Integrated Covolatility Matrix Estimation for High Dimensional Diffusion Processes in the Presence of Microstructure Noise

Xinghua Zheng

Department of ISOM, HKUST

http://ihome.ust.hk/∼xhzheng/

Random matrices and their applications workshop
HKU, Jan 2015

Based on Joint Work with Ningning Xia

Outline

Background: Integrated Covariance Matrix

- $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^T$ denotes a p-dimensional log price process
- Model:

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{W}_t, \quad t \in [0, 1]$$

  where
  1. $\boldsymbol{\mu}_t$ is a p-dimensional drift process;
  2. $\boldsymbol{\Theta}_t$ is a $p \times p$ matrix-valued covolatility process;
  3. $\mathbf{W}_t$ is a p-dimensional standard Brownian motion.
- Both μ_t and Θ_t can be *stochastic*, *discontinuous*, and *dependent on* $\mathbf{W}_t$
- The integrated covariance matrix (ICV):

$$\Sigma^{ICV} := \int_0^1 \Theta_t \Theta_t^T dt.$$

- ICV is the key in risk management and portfolio optimization

Background: Integrated Covariance Matrix

- $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^T$ denotes a p-dimensional log price process
- Model:

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{W}_t, \quad t \in [0, 1]$$

  where
  1. $\boldsymbol{\mu}_t$ is a p-dimensional drift process;
  2. $\boldsymbol{\Theta}_t$ is a $p \times p$ matrix-valued covolatility process;
  3. $\mathbf{W}_t$ is a p-dimensional standard Brownian motion.
- Both $\boldsymbol{\mu}_t$ and $\boldsymbol{\Theta}_t$ can be *stochastic*, *discontinuous*, and *dependent on* $\mathbf{W}_t$
- The integrated covariance matrix (ICV):

$$\Sigma^{ICV} := \int_0^1 \Theta_t \Theta_t^T dt.$$

- ICV is the key in risk management and portfolio optimization

Background: Integrated Covariance Matrix

- $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^T$ denotes a p-dimensional log price process
- Model:

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{W}_t, \quad t \in [0, 1]$$

  where
  1. $\boldsymbol{\mu}_t$ is a p-dimensional drift process;
  2. $\boldsymbol{\Theta}_t$ is a $p \times p$ matrix-valued covolatility process;
  3. $\mathbf{W}_t$ is a p-dimensional standard Brownian motion.
- Both $\boldsymbol{\mu}_t$ and $\boldsymbol{\Theta}_t$ can be *stochastic*, *discontinuous*, and *dependent on* $\mathbf{W}_t$
- The integrated covariance matrix (ICV):

$$\boldsymbol{\Sigma}^{ICV} := \int_0^1 \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^T dt.$$

- ICV is the key in risk management and portfolio optimization

Background: Integrated Covariance Matrix

- $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^T$ denotes a p-dimensional log price process
- Model:

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{W}_t, \quad t \in [0, 1]$$

  where
  1. $\boldsymbol{\mu}_t$ is a p-dimensional drift process;
  2. $\boldsymbol{\Theta}_t$ is a $p \times p$ matrix-valued covolatility process;
  3. $\mathbf{W}_t$ is a p-dimensional standard Brownian motion.
- Both $\boldsymbol{\mu}_t$ and $\boldsymbol{\Theta}_t$ can be *stochastic*, *discontinuous*, and *dependent on* $\mathbf{W}_t$
- The integrated covariance matrix (ICV):

$$\boldsymbol{\Sigma}^{ICV} := \int_0^1 \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^T dt.$$

- ICV is the key in risk management and portfolio optimization

Estimate ICV: Realized Covariance (RCV) Matrix

- Suppose one observes $(\mathbf{X}_t)$ at times $0 = t_0^n < t_1^n < \ldots < t_n^n = 1$
- The realized covariance matrix (RCV):

$$\mathbf{\Sigma}^{RCV} := \sum_{i=1}^{n} \Delta\mathbf{X}_i (\Delta\mathbf{X}_i)^T$$

  where $\Delta\mathbf{X}_i = \mathbf{X}_{t_i^n} - \mathbf{X}_{t_{i-1}^n}$.
- When dimension p is fixed and observation frequency n tends to infinity,

$$\|\mathbf{\Sigma}^{RCV} - \mathbf{\Sigma}^{ICV}\| \xrightarrow{p} 0,$$

  where $\|\cdot\|$ can be any matrix norm.
- In practical applications, p is often comparable with n
- RCV is a poor estimator in such a high-dimensional setting.

Estimate ICV: Realized Covariance (RCV) Matrix

- Suppose one observes $(\mathbf{X}_t)$ at times $0 = t_0^n < t_1^n < \ldots < t_n^n = 1$
- The realized covariance matrix (RCV):

$$\mathbf{\Sigma}^{RCV} := \sum_{i=1}^{n} \Delta\mathbf{X}_i (\Delta\mathbf{X}_i)^T$$

where $\Delta\mathbf{X}_i = \mathbf{X}_{t_i^n} - \mathbf{X}_{t_{i-1}^n}$.
- When dimension p is fixed and observation frequency n tends to infinity,

$$\|\mathbf{\Sigma}^{RCV} - \mathbf{\Sigma}^{ICV}\| \xrightarrow{p} 0,$$

where $\|\cdot\|$ can be any matrix norm.
- In practical applications, p is often comparable with n
- RCV is a poor estimator in such a high-dimensional setting.

Limiting Behaviour of RCV Matrix

- If $\mu_t \equiv 0, \Theta_t \equiv \Theta$, $t_i^n = i/n$, then $\Sigma^{ICV} = \Theta\Theta^T$, and

$$\Delta \mathbf{X}_i = \int_{(i-1)/n}^{i/n} \Theta d\mathbf{W}_t \stackrel{d}{=} \frac{1}{\sqrt{n}} \mathbf{Y}_i,$$

  where $\mathbf{Y}_i \stackrel{i.i.d.}{\sim} N(0, \Sigma^{ICV})$
- Hence

$$\Sigma^{RCV} = \sum_{i=1}^{n} \Delta \mathbf{X}_i (\Delta \mathbf{X}_i)^T \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i \mathbf{Y}_i^T,$$

  a sample covariance matrix with population covariance Σ^{ICV}.

