ℓ^∞ Eigenvector Perturbation Bound and Robust Covariance Estimation

Yiqiao Zhong

Princeton University
with **Jianqing Fan and Weichen Wang**

May 8, 2016

Outline

1. Matrix Eigenvector Perturbation
   - Davis-Kahan Theorem
   - Our result: ℓ^∞ Eigenvector Perturbation
2. Robust Covariance Estimation via Factor Models
   - Known factors
   - Factors unknown
3. Simulations
4. Real Data Analysis: Portfolio Risk Estimation

Outline

1. Matrix Eigenvector Perturbation
   - Davis-Kahan Theorem
   - Our result: ℓ^∞ Eigenvector Perturbation
2. Robust Covariance Estimation via Factor Models
   - Known factors
   - Factors unknown
3. Simulations
4. Real Data Analysis: Portfolio Risk Estimation

Matrix Eigenvector Perturbation

- Matrix eigenvectors (or singular vectors) are often used to uncover underlying structure.
- Let A be a $d \times d$ symmetric matrix.

$$A = V \Sigma V^T = \sum_{i=1}^{r} \lambda_i v_i v_i^T,$$

where $V = [v_1, \ldots, v_r]$ and $r \leq d$.

- Examples:
  - PCA and factor models;
  - Network analysis;
  - Classical MDS.

Matrix Eigenvector Perturbation

- Matrix eigenvectors (or singular vectors) are often used to uncover underlying structure.
- Let A be a $d \times d$ symmetric matrix.

$$A = V \Sigma V^T = \sum_{i=1}^{r} \lambda_i v_i v_i^T,$$

where $V = [v_1, \ldots, v_r]$ and $r \leq d$.

- Examples:
  - PCA and factor models;
  - Network analysis;
  - Classical MDS.

Matrix Eigenvector Perturbation

- Matrix eigenvectors (or singular vectors) are often used to uncover underlying structure.
- Let A be a $d \times d$ symmetric matrix.

$$A = V\Sigma V^T = \sum_{i=1}^{r} \lambda_i v_i v_i^T,$$

where $V = [v_1, \ldots, v_r]$ and $r \leq d$.

- Examples:
  - PCA and factor models;
  - Network analysis;
  - Classical MDS.

Matrix Eigenvector Perturbation

- However...our data are usually not perfect.
- Let $\widetilde{A}$ be our data matrix,

$$\widetilde{A} = A + E,$$

where symmetric $E \in \mathbf{R}^{d \times d}$ is perturbation.

- Similar decomposition for $\widetilde{A}$.

$$\widetilde{A} = \sum_{i=1}^{d} \widetilde{\lambda}_i \widetilde{v}_i \widetilde{v}_i^T,$$

where $V = [\widetilde{v}_1, \ldots, \widetilde{v}_r]$.

- Want to bound $\|v_i - \widetilde{v}_i\|$ for some norm $\|\cdot\|$.

Matrix Eigenvector Perturbation

- However...our data are usually not perfect.
- Let $\widetilde{A}$ be our data matrix,

$$\widetilde{A} = A + E,$$

where symmetric $E \in \mathbf{R}^{d \times d}$ is perturbation.

- Similar decomposition for $\widetilde{A}$.

$$\widetilde{A} = \sum_{i=1}^{d} \widetilde{\lambda}_i \widetilde{v}_i \widetilde{v}_i^T,$$

where $V = [\widetilde{v}_1, \ldots, \widetilde{v}_r]$.

- Want to bound $\|v_i - \widetilde{v}_i\|$ for some norm $\| \cdot \|$.

Matrix Eigenvector Perturbation

- However...our data are usually not perfect.
- Let $\widetilde{A}$ be our data matrix,

$$\widetilde{A} = A + E,$$

where symmetric $E \in \mathbf{R}^{d \times d}$ is perturbation.

- Similar decomposition for $\widetilde{A}$.

$$\widetilde{A} = \sum_{i=1}^{d} \widetilde{\lambda}_i \widetilde{v}_i \widetilde{v}_i^T,$$

where $V = [\widetilde{v}_1, \ldots, \widetilde{v}_r]$.

- Want to bound $\|v_i - \widetilde{v}_i\|$ for some norm $\|\cdot\|$.

Matrix Eigenvector Perturbation

- Eigengap γ: the smallest gap of $\{\lambda_1, \ldots, \lambda_r, 0\}$.
- Weyl's inequality matches eigenvalues: $\max_i |\lambda_i - \widetilde{\lambda}_i| \leq \|E\|_2$.

