^n = \frac{1}{2 h_n \Delta_n^{1/2}} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} \lambda_r \left( \widehat{c}_i^n \right) \Gamma \left( \widehat{c}_i^n, \widehat{\Psi}_i^n \right)^{jk,lm}. \qquad (2)$$

# CLT

## Theorem

*Suppose that the following conditions hold for the rates:*

$$h \in \left(\tfrac{23}{31}, \tfrac{3}{4}\right),\ \ell \in \left(\tfrac{1}{4(v-1)}, \tfrac{1}{12}\right),\ m \in \left(\ell, \tfrac{1}{12}\right),\ \varpi \in \left(\tfrac{1}{4-2r}, \tfrac{2\lfloor v \rfloor - 3}{4\lfloor v \rfloor - 4}\right),\ r < \tfrac{2\lfloor v \rfloor - 4}{2\lfloor v \rfloor - 3}.$$

*Then, under Assumptions (V), (N) and (SL), for any* $t \in (0, T]$*, as* $\Delta_n \to 0$*, it holds that*

$$\Delta_n^{-1/4}\left[\widehat{V}(\lambda_r)_t^n - V(\lambda_r)_t\right] \xrightarrow{d_s} MN(0, S(\lambda_r)_t),$$

*where the conditional variance is given by*

$$S(\lambda_r)_t = \int_0^t \sum_{j,k,l,m=1}^{d} \partial_{jk}\lambda_r\left(c_s\right)\partial_{lm}\lambda_r\left(c_s\right)\Gamma\left(c_s, \omega_s \mathcal{R} \omega_s\right)^{jk,lm} \mathrm{d}s.$$

1. Introduction
2. Setting (Target, Estimators, Assumptions)
3. Theoretical Results
4. Simulations
5. Empirical Analysis

# Simulation study

We work with $d = 10$ assets and suppose $r = 3$ factors. We simulate the following $d$-dimensional model:

$$\mathrm{d}X_{i,t} = \sum_{j=1}^{r} \beta_{ij,t} \mathrm{d}F_{j,t} + \mathrm{d}Z_{i,t},$$

where for $i = 1, \dots, d$ and $j = 1, \dots, r$.

$$\mathrm{d}F_{j,t} = \mu_j \mathrm{d}t + \sigma_{j,t} \mathrm{d}W_{j,t} \qquad \text{and} \qquad \mathrm{d}Z_{i,t} = \gamma_{i,t} \mathrm{d}B_{i,t}.$$

In the above, the coefficient processes are described by:

$$\begin{aligned}
\mathrm{d}\sigma^2_{j,t} &= \kappa(\theta - \sigma^2_{j,t})\mathrm{d}t + \eta\sigma_{j,t}\mathrm{d}\widetilde{W}_{j,t}, \\
\mathrm{d}\gamma^2_{i,t} &= \kappa_Z(\theta_Z - \gamma^2_{i,t})\mathrm{d}t + \eta_Z\gamma_{i,t}\mathrm{d}\bar{B}_{i,t}, \\
\mathrm{d}\beta_{i1,t} &= \tilde{\kappa}_1(\tilde{\theta}_{i1} - \beta_{i1,t})\mathrm{d}t + \tilde{\xi}_1\sqrt{\beta_{i1,t}}\mathrm{d}\widetilde{B}_{i1,t}, \\
\mathrm{d}\beta_{ij,t} &= \tilde{\kappa}_j(\tilde{\theta}_{ij} - \beta_{ij,t})\mathrm{d}t + \tilde{\xi}_j\mathrm{d}\widetilde{B}_{ij,t}, \quad j \geq 2.
\end{aligned}$$

# Three Noise Settings

We have

$$\epsilon_i^{n,j} = \omega_i^{n,j} \chi_i$$

where

$$\omega_i^{n,j} = \sqrt{\rho} n \sigma_i^{n,j} \text{ and } \chi_i = \sum_{j=0}^{q} \theta_j u_j,$$

with i.i.d. normal $u_i$ and $\rho = 0.5$

**We consider three types of noise:**

- Uncorrelated Noise with $q = 0$.
- Moving average noise with $q = 100$ following Jacod, Li and Zheng (2017).
- Autoregressive noise with $q = \infty$ and $\theta_j = 0.8^j$ following Li and Linton (2024).

# Simulation Results-Integrated eigenvalue estimation

| Eigenvalue | 1st | 2nd | 3rd |
|---|---|---|---|
| True value | 0.907 | 0.408 | 0.336 |
| Explained variation | 0.242 | 0.109 | 0.090 |
| | | | |
| Panel A: Uncorrelated noise | | | |
| *Consistent estimator* | | | |
| Relative bias | 1.026 | 1.154 | 1.196 |
| Estimation Error | 0.025 | 0.061 | 0.066 |
| Standard deviation | 0.067 | 0.022 | 0.013 |
| | | | |
| *Bias-corrected estimator* | | | |
| Relative bias | 0.993 | 1.045 | 1.113 |
| Estimation Error | -0.002 | 0.017 | 0.038 |
| Standard deviation | 0.069 | 0.024 | 0.016 |