Limiting Behaviour of RCV Matrix

- If $\boldsymbol{\mu}_t \equiv 0, \boldsymbol{\Theta}_t \equiv \boldsymbol{\Theta}$, $t_i^n = i/n$, then $\boldsymbol{\Sigma}^{ICV} = \boldsymbol{\Theta}\boldsymbol{\Theta}^T$, and

$$\Delta \mathbf{X}_i = \int_{(i-1)/n}^{i/n} \boldsymbol{\Theta} d\mathbf{W}_t \stackrel{d}{=} \frac{1}{\sqrt{n}} \mathbf{Y}_i,$$

  where $\mathbf{Y}_i \stackrel{i.i.d.}{\sim} N(0, \boldsymbol{\Sigma}^{ICV})$

- Hence

$$\boldsymbol{\Sigma}^{RCV} = \sum_{i=1}^{n} \Delta \mathbf{X}_i (\Delta \mathbf{X}_i)^T \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i \mathbf{Y}_i^T,$$

  a sample covariance matrix with population covariance $\boldsymbol{\Sigma}^{ICV}$.

Limiting Behaviour of RCV Matrix

- If $\boldsymbol{\mu}_t \equiv 0, \boldsymbol{\Theta}_t \equiv \boldsymbol{\Theta}$, $t_i^n = i/n$, then $\boldsymbol{\Sigma}^{ICV} = \boldsymbol{\Theta}\boldsymbol{\Theta}^T$, and

$$\Delta \mathbf{X}_i = \int_{(i-1)/n}^{i/n} \boldsymbol{\Theta}\, d\mathbf{W}_t \stackrel{d}{=} \frac{1}{\sqrt{n}} \mathbf{Y}_i,$$

  where $\mathbf{Y}_i \stackrel{i.i.d.}{\sim} N(0, \boldsymbol{\Sigma}^{ICV})$

- Hence

$$\boldsymbol{\Sigma}^{RCV} = \sum_{i=1}^{n} \Delta \mathbf{X}_i (\Delta \mathbf{X}_i)^T \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i \mathbf{Y}_i^T,$$

  a sample covariance matrix with population covariance $\boldsymbol{\Sigma}^{ICV}$.

Marčenko-Pastur Theorem

- n i.i.d. p-dim observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$, with mean 0 and covariance matrix $\mathbf{\Sigma}$
- Sample covariance matrix

$$S = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i (\mathbf{Y}_i)^T$$

- If (1) the empirical spectral distribution (ESD) of $\mathbf{\Sigma}$, $F^{\mathbf{\Sigma}}$, converges to H, and (2) $p/n \to y \in (0, \infty)$, then the ESD of $\mathbf{S}$ converges to a nonrandom limit F, whose Stieltjes transform $m_F(\cdot)$ relates to H through

$$m_F(z) = \int_{\tau\in\mathbb{R}} \frac{dH(\tau)}{\tau(1 - y - yzm_F(z))}, \quad \forall z \in \mathbb{C}^+.$$

Marčenko-Pastur Theorem

- n i.i.d. p-dim observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$, with mean 0 and covariance matrix $\mathbf{\Sigma}$
- Sample covariance matrix

$$S = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i (\mathbf{Y}_i)^T$$

- If (1) the empirical spectral distribution (ESD) of $\mathbf{\Sigma}$, $F^{\mathbf{\Sigma}}$, converges to H, and (2) $p/n \to y \in (0, \infty)$, then the ESD of $\mathbf{S}$ converges to a nonrandom limit F, whose Stieltjes transform $m_F(\cdot)$ relates to H through

$$m_F(z) = \int_{\tau \in \mathbb{R}} \frac{dH(\tau)}{\tau(1 - y - yzm_F(z))}, \quad \forall z \in \mathbb{C}^+.$$

Implications

- If indeed $d\mathbf{X}_t = \Theta d\mathbf{W}_t$, then based on the M-P theorem,
  - Knowing limit of ESD (LSD) of $\mathbf{\Sigma}^{ICV}$, one can "predict" the ESD(LSD) of $\mathbf{\Sigma}^{RCV}$
  - Starting from the observable ESD of $\mathbf{\Sigma}^{RCV}$, one can "recover" the ESD of $\mathbf{\Sigma}^{ICV}$ ([Bai, Chen, and Yao(2010)], [El Karoui(2008)], [Mestre(2008)], $\cdots$).
- However, in practice, $\mu_t \neq 0, \Theta_t \neq \Theta$, and

$$\Delta\mathbf{X}_i = \int_{(i-1)/n}^{i/n} (\mu_t dt + \Theta_t d\mathbf{W}_t)$$

  can be far from i.i.d..
- [Zheng and Li(2011)] show that the LSD of $\mathbf{\Sigma}^{RCV}$ depends on time variability of (Θ_t)

Implications

- If indeed $d\mathbf{X}_t = \Theta d\mathbf{W}_t$, then based on the M-P theorem,
  - Knowing limit of ESD (LSD) of $\mathbf{\Sigma}^{ICV}$, one can "predict" the ESD(LSD) of $\mathbf{\Sigma}^{RCV}$
  - Starting from the observable ESD of $\mathbf{\Sigma}^{RCV}$, one can "recover" the ESD of $\mathbf{\Sigma}^{ICV}$ ([Bai, Chen, and Yao(2010)], [El Karoui(2008)], [Mestre(2008)], $\cdots$).
- However, in practice, $\mu_t \neq 0, \Theta_t \neq \Theta$, and

$$\Delta\mathbf{X}_i = \int_{(i-1)/n}^{i/n} (\mu_t dt + \Theta_t d\mathbf{W}_t)$$

  can be far from i.i.d..
- [Zheng and Li(2011)] show that the LSD of $\mathbf{\Sigma}^{RCV}$ depends on time variability of (Θ_t)