Theorem (a simpler version of Davis and Kahan [1970])

Suppose $\gamma > 0$, *and we have*

$$\|v_i - \eta_i \widetilde{v}_i\|_2 \leq \frac{2\sqrt{2}\,\|E\|_2}{\gamma}, \quad i = 1, \ldots, r$$

where $\eta_i \in \{\pm 1\}$ *are suitable signs.*

- Extends to general case (SVD) easily. (Wedin 72').
- Nontrivial only when $\gamma \gg \|E\|_2$

Matrix Eigenvector Perturbation

- Eigengap γ: the smallest gap of $\{\lambda_1, \ldots, \lambda_r, 0\}$.
- Weyl's inequality matches eigenvalues: $\max_i |\lambda_i - \widetilde{\lambda}_i| \le \|E\|_2$.

Theorem (a simpler version of Davis and Kahan [1970])

Suppose $\gamma > 0$, *and we have*

$$\|v_i - \eta_i \widetilde{v}_i\|_2 \le \frac{2\sqrt{2}\,\|E\|_2}{\gamma}, \quad i = 1, \ldots, r$$

where $\eta_i \in \{\pm 1\}$ *are suitable signs.*

- Extends to general case (SVD) easily. (Wedin 72').
- Nontrivial only when $\gamma \gg \|E\|_2$

Question:

- Can we go beyond ℓ^2 norm perturbation?
- Trivial bound $\|\cdot\|_\infty \leq \|\cdot\|_2$ too loose. A sharper bound?

Question:

- Can we go beyond ℓ^2 norm perturbation?
- Trivial bound $\|\cdot\|_\infty \leq \|\cdot\|_2$ too loose. A sharper bound?

Question:

- Can we go beyond ℓ^2 norm perturbation?
- Trivial bound $\|\cdot\|_\infty \leq \|\cdot\|_2$ too loose. A sharper bound?

Outline

1. Matrix Eigenvector Perturbation
   - Davis-Kahan Theorem
   - Our result: ℓ^∞ Eigenvector Perturbation
2. Robust Covariance Estimation via Factor Models
   - Known factors
   - Factors unknown
3. Simulations
4. Real Data Analysis: Portfolio Risk Estimation

ℓ^∞ Eigenvector Perturbation

- A first sight: mission impossible!
- Consider a $d \times d$ low rank matrix $A = d(1,0,\dots,0)^T(1,0,\dots,0)$. Fix $r = 1$. Its top eigenvector is just $v_1 = (1,0,\dots,0)^T$, and $\gamma = d$. Perturbation matrix E: $E_{12} = E_{21} = d/4$ and 0 elsewhere.
- Simple calculation shows $\|v_i - \widetilde{v}_i\|_\infty \asymp 1$. Already sharp!
- Can we do better than a trivial bound?
- Answer: Yes! But for structured low rank matrix A.

ℓ^∞ Eigenvector Perturbation

- A first sight: mission impossible!
- Consider a $d \times d$ low rank matrix $A = d(1,0,\dots,0)^T(1,0,\dots,0)$. Fix $r = 1$. Its top eigenvector is just $v_1 = (1,0,\dots,0)^T$, and $\gamma = d$. Perturbation matrix E: $E_{12} = E_{21} = d/4$ and 0 elsewhere.
- Simple calculation shows $\|v_i - \widetilde{v}_i\|_\infty \asymp 1$. Already sharp!
- Can we do better than a trivial bound?
- Answer: Yes! But for structured low rank matrix A.

ℓ^∞ Eigenvector Perturbation

- A first sight: mission impossible!
- Consider a $d \times d$ low rank matrix $A = d(1,0,\ldots,0)^T(1,0,\ldots,0)$. Fix $r = 1$. Its top eigenvector is just $v_1 = (1,0,\ldots,0)^T$, and $\gamma = d$. Perturbation matrix E: $E_{12} = E_{21} = d/4$ and 0 elsewhere.
- Simple calculation shows $\|v_i - \widetilde{v}_i\|_\infty \asymp 1$. Already sharp!
- Can we do better than a trivial bound?
- Answer: Yes! But for structured low rank matrix A.