# Simulation Results-MA Noise

| Eigenvalue | 1st | 2nd | 3rd |
|---|---|---|---|
| True value | 0.907 | 0.408 | 0.336 |
| Explained variation | 0.242 | 0.109 | 0.090 |
| | | | |
| Panel B: Moving average noise | | | |
| *Consistent estimator* | | | |
| Relative bias | 1.027 | 1.156 | 1.199 |
| Estimation Error | 0.026 | 0.062 | 0.067 |
| Standard deviation | 0.067 | 0.022 | 0.013 |
| | | | |
| *Bias-corrected estimator* | | | |
| Relative bias | 0.994 | 1.047 | 1.116 |
| Estimation Error | -0.001 | 0.018 | 0.039 |
| Standard deviation | 0.069 | 0.024 | 0.016 |

# Simulation Results-AR Noise

| Eigenvalue | 1st | 2nd | 3rd |
|---|---|---|---|
| True value | 0.907 | 0.408 | 0.336 |
| Explained variation | 0.242 | 0.109 | 0.090 |

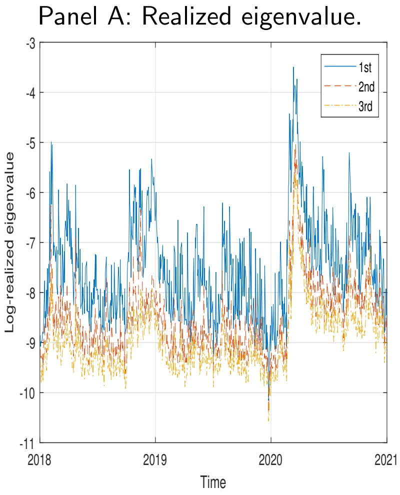
| Panel C: Autoregressive noise | | | |
|---|---|---|---|
| *Consistent estimator* | | | |
| Relative bias | 1.028 | 1.157 | 1.200 |
| Estimation Error | 0.026 | 0.062 | 0.067 |
| Standard deviation | 0.067 | 0.022 | 0.013 |
| | | | |
| *Bias-corrected estimator* | | | |
| Relative bias | 0.995 | 1.048 | 1.117 |
| Estimation Error | -0.001 | 0.019 | 0.039 |
| Standard deviation | 0.069 | 0.024 | 0.016 |

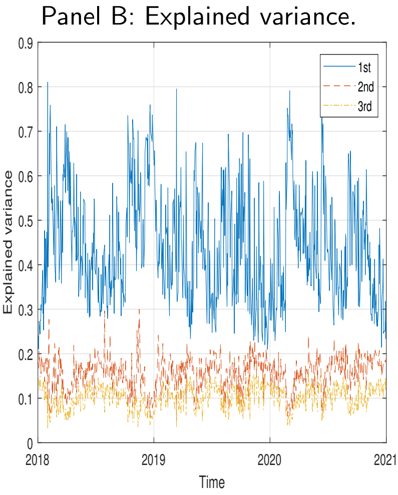
1. Introduction
2. Setting (Target, Estimators, Assumptions)
3. Theoretical Results
4. Simulations
5. Empirical Analysis

# Dataset

- $d = 10$ stocks from NYSE TAQ dataset.
- Eight liquid Dow Jones index members:
  Apple (AAPL), Caterpillar (CAT), Walt Disney (DIS), Microsoft (MSFT), Nike (NKE), Pfizer (PFE), Walmart (WMT), and Exxon Mobil (XOM).
- Two less liquid companies:
  Autodesk (ADSK) and Edison International (EIX).
- January 2, 2018 to December 31, 2020 ($T = 756$ days).
- 1 second sampling during 9:30-16:00 ($n = 23400$ per day).

# Dynamics of daily Realized Eigenvalue

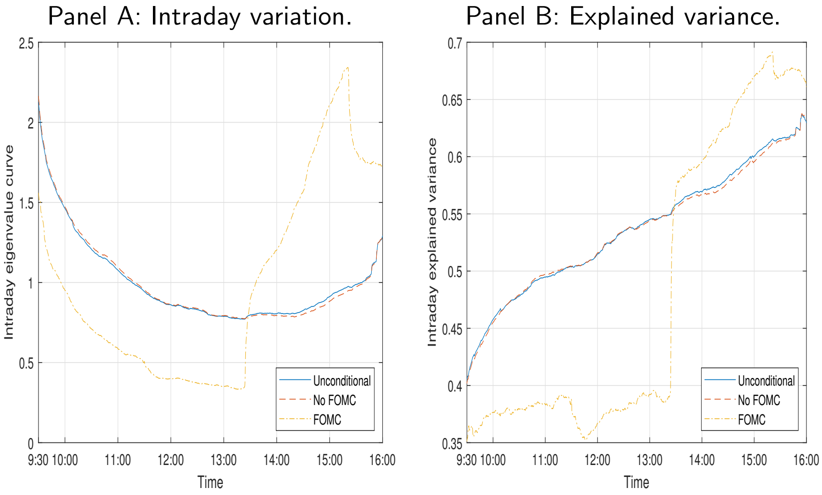


# Intraday Periodicity

- Is there intraday variation in eigenvalues or explained variance? Next, we are interested in the intraday periodicity.
- We look are Federal Open Market Committee (FOMC) meetings, where there eight ordinary meetings per year.
- This is announced during 14.00-14.15.
- On FOMC days, we observe a drastic change in eigenvalues and variation around announcement time.

# Intraday dynamic of principal eigenvalue



**Thanks for your attention!**