Implications

- If indeed $d\mathbf{X}_t = \Theta d\mathbf{W}_t$, then based on the M-P theorem,
  - Knowing limit of ESD (LSD) of $\mathbf{\Sigma}^{ICV}$, one can "predict" the ESD(LSD) of $\mathbf{\Sigma}^{RCV}$
  - Starting from the observable ESD of $\mathbf{\Sigma}^{RCV}$, one can "recover" the ESD of $\mathbf{\Sigma}^{ICV}$ ([Bai, Chen, and Yao(2010)], [El Karoui(2008)], [Mestre(2008)], $\cdots$).
- However, in practice, $\boldsymbol{\mu}_t \neq 0, \mathbf{\Theta}_t \neq \Theta$, and

$$\Delta\mathbf{X}_i = \int_{(i-1)/n}^{i/n} (\boldsymbol{\mu}_t dt + \mathbf{\Theta}_t d\mathbf{W}_t)$$

  can be far from i.i.d..
- [Zheng and Li(2011)] show that the LSD of $\mathbf{\Sigma}^{RCV}$ depends on time variability of $(\mathbf{\Theta}_t)$

Implications

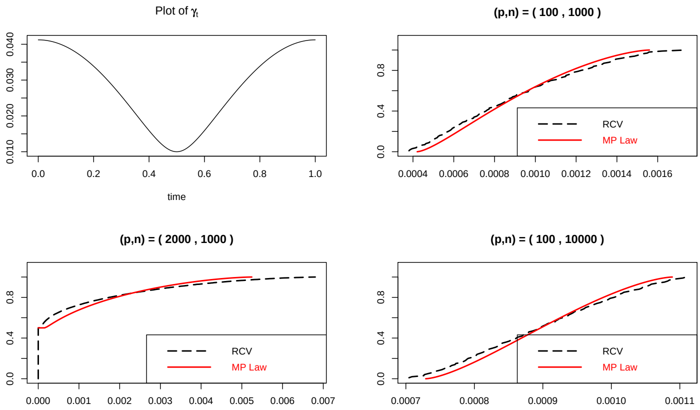
- If indeed $d\mathbf{X}_t = \Theta d\mathbf{W}_t$, then based on the M-P theorem,
  - Knowing limit of ESD (LSD) of $\mathbf{\Sigma}^{ICV}$, one can "predict" the ESD(LSD) of $\mathbf{\Sigma}^{RCV}$
  - Starting from the observable ESD of $\mathbf{\Sigma}^{RCV}$, one can "recover" the ESD of $\mathbf{\Sigma}^{ICV}$ ([Bai, Chen, and Yao(2010)], [El Karoui(2008)], [Mestre(2008)], $\cdots$).
- However, in practice, $\boldsymbol{\mu}_t \neq 0, \mathbf{\Theta}_t \neq \Theta$, and

$$\Delta \mathbf{X}_i = \int_{(i-1)/n}^{i/n} (\boldsymbol{\mu}_t dt + \mathbf{\Theta}_t d\mathbf{W}_t)$$

  can be far from i.i.d..
- [Zheng and Li(2011)] show that the LSD of $\mathbf{\Sigma}^{RCV}$ depends on time variability of $(\mathbf{\Theta}_t)$

Illustration: ESD of Σ^{RCV} depends on time variability of Θ_t

An example of ESD of RCV with a time varying γ_t:



Yet Another Challenge

- In practice, another challenge is that the observations are contaminated:

$$\mathbf{Y}_{t_i} = \mathbf{X}_{t_i} + \varepsilon_i$$

- Fundamental questions: in the high-dimensional setting, with the *noisy* high-frequency observations $(\mathbf{Y}_{t_i})$,
    - How well can we estimate the ICV?
    - In particular, how well can we estimate the eigenvalues of ICV?

Yet Another Challenge

- In practice, another challenge is that the observations are contaminated:

$$\mathbf{Y}_{t_i} = \mathbf{X}_{t_i} + \varepsilon_i$$

- Fundamental questions: in the high-dimensional setting, with the *noisy* high-frequency observations $(\mathbf{Y}_{t_i})$,
  - How well can we estimate the ICV?
  - In particular, how well can we estimate the eigenvalues of ICV?

Yet Another Challenge

- In practice, another challenge is that the observations are contaminated:

$$\mathbf{Y}_{t_i} = \mathbf{X}_{t_i} + \varepsilon_i$$

- Fundamental questions: in the high-dimensional setting, with the *noisy* high-frequency observations $(\mathbf{Y}_{t_i})$,
  - How well can we estimate the ICV?
  - In particular, how well can we estimate the eigenvalues of ICV?

Consequence of Microstructure Noise

- Suppose $\boldsymbol{\mu}_t \equiv 0, \boldsymbol{\Theta}_t \equiv \mathbf{I}$, $t_i^n = \dfrac{i}{n}$ for $i = 0, 1, \cdots, n$, and

$$\mathbf{Y}_{t_i} = \mathbf{X}_{t_i} + \varepsilon_i, \quad \varepsilon_i \overset{i.i.d.}{\sim} (0, \sigma^2 \mathbf{I})$$

- Then $\Delta \mathbf{Y}_i := \Delta \mathbf{X}_i + \Delta \varepsilon_i$, where

$$\begin{aligned} \Delta \mathbf{X}_i &= \mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}} \overset{d}{=} \frac{1}{\sqrt{n}} \mathbf{Z}_i = O_p\left(\frac{1}{\sqrt{n}}\right) \\ \Delta \varepsilon_i &= \varepsilon_i - \varepsilon_{i-1} \overset{d}{=} \sqrt{2}\sigma \mathbf{e}_i \quad = O_p(1) \end{aligned}$$

$$\mathbf{Z}_i \overset{i.i.d.}{\sim} N(0, \mathbf{I}), \qquad \mathbf{e}_i \overset{i.i.d.}{\sim} (0, \mathbf{I}).$$

- Noise dominates signal!

Consequence of Microstructure Noise

- Suppose $\boldsymbol{\mu}_t \equiv 0, \boldsymbol{\Theta}_t \equiv \mathbf{I}$, $t_i^n = \dfrac{i}{n}$ for $i = 0, 1, \cdots, n$, and

$$\mathbf{Y}_{t_i} = \mathbf{X}_{t_i} + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} (0, \sigma^2 \mathbf{I})$$

- Then $\Delta \mathbf{Y}_i := \Delta \mathbf{X}_i + \Delta \boldsymbol{\varepsilon}_i$, where

$$\begin{aligned} \Delta \mathbf{X}_i &= \mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}} \overset{d}{=} \frac{1}{\sqrt{n}} \mathbf{Z}_i = O_p\left(\frac{1}{\sqrt{n}}\right) \\ \Delta \boldsymbol{\varepsilon}_i &= \boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_{i-1} \overset{d}{=} \sqrt{2}\sigma \mathbf{e}_i = O_p(1) \end{aligned}$$

$$\mathbf{Z}_i \overset{i.i.d.}{\sim} N(0, \mathbf{I}), \qquad \mathbf{e}_i \overset{i.i.d.}{\sim} (0, \mathbf{I}).$$

- Noise dominates signal!