ℓ^∞ Eigenvector Perturbation

- A first sight: mission impossible!
- Consider a $d \times d$ low rank matrix $A = d(1, 0, \ldots, 0)^T(1, 0, \ldots, 0)$. Fix $r = 1$. Its top eigenvector is just $v_1 = (1, 0, \ldots, 0)^T$, and $\gamma = d$. Perturbation matrix E: $E_{12} = E_{21} = d/4$ and 0 elsewhere.
- Simple calculation shows $\|v_i - \widetilde{v}_i\|_\infty \asymp 1$. Already sharp!
- Can we do better than a trivial bound?
- Answer: Yes! But for structured low rank matrix A.

Our result: ℓ^∞ Eigenvector Perturbation

- What structure? *low rank* and *incoherent* matrices.

- Low rank: $r = \text{rank}(A)$ is small, e.g. bounded by a constant.

Definition (Candès and Recht [2009])

Let $V = [v_1, \ldots, v_r]$ be r columns of orthonormal vectors in $\mathbf{R}^d$. The coherence of V is:

$$\mu(V) = \frac{d}{r} \max_i \sum_{j=1}^{r} V_{ij}^2.$$

- Incoherence: $\mu(V)$ is small, e.g. logarithmic in d.

- When v_i is a uinform vector, $\mu(V) = O(\log d)$.

Our result: ℓ^∞ Eigenvector Perturbation

- What structure? *low rank* and *incoherent* matrices.
- Low rank: $r = \text{rank}(A)$ is small, e.g. bounded by a constant.

Definition (Candès and Recht [2009])

Let $V = [v_1, \ldots, v_r]$ be r columns of orthonormal vectors in $\mathbf{R}^d$. The coherence of V is:

$$\mu(V) = \frac{d}{r} \max_i \sum_{j=1}^{r} V_{ij}^2 .$$

- Incoherence: $\mu(V)$ is small, e.g. logarithmic in d.
- When v_i is a uinform vector, $\mu(V) = O(\log d)$.

Our result: ℓ^∞ Eigenvector Perturbation

- What structure? *low rank* and *incoherent* matrices.
- Low rank: $r = \mathrm{rank}(A)$ is small, e.g. bounded by a constant.

Definition (Candès and Recht [2009])

Let $V = [v_1, \ldots, v_r]$ be r columns of orthonormal vectors in $\mathbf{R}^d$. The coherence of V is:

$$\mu(V) = \frac{d}{r} \max_i \sum_{j=1}^{r} V_{ij}^2 .$$

- Incoherence: $\mu(V)$ is small, e.g. logarithmic in d.
- When v_i is a uinform vector, $\mu(V) = O(\log d)$.

Our result: ℓ^∞ Eigenvector Perturbation

- One more thing: need different matrix norms for perturbation E.

- For symmetric $E \in \mathbf{R}^{d \times d}$, define

$$\|E\|_\infty = \max_i \sum_j^d |E_{ij}|.$$

- A trivial bound $\|E\|_2 \leq \|E\|_\infty$.

- The two norms usually have the same scale, e.g. all-one matrices.

Our result: ℓ^∞ Eigenvector Perturbation

- One more thing: need different matrix norms for perturbation E.
- For symmetric $E \in \mathbf{R}^{d\times d}$, define

$$\|E\|_\infty = \max_i \sum_j^d |E_{ij}|.$$

- A trivial bound $\|E\|_2 \leq \|E\|_\infty$.
- The two norms usually have the same scale, e.g. all-one matrices.

Our result: ℓ^∞ Eigenvector Perturbation

- One more thing: need different matrix norms for perturbation E.
- For symmetric $E \in \mathbf{R}^{d\times d}$, define

$$\|E\|_\infty = \max_i \sum_j^d |E_{ij}|.$$

- A trivial bound $\|E\|_2 \leq \|E\|_\infty$.
- The two norms usually have the same scale, e.g. all-one matrices.

Our result: ℓ^∞ Eigenvector Perturbation

- Suppose A, E are symmetric. Recall that γ is the eigengap.

Theorem (Fan, Wang, and Zhong [2016b])

There exists $C = C(\mu(V), r) = O(\mu(V)^{1.5} r^{3.5})$ such that

$$\max_{1 \le k \le r} \|v_k - \eta_k \widetilde{v}_k\|_\infty \le C(\mu(V), r) \frac{\|E\|_\infty}{\gamma \sqrt{d}},$$

when $\gamma > C\|E\|_\infty$, for suitable signs $\eta_k \in \{\pm 1\}$.