Consequence of Microstructure Noise

- Suppose $\boldsymbol{\mu}_t \equiv 0, \boldsymbol{\Theta}_t \equiv \mathbf{I}$, $t_i^n = \frac{i}{n}$ for $i = 0, 1, \cdots, n$, and

$$\mathbf{Y}_{t_i} = \mathbf{X}_{t_i} + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} (0, \sigma^2 \mathbf{I})$$

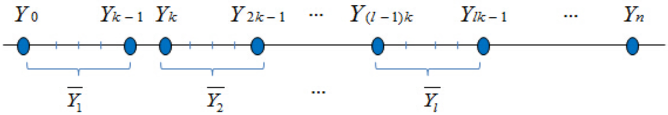
- Then $\Delta \mathbf{Y}_i := \Delta \mathbf{X}_i + \Delta \boldsymbol{\varepsilon}_i$, where

$$\begin{aligned}
\Delta \mathbf{X}_i &= \mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}} \overset{d}{=} \frac{1}{\sqrt{n}} \mathbf{Z}_i = O_p\left(\frac{1}{\sqrt{n}}\right) \\
\Delta \boldsymbol{\varepsilon}_i &= \boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_{i-1} \overset{d}{=} \sqrt{2}\sigma \mathbf{e}_i = O_p(1)
\end{aligned}$$

$$\mathbf{Z}_i \overset{i.i.d.}{\sim} N(0, \mathbf{I}), \qquad \mathbf{e}_i \overset{i.i.d.}{\sim} (0, \mathbf{I}).$$

- Noise dominates signal!

Pre-averaging Approach [Jacod et al.(2009)]

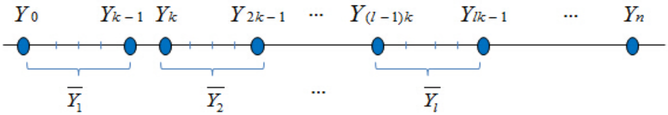


Define moving averages

$$\overline{\mathbf{Y}}_\ell = \frac{1}{k}\sum_{j=(\ell-1)k}^{\ell k-1} \mathbf{Y}_{t_j} = \overline{\mathbf{X}}_\ell + \bar{\varepsilon}_\ell \quad \ell = 1, 2, \cdots, [n/k].$$

- Main intuition: averaging reduces the variance of the noise in $\overline{\mathbf{Y}}_\ell$ by a factor of $1/k$.
- RCV based on $(\overline{\mathbf{Y}}_\ell)$ may be more relevant.

Pre-averaging Approach [Jacod et al.(2009)]

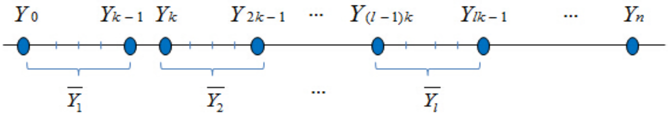


Define moving averages

$$\overline{\mathbf{Y}}_\ell = \frac{1}{k} \sum_{j=(\ell-1)k}^{\ell k-1} \mathbf{Y}_{t_j} = \overline{\mathbf{X}}_\ell + \bar{\varepsilon}_\ell \quad \ell = 1, 2, \cdots, [n/k].$$

- Main intuition: averaging reduces the variance of the noise in $\overline{\mathbf{Y}}_\ell$ by a factor of $1/k$.
- RCV based on $(\overline{\mathbf{Y}}_\ell)$ may be more relevant.

Pre-averaging Approach [Jacod et al.(2009)]



Define moving averages

$$\overline{\mathbf{Y}}_\ell = \frac{1}{k}\sum_{j=(\ell-1)k}^{\ell k-1} \mathbf{Y}_{t_j} = \overline{\mathbf{X}}_\ell + \overline{\varepsilon}_\ell \quad \ell = 1, 2, \cdots, [n/k].$$

- Main intuition: averaging reduces the variance of the noise in $\overline{\mathbf{Y}}_\ell$ by a factor of $1/k$.
- RCV based on $(\overline{\mathbf{Y}}_\ell)$ may be more relevant.

PA-RCV

- Choose a $\theta \in (0, \infty)$ and let moving window length be $k = [\theta\sqrt{n}]$.
- The observations $(\mathbf{Y}_t)$ can be grouped into $m = [n/(2k)]$ pairs of non-overlapping windows.
- Define the PA-RCV matrix as

$$\begin{aligned}\Sigma^{PARCV} &:= \sum_{\ell=1}^{m} (\Delta_{2\ell}\bar{\mathbf{Y}})(\Delta_{2\ell}\bar{\mathbf{Y}})^T \\ &= \sum_{\ell=1}^{m} (\Delta_{2\ell}\bar{\mathbf{X}} + \Delta_{2\ell}\bar{\varepsilon})(\Delta_{2\ell}\bar{\mathbf{X}} + \Delta_{2\ell}\bar{\varepsilon})^T,\end{aligned}$$

where

$$\Delta_{2\ell}\bar{\mathbf{V}} = \bar{\mathbf{V}}_{2\ell} - \bar{\mathbf{V}}_{2\ell-1}, \quad \bar{\mathbf{V}}_\ell = \frac{1}{k}\sum_{j=(\ell-1)k}^{\ell k-1} \mathbf{V}_j,$$

for any process $\mathbf{V} = (\mathbf{V}_t)_{t\geq 0}$.

PA-RCV

- Choose a $\theta \in (0, \infty)$ and let moving window length be $k = [\theta\sqrt{n}]$.
- The observations $(\mathbf{Y}_t)$ can be grouped into $m = [n/(2k)]$ pairs of non-overlapping windows.
- Define the PA-RCV matrix as

$$\begin{aligned}\boldsymbol{\Sigma}^{PARCV} &:= \sum_{\ell=1}^{m} (\Delta_{2\ell}\bar{\mathbf{Y}})(\Delta_{2\ell}\bar{\mathbf{Y}})^T \\ &= \sum_{\ell=1}^{m} (\Delta_{2\ell}\bar{\mathbf{X}} + \Delta_{2\ell}\bar{\varepsilon})(\Delta_{2\ell}\bar{\mathbf{X}} + \Delta_{2\ell}\bar{\varepsilon})^T,\end{aligned}$$

where

$$\Delta_{2\ell}\bar{\mathbf{V}} = \bar{\mathbf{V}}_{2\ell} - \bar{\mathbf{V}}_{2\ell-1}, \quad \bar{\mathbf{V}}_{\ell} = \frac{1}{k}\sum_{j=(\ell-1)k}^{\ell k-1} \mathbf{V}_j,$$

for any process $\mathbf{V} = (\mathbf{V}_t)_{t\geq 0}$.

PA-RCV

- The matrix $\mathbf{\Sigma}^{PARCV}$ can be viewed as the sample covariance matrix based on *noisy* observations $\Delta_{2i}\bar{\mathbf{X}} + \Delta_{2i}\bar{\varepsilon}$;
- [Dozier and Silverstein(2007)] consider such information-plus-noise-type sample covariance matrices as

$$\mathbf{S}_n = \frac{1}{n}(\mathbf{A}_n + \sigma\varepsilon_n)(\mathbf{A}_n + \sigma\varepsilon_n)^T,$$

where ε_n is independent of $\mathbf{A}_n$ and consists of i.i.d. entries with zero mean and unit variance.

PA-RCV

- The matrix $\mathbf{\Sigma}^{PARCV}$ can be viewed as the sample covariance matrix based on *noisy* observations $\Delta_{2i}\bar{\mathbf{X}} + \Delta_{2i}\bar{\varepsilon}$;
- [Dozier and Silverstein(2007)] consider such information-plus-noise-type sample covariance matrices as

$$\mathbf{S}_n = \frac{1}{n}(\mathbf{A}_n + \sigma\boldsymbol{\varepsilon}_n)(\mathbf{A}_n + \sigma\boldsymbol{\varepsilon}_n)^T,$$

where ε_n is independent of $\mathbf{A}_n$ and consists of i.i.d. entries with zero mean and unit variance.

Sample covariance matrices based on

- [Dozier and Silverstein(2007)] show that if (i) $F^{\mathcal{A}_n} \to H$ where $\mathcal{A}_n = \mathbf{A}_n \mathbf{A}_n / n$, and (ii) $p/n \to y > 0$, then the ESD of $\mathbf{S}_n$ converges to a nonrandom p.d.f. F whose Stieltjes transform $m = m(z)$ satisfies

$$m = \int \frac{dH(t)}{\frac{t}{1+\sigma^2 ym} - (1+\sigma^2 ym)z + \sigma^2(1-y)}, \quad \forall z \in \mathbb{C}^+.$$

- This relationship shows how the LSD of $\mathbf{S}_n$ depends on that of $\mathcal{A}_n$.
- In practice, we are often more interested in making inference about signals $\mathbf{A}_n$ based on noisy observation $\mathbf{A}_n + \sigma\varepsilon_n$.
- Our first result establishes a relationship that describes how the LSD of $\mathcal{A}_n$ depends on that of $\mathbf{S}_n$.

Sample covariance matrices based on

- [Dozier and Silverstein(2007)] show that if (i) $F^{\mathcal{A}_n} \to H$ where $\mathcal{A}_n = \mathbf{A}_n\mathbf{A}_n/n$, and (ii) $p/n \to y > 0$, then the ESD of $\mathbf{S}_n$ converges to a nonrandom p.d.f. F whose Stieltjes transform $m = m(z)$ satisfies

$$m = \int \frac{dH(t)}{\frac{t}{1+\sigma^2 ym} - (1+\sigma^2 ym)z + \sigma^2(1-y)}, \quad \forall z \in \mathbb{C}^+.$$

- This relationship shows how the LSD of $\mathbf{S}_n$ depends on that of $\mathcal{A}_n$.
- In practice, we are often more interested in making inference about signals $\mathbf{A}_n$ based on noisy observation $\mathbf{A}_n + \sigma\varepsilon_n$.
- Our first result establishes a relationship that describes how the LSD of $\mathcal{A}_n$ depends on that of $\mathbf{S}_n$.

Inversion Theorem

Theorem (1)

[Xia and Zheng(2014)] Under the assumptions above and if F admits a bounded density over a finite interval and possibly a point mass at 0, then $m_{\mathcal{A}}(z)$ is determined by F in that it uniquely solves the following equation

$$m_{\mathcal{A}}(z) = \int \frac{dF(\tau)}{\frac{\tau}{1 - y\sigma^2 m_{\mathcal{A}}(z)} - z(1 - y\sigma^2 m_{\mathcal{A}}(z)) + \sigma^2(y-1)}. \quad (2.1)$$

- F is observable, solving for $m_{\mathcal{A}}(z)$ allowing us to make inferences about the spectrum of the covariance structure of the underlying signals.

Inversion Theorem

Theorem (1)

[Xia and Zheng(2014)] Under the assumptions above and if F admits a bounded density over a finite interval and possibly a point mass at 0, then $m_{\mathcal{A}}(z)$ is determined by F in that it uniquely solves the following equation

$$m_{\mathcal{A}}(z) = \int \frac{dF(\tau)}{\frac{\tau}{1 - y\sigma^2 m_{\mathcal{A}}(z)} - z(1 - y\sigma^2 m_{\mathcal{A}}(z)) + \sigma^2(y-1)}. \quad (2.1)$$

- F is observable, solving for $m_{\mathcal{A}}(z)$ allowing us to make inferences about the spectrum of the covariance structure of the underlying signals.