- A similar result for SVD.
- The condition $\gamma > C\|E\|_\infty$ is mild: similar to ℓ^2 bound.

Our result: ℓ^∞ Eigenvector Perturbation

- Suppose A, E are symmetric. Recall that γ is the eigengap.

Theorem (Fan, Wang, and Zhong [2016b])

There exists $C = C(\mu(V), r) = O(\mu(V)^{1.5} r^{3.5})$ such that

$$\max_{1 \le k \le r} \|v_k - \eta_k \widetilde{v}_k\|_\infty \le C(\mu(V), r) \frac{\|E\|_\infty}{\gamma\sqrt{d}},$$

when $\gamma > C\|E\|_\infty$, for suitable signs $\eta_k \in \{\pm 1\}$.

- A similar result for SVD.

- The condition $\gamma > C\|E\|_\infty$ is mild: similar to ℓ^2 bound.

Our result: ℓ^∞ Eigenvector Perturbation

- Suppose A, E are symmetric. Recall that γ is the eigengap.

Theorem (Fan, Wang, and Zhong [2016b])

There exists $C = C(\mu(V), r) = O(\mu(V)^{1.5} r^{3.5})$ *such that*

$$\max_{1 \le k \le r} \|v_k - \eta_k \widetilde{v}_k\|_\infty \le C(\mu(V), r) \frac{\|E\|_\infty}{\gamma \sqrt{d}},$$

when $\gamma > C\|E\|_\infty$, *for suitable signs* $\eta_k \in \{\pm 1\}$.

- A similar result for SVD.
- The condition $\gamma > C\|E\|_\infty$ is mild: similar to ℓ^2 bound.

Simulation: Eigenvector Perturbation

- Setting:
  - d run from 200 to 2000 with increment 200.
  - $A = \sum_{k=1}^{3}(4-k)\gamma v_k v_k^T$; v_k is an eigenvector of an iid normal random matrix.
  - Generating E: (a) random number in $[0, L]$ by randomly selecting s entries each row; (b) $E_{ij} = L'\rho^{|i-j|}$.

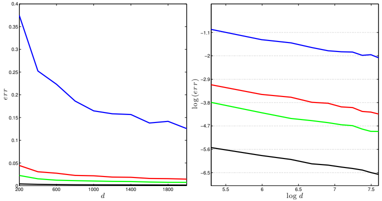

Figure: *The slope is around* -0.5. *Blue:* $\gamma = 10$*; red:* $\gamma = 50$*; green:* $\gamma = 100$*; and black:* $\gamma = 500$. *We report the largest error over* 100 *runs.*

Simulation: Eigenvector Perturbation

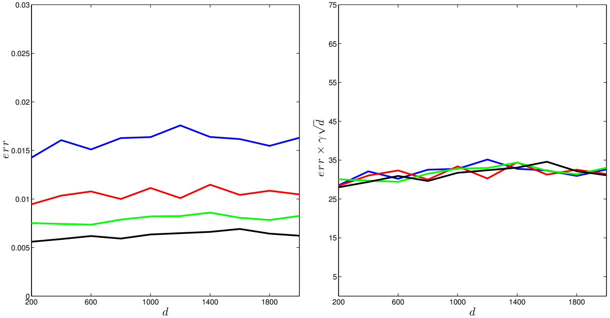

Figure: $\gamma\sqrt{d}$ *is fixed for each line. The right plot shows the error multiplied by* $\gamma\sqrt{d}$ *against* d*. Blue:* $\gamma\sqrt{d} = 2000$*; red:* $\gamma\sqrt{d} = 3000$*; green:* $\gamma\sqrt{d} = 4000$*; and black:* $\gamma\sqrt{d} = 5000$*. We report the largest error over* 100 *runs.*

Simulation: Eigenvector Perturbation

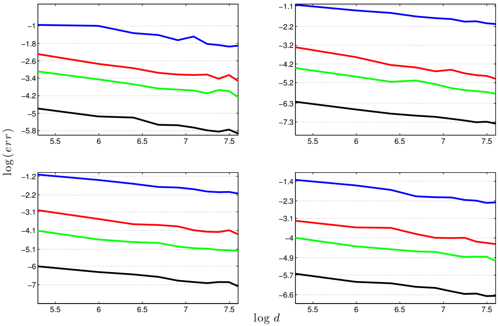

Figure: $\log(err)$ *vs.* $\log(d)$. *Top left:* $L = 10, s = 3$*; top right:* $L = 0.6, s = 50$*; bottom left:* $L' = 1.5, \rho = 0.9$*; bottom right:* $L' = 7.5, \rho = 0.5$*. The slopes are around* -0.5*. Blue:* $\gamma = 10$*; red:* $\gamma = 50$*; green:* $\gamma = 100$*; black:* $\gamma = 500$.