The Class $\mathcal{C}$

- Say that $(\mathbf{X}_t)$ belongs to Class $\mathcal{C}$ if its covolatility process $(\boldsymbol{\Theta}_t)$ has the form

$$\boldsymbol{\Theta}_t = \gamma_t \boldsymbol{\Lambda}$$

where $(\gamma_t) \in D([0,1];\mathbb{R})$ and $\boldsymbol{\Lambda}$ is a $p \times p$ matrix.

LSD of PA-RCV for Class $\mathcal{C}$

Theorem (2)

Suppose that

- $(\mathbf{X}_t)$ *belongs to Class* $\mathcal{C}$ *with a covolatility process* $\mathbf{\Theta}_t = \gamma_t \mathbf{\Lambda}$
- *Observe* $\mathbf{Y}_{i/n} = \mathbf{X}_{i/n} + \varepsilon_i$ *where* (ε_i) *are i.i.d. with* $\mathrm{E}(\varepsilon_i) = 0$ *and* $\mathrm{cov}(\varepsilon_i) = \sigma_e^2 \mathbf{I}_p$.
- $\breve{\mathbf{\Sigma}}_p = \mathbf{\Lambda}\mathbf{\Lambda}^T$ *with an LSD* $\breve{H}$.

Assume $k = [\theta\sqrt{n}]$ *for some* $\theta \in (0, \infty)$ *and* $m = [n/(2k)]$ *satisfies that* $\lim_{p\to\infty} p/m = y > 0$. *Then as* $p \to \infty$,

- *ESDs of* $\mathbf{\Sigma}^{ICV}$ *and* $\mathbf{\Sigma}^{PARCV}$ *converge to* H *and* F, *respectively, where*

$$H(x) = \breve{H}(x/\zeta), \quad \text{for all } x \geq 0 \quad \textit{and} \quad \zeta = \lim \int_0^1 (\gamma_t)^2 \, dt.$$

- *Moreover, if* F *admits a bounded density over a finite interval and possibly a point mass at 0, then we have the following relationships*

$$m_A(z) = -\frac{1}{z} \int \frac{\zeta}{\tau M(z) + \zeta} \, dH(\tau), \qquad (2.2)$$

LSD of PA-RCV for Class $\mathcal{C}$

Theorem (2)

Suppose that

- $(\mathbf{X}_t)$ *belongs to Class* $\mathcal{C}$ *with a covolatility process* $\mathbf{\Theta}_t = \gamma_t \mathbf{\Lambda}$
- *Observe* $\mathbf{Y}_{i/n} = \mathbf{X}_{i/n} + \varepsilon_i$ *where* (ε_i) *are i.i.d. with* $\mathrm{E}(\varepsilon_i) = 0$ *and* $\mathrm{cov}(\varepsilon_i) = \sigma_e^2 \mathbf{I}_p$.
- $\breve{\mathbf{\Sigma}}_p = \mathbf{\Lambda}\mathbf{\Lambda}^T$ *with an LSD* $\breve{H}$.

Assume $k = [\theta\sqrt{n}]$ *for some* $\theta \in (0, \infty)$ *and* $m = [n/(2k)]$ *satisfies that* $\lim_{p\to\infty} p/m = y > 0$. *Then as* $p \to \infty$,

- *ESDs of* $\mathbf{\Sigma}^{ICV}$ *and* $\mathbf{\Sigma}^{PARCV}$ *converge to* H *and* F, *respectively, where*

$$H(x) = \breve{H}(x/\zeta), \quad \text{for all } x \geq 0 \quad \textit{and} \quad \zeta = \lim \int_0^1 (\gamma_t)^2 \, dt.$$

- *Moreover, if* F *admits a bounded density over a finite interval and possibly a point mass at 0, then we have the following relationships*

$$m_{\mathcal{A}}(z) = -\frac{1}{z} \int \frac{\zeta}{\tau M(z) + \zeta} \, dH(\tau), \tag{2.2}$$

LSD of PA-RCV for Class $\mathcal{C}$, *ctd*

- *where* $m_{\mathcal{A}}(z)$ *denotes the Stieltjes transform of the LSD of*

$$\sum_{i=1}^{m} \Delta_{2i}\bar{\mathbf{X}}(\Delta_{2i}\bar{\mathbf{X}})^T,$$

and is the unique solution to equation

$$m_{\mathcal{A}}(z) = \int \frac{dF(\tau)}{\frac{\tau}{1 - y\theta^{-2}\sigma_e^2 m_{\mathcal{A}}(z)} - z(1 - y\theta^{-2}\sigma_e^2 m_{\mathcal{A}}(z)) + \theta^{-2}\sigma_e^2(y-1)} \tag{2.3}$$

- *and* $M(z)$, *together with another function* $\widetilde{m}(z)$, *uniquely solve the following equations in* $\mathbb{C}^+ \times \mathbb{C}^+$

$$\begin{cases} M(z) = -\frac{1}{z}\int_0^1 \frac{(1/3)(\gamma_s^*)^2}{1 + y\widetilde{m}(z)(1/3)(\gamma_s^*)^2}\,ds, \\ \widetilde{m}(z) = -\frac{1}{z}\int \frac{\tau}{\tau M(z) + \zeta}\,dH(\tau). \end{cases} \tag{2.4}$$

Implications of the Convergence

- Σ^{PARCV} and hence F is observable
- Use (2.3) to estimate $m_{\mathcal{A}}$
- Further use (2.2) and (2.4) to estimate H, the LSD of Σ^{ICV}, the object of interest
- One challenge: the process (γ_t) in (2.4) is not observable
- $\rightarrow$ alternative estimator that circumvents this challenge

Implications of the Convergence

- $\boldsymbol{\Sigma}^{PARCV}$ and hence F is observable
- Use (2.3) to estimate $m_{\mathcal{A}}$
- Further use (2.2) and (2.4) to estimate H, the LSD of Σ^{ICV}, the object of interest
- One challenge: the process (γ_t) in (2.4) is not observable
- $\rightarrow$ alternative estimator that circumvents this challenge

Implications of the Convergence

- $\mathbf{\Sigma}^{PARCV}$ and hence F is observable
- Use (2.3) to estimate $m_{\mathcal{A}}$
- Further use (2.2) and (2.4) to estimate H, the LSD of $\mathbf{\Sigma}^{ICV}$, the object of interest
- One challenge: the process (γ_t) in (2.4) is not observable
- $\rightarrow$ alternative estimator that circumvents this challenge

Implications of the Convergence

- $\mathbf{\Sigma}^{PARCV}$ and hence F is observable
- Use (2.3) to estimate $m_{\mathcal{A}}$
- Further use (2.2) and (2.4) to estimate H, the LSD of $\mathbf{\Sigma}^{ICV}$, the object of interest
- One challenge: the process (γ_t) in (2.4) is not observable
- $\rightarrow$ alternative estimator that circumvents this challenge

Implications of the Convergence

- $\boldsymbol{\Sigma}^{PARCV}$ and hence F is observable
- Use (2.3) to estimate $m_{\mathcal{A}}$
- Further use (2.2) and (2.4) to estimate H, the LSD of $\boldsymbol{\Sigma}^{ICV}$, the object of interest
- One challenge: the process (γ_t) in (2.4) is not observable
- $\rightarrow$ alternative estimator that circumvents this challenge

Alternative Estimator: Pre-averaging Time-Variation Adjusted RCV (PA-TVARCV)

- Fix an $\alpha \in (1/2, 1)$ and $\theta \in (0, \infty)$, let $k = [\theta n^{\alpha}]$, $m = [n/(2k)]$.
- Define PA-TVARCV as

$$\boldsymbol{\Sigma}^{PATVARCV} := \frac{\operatorname{tr}(\mathbf{S}_p)}{m} \sum_{i=1}^{m} \frac{\Delta_{2i}\bar{\mathbf{Y}}(\Delta_{2i}\bar{\mathbf{Y}})^T}{|\Delta_{2i}\bar{\mathbf{Y}}|^2},$$

where $\mathbf{S}_p$ is a standard pre-averaging estimator in [Jacod et al.(2009)]

$$\mathbf{S}_p := \frac{12}{\nu\sqrt{n}} \sum_{i=0}^{n-\ell_n+1} \Delta\bar{\mathbf{Y}}_i(\Delta\bar{\mathbf{Y}}_i)^T - \frac{6}{\nu^2 n} \sum_{i=1}^{n} \Delta_i\mathbf{Y}(\Delta_i\mathbf{Y})^T,$$

where $\ell_n = [\nu\sqrt{n}]$ for some $\nu \in (0, \infty)$,

$$\Delta\bar{\mathbf{Y}}_i = \frac{1}{\ell_n}\left(\sum_{j=[\ell_n/2]}^{\ell_n-1} \mathbf{Y}_{(i+j)/n} - \sum_{j=0}^{[\ell_n/2]-1} \mathbf{Y}_{(i+j)/n}\right), \ \Delta_i\mathbf{Y} = \mathbf{Y}_{i/n} - \mathbf{Y}_{(i-1)/n}.$$

LSD of PA-TVARCV for Class $\mathcal{C}$

Theorem (3)

Suppose that

- $(\mathbf{X}_t)$ *belongs to Class* $\mathcal{C}$ *with a covolatility process* $\mathbf{\Theta}_t = \gamma_t \mathbf{\Lambda}$;
- *Observe* $\mathbf{Y}_{i/n} = \mathbf{X}_{i/n} + \boldsymbol{\varepsilon}_i$ *where* $(\boldsymbol{\varepsilon}_i)$ *are i.i.d. with* $\mathrm{E}(\boldsymbol{\varepsilon}_i) = 0$ *and* $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \mathrm{diag}(d_1^2, \cdots, d_p^2)$;
- p *and* m *satisfy* $p/m \to y \in (0, \infty)$ *as* $p \to \infty$.

Then the LSD of $\mathbf{\Sigma}^{PATVARCV}$ *is uniquely determined by that of ICV through Stieltjes transforms via the standard M-P equation*

$$m_F(z) = \int_{\tau\in\mathbb{R}} \frac{dH(\tau)}{\tau(1 - y(1 + zm_F(z)) - z)}, \quad \forall z \in \mathbb{C}^+.$$

- No (γ_t) involved
- More importantly, noise is also eliminated!

LSD of PA-TVARCV for Class $\mathcal{C}$

Theorem (3)

Suppose that

- $(\mathbf{X}_t)$ *belongs to Class* $\mathcal{C}$ *with a covolatility process* $\boldsymbol{\Theta}_t = \gamma_t \boldsymbol{\Lambda}$;
- *Observe* $\mathbf{Y}_{i/n} = \mathbf{X}_{i/n} + \boldsymbol{\varepsilon}_i$ *where* $(\boldsymbol{\varepsilon}_i)$ *are i.i.d. with* $\mathrm{E}(\boldsymbol{\varepsilon}_i) = 0$ *and* $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \mathrm{diag}(d_1^2, \cdots, d_p^2)$;
- p *and* m *satisfy* $p/m \to y \in (0, \infty)$ *as* $p \to \infty$.

Then the LSD of $\boldsymbol{\Sigma}^{PATVARCV}$ *is uniquely determined by that of ICV through Stieltjes transforms via the standard M-P equation*

$$m_F(z) = \int_{\tau \in \mathbb{R}} \frac{dH(\tau)}{\tau(1 - y(1 + z m_F(z)) - z)}, \quad \forall z \in \mathbb{C}^+.$$

- No (γ_t) involved
- More importantly, noise is also eliminated!

LSD of PA-TVARCV for Class $\mathcal{C}$

Theorem (3)

Suppose that

- $(\mathbf{X}_t)$ *belongs to Class* $\mathcal{C}$ *with a covolatility process* $\mathbf{\Theta}_t = \gamma_t \mathbf{\Lambda}$*;*
- *Observe* $\mathbf{Y}_{i/n} = \mathbf{X}_{i/n} + \boldsymbol{\varepsilon}_i$ *where* $(\boldsymbol{\varepsilon}_i)$ *are i.i.d. with* $\mathrm{E}(\boldsymbol{\varepsilon}_i) = 0$ *and* $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \mathrm{diag}(d_1^2, \cdots, d_p^2)$*;*
- p *and* m *satisfy* $p/m \to y \in (0, \infty)$ *as* $p \to \infty$*.*

Then the LSD of $\mathbf{\Sigma}^{PATVARCV}$ *is uniquely determined by that of ICV through Stieltjes transforms via the standard M-P equation*

$$m_F(z) = \int_{\tau \in \mathbb{R}} \frac{dH(\tau)}{\tau(1 - y(1 + z m_F(z)) - z)}, \quad \forall z \in \mathbb{C}^+.$$

- No (γ_t) involved
- More importantly, noise is also eliminated!