Outline

Robust Covariance Estimation: Factors Known

- Factor Model:
$$y_i = Bf_i + u_i$$
where $y_i, u_i \in \mathbf{R}^d$, $f \in \mathbf{R}^r$ and $B \in \mathbf{R}^{d \times r}$.

- $\{y_i\}_{i=1}^n, \{f_i\}_{i=1}^n$ are i.i.d. observed over $i = 1, \ldots, n$.

- Assuming $\mathrm{Cov}(f_i, u_i) = 0$:
$$\Sigma = B\Sigma_f B^T + \Sigma_u, \quad B \in \mathbf{R}^{d\times r}, \Sigma \in \mathbf{R}^{d\times d}, \Sigma_f \in \mathbf{R}^{r\times r},$$
where r is a constant and Σ_u is sparse.

- Goal: estimate Σ from observations.
- **Challenge:** y_i and u_i may possibly have heavy tails.

Robust Covariance Estimation: Factors Known

- Factor Model:
$$y_i = Bf_i + u_i$$
where $y_i, u_i \in \mathbf{R}^d$, $f \in \mathbf{R}^r$ and $B \in \mathbf{R}^{d\times r}$.

- $\{y_i\}_{i=1}^n, \{f_i\}_{i=1}^n$ are i.i.d. observed over $i = 1, \ldots, n$.

- Assuming $\mathrm{Cov}(f_i, u_i) = 0$:
$$\Sigma = B\Sigma_f B^T + \Sigma_u, \quad B \in \mathbf{R}^{d\times r}, \Sigma \in \mathbf{R}^{d\times d}, \Sigma_f \in \mathbf{R}^{r\times r},$$
where r is a constant and Σ_u is sparse.

- Goal: estimate Σ from observations.
- **Challenge:** y_i and u_i may possibly have heavy tails.

Robust Covariance Estimation: Factors Known

- Factor Model:
$$y_i = Bf_i + u_i$$
where $y_i, u_i \in \mathbf{R}^d$, $f \in \mathbf{R}^r$ and $B \in \mathbf{R}^{d\times r}$.

- $\{y_i\}_{i=1}^n, \{f_i\}_{i=1}^n$ are i.i.d. observed over $i = 1, \dots, n$.

- Assuming $\text{Cov}(f_i, u_i) = 0$:
$$\Sigma = B\Sigma_f B^T + \Sigma_u, \quad B \in \mathbf{R}^{d\times r}, \Sigma \in \mathbf{R}^{d\times d}, \Sigma_f \in \mathbf{R}^{r\times r},$$
where r is a constant and Σ_u is sparse.

- Goal: estimate Σ from observations.
- **Challenge:** y_i and u_i may possibly have heavy tails.

Robust Covariance Estimation: Factors Known

- Factor Model:
$$y_i = Bf_i + u_i$$
where $y_i, u_i \in \mathbf{R}^d$, $f \in \mathbf{R}^r$ and $B \in \mathbf{R}^{d\times r}$.

- $\{y_i\}_{i=1}^n, \{f_i\}_{i=1}^n$ are i.i.d. observed over $i = 1, \ldots, n$.

- Assuming $\mathrm{Cov}(f_i, u_i) = 0$:
$$\Sigma = B\Sigma_f B^T + \Sigma_u, \quad B \in \mathbf{R}^{d\times r}, \Sigma \in \mathbf{R}^{d\times d}, \Sigma_f \in \mathbf{R}^{r\times r},$$
where r is a constant and Σ_u is sparse.

- Goal: estimate Σ from observations.
- **Challenge:** y_i and u_i may possibly have heavy tails.

Robust Covariance Estimation: Factors Known

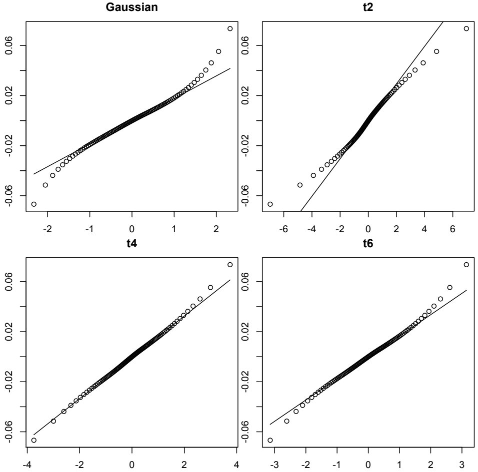

Figure: *Q-Q plots show heavy tails of data: x-axis: base distribution quantiles, y-axis: return data quantiles.*

- **Antidote**: Huber's M-estimator with suitable a diverging parameter; bias-variance tradeoff.
- For any i.i.d. $Z_1, \dots, Z_n$ with $\mu^\star = \mathbb{E} Z_i$. Let $l_\alpha(x) = 2\alpha|x| - \alpha^2$ when $|x| \geq \alpha$ and x^2 when $|x| \leq \alpha$.

$$\widehat{\mu} = \operatorname{argmin}_\mu \sum_{t=1}^{n} l_\alpha(Z_t - \mu)$$

Theorem (Fan, Li, and Wang [2014])

Suppose $\varepsilon \in (0,1)$ *and* $n \geq 8\log(\varepsilon^{-1})$*. Choose* $\alpha = \sqrt{(nv^2)/\log(\varepsilon^{-1})}$*, where* v^2 *is an upper bound of* $\mathrm{cov}(Z_t)$*. Then,*

$$P\left(|\widehat{\mu} - \mu^\star| \leq 4v\sqrt{\frac{\log(\varepsilon^{-1})}{n}}\right) \geq 1 - 2\varepsilon. \tag{1}$$

- Similar result: Catoni [2012]

Robust Covariance Estimation: Factors Known

- **Antidote**: Huber's *M*-estimator with suitable a diverging parameter; bias-variance tradeoff.
- For any i.i.d. $Z_1, \dots, Z_n$ with $\mu^\star = \mathbb{E}Z_i$. Let $l_\alpha(x) = 2\alpha|x| - \alpha^2$ when $|x| \geq \alpha$ and x^2 when $|x| \leq \alpha$.

$$\widehat{\mu} = \text{argmin}_\mu \sum_{t=1}^{n} l_\alpha(Z_t - \mu)$$

Theorem (Fan, Li, and Wang [2014])

Suppose $\varepsilon \in (0,1)$ *and* $n \geq 8\log(\varepsilon^{-1})$. *Choose* $\alpha = \sqrt{(nv^2)/\log(\varepsilon^{-1})}$, *where* v^2 *is an upper bound of* $\text{cov}(Z_t)$. *Then,*

$$P\left(|\widehat{\mu} - \mu^\star| \leq 4v\sqrt{\frac{\log(\varepsilon^{-1})}{n}}\right) \geq 1 - 2\varepsilon. \qquad (1)$$

- Similar result: Catoni [2012]

Robust Covariance Estimation: Factors Known

- **Antidote**: Huber's *M*-estimator with suitable a diverging parameter; bias-variance tradeoff.
- For any i.i.d. $Z_1, \ldots, Z_n$ with $\mu^\star = \mathbb{E} Z_i$. Let $l_\alpha(x) = 2\alpha|x| - \alpha^2$ when $|x| \geq \alpha$ and x^2 when $|x| \leq \alpha$.

$$\widehat{\mu} = \text{argmin}_\mu \sum_{t=1}^{n} l_\alpha(Z_t - \mu)$$

Theorem (Fan, Li, and Wang [2014])

Suppose $\varepsilon \in (0,1)$ *and* $n \geq 8\log(\varepsilon^{-1})$. *Choose* $\alpha = \sqrt{(nv^2)/\log(\varepsilon^{-1})}$, *where* v^2 *is an upper bound of* $cov(Z_t)$. *Then,*

$$P\left(|\widehat{\mu} - \mu^\star| \leq 4v\sqrt{\frac{\log(\varepsilon^{-1})}{n}}\right) \geq 1 - 2\varepsilon. \qquad (1)$$

- Similar result: Catoni [2012]

Robust Covariance Estimation: Factors Known

- **Antidote**: Huber's *M*-estimator with suitable a diverging parameter; bias-variance tradeoff.
- For any i.i.d. $Z_1, \ldots, Z_n$ with $