Simulation Studies

- To compare the ESDs of PA-RCV, PA-TVARCV matrices and reference matrix

$$\mathbf{S}_p := \frac{1}{m}(\boldsymbol{\Sigma}^{ICV})^{1/2}\mathbf{Z}_m\mathbf{Z}_m^T(\boldsymbol{\Sigma}^{ICV})^{1/2},$$

- According to Theorem 3, the ESDs of PA-TVARCV and the reference matrix should be similar
- In contrast, according to Theorem 2, that of PA-RCV should be distinguishably different from theirs
- For different p's, with
$n = 23400, \quad k = 250 \left(\approx 1.63\sqrt{n} \approx n^{0.55}\right), \quad m = [n/(2k)],$

Simulation Studies

- To compare the ESDs of PA-RCV, PA-TVARCV matrices and reference matrix

$$\mathbf{S}_p := \frac{1}{m}(\mathbf{\Sigma}^{ICV})^{1/2}\mathbf{Z}_m\mathbf{Z}_m^T(\mathbf{\Sigma}^{ICV})^{1/2},$$

- According to Theorem 3, the ESDs of PA-TVARCV and the reference matrix should be similar
- In contrast, according to Theorem 2, that of PA-RCV should be distinguishably different from theirs
- For different p's, with
$n = 23400, \quad k = 250 \left(\approx 1.63\sqrt{n} \approx n^{0.55}\right), \quad m = [n/(2k)],$

Simulation Studies

- To compare the ESDs of PA-RCV, PA-TVARCV matrices and reference matrix

$$\mathbf{S}_p := \frac{1}{m}(\boldsymbol{\Sigma}^{ICV})^{1/2}\mathbf{Z}_m\mathbf{Z}_m^T(\boldsymbol{\Sigma}^{ICV})^{1/2},$$

- According to Theorem 3, the ESDs of PA-TVARCV and the reference matrix should be similar
- In contrast, according to Theorem 2, that of PA-RCV should be distinguishably different from theirs
- For different p's, with $n = 23400, \quad k = 250 \left(\approx 1.63\sqrt{n} \approx n^{0.55}\right), \quad m = [n/(2k)],$

Simulation Studies

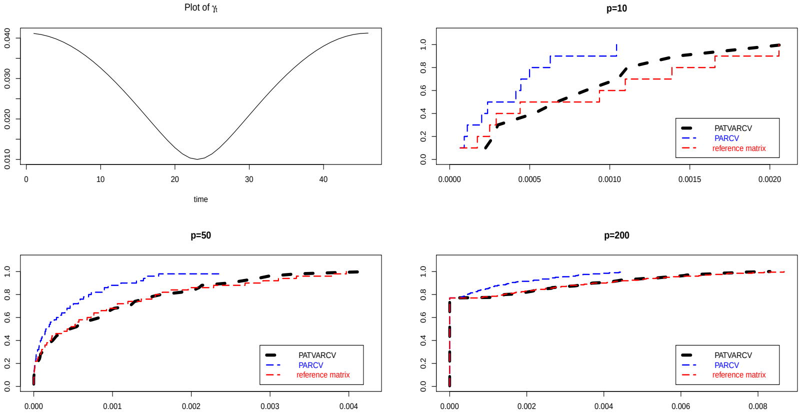
- To compare the ESDs of PA-RCV, PA-TVARCV matrices and reference matrix

$$\mathbf{S}_p := \frac{1}{m}(\boldsymbol{\Sigma}^{ICV})^{1/2}\mathbf{Z}_m\mathbf{Z}_m^T(\boldsymbol{\Sigma}^{ICV})^{1/2},$$

- According to Theorem 3, the ESDs of PA-TVARCV and the reference matrix should be similar
- In contrast, according to Theorem 2, that of PA-RCV should be distinguishably different from theirs
- For different p's, with $n = 23400, \quad k = 250\left(\approx 1.63\sqrt{n} \approx n^{0.55}\right), \quad m = [n/(2k)],$

Simulation Studies, *ctd*

ESDs of PA-RCV and PA-TVARCV, with a continuous (γ_t)



Summary

1. Under high-dimensional noisy setting, we propose PA-RCV estimator and PA-TVARCV estimator, both of which can be used to recover the ESD of ICV matrix.
2. In order to use PA-RCV, one needs to estimate the stochastic volatility process (γ_t).
3. PA-TVARCV has the advantage of eliminating the impacts of both stochastic volatility and the noise!

Summary

1. Under high-dimensional noisy setting, we propose PA-RCV estimator and PA-TVARCV estimator, both of which can be used to recover the ESD of ICV matrix.
2. In order to use PA-RCV, one needs to estimate the stochastic volatility process (γ_t).
3. PA-TVARCV has the advantage of eliminating the impacts of both stochastic volatility and the noise!

Bai, Z., Chen, J., and Yao, J. (2010), "On estimation of the population spectral distribution from high-dimensional sample covariance matrix," *Australian & New Zealand Journal of Statistics*, 52, 423–437.

Dozier, R. B. and Silverstein, J. W. (2007), "Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices," *J. Multivariate Anal.* , 98, 678–694.

El Karoui, N. (2008), "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *Ann. Statist.*, 36, 2757–2790.

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

- Marčenko, V. A. and Pastur, L. A. (1967), "Distribution of eigenvalues in certain sets of random matrices," *Mat. Sb. (N.S.)*, 72 (114), 507–536.
- Mestre, X. (2008), "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Trans. Inform. Theory*, 54, 5113–5129.
- Silverstein, J. W. (1995), "Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices," *J. Multivariate Anal.*, 55, 331–339.
- Xia, N. and Zheng, X. (2014), "Integrated covariance matrix estimation for high-dimensional diffusion processes in the presence of microstructure noise," *working paper*.
- Zheng, X. and Li, Y. (2011), "On the Estimation of Integrated Covariance Matrices of High Dimensional Diffusion Processes," *the Annals of Statistics*, 39, 3121–3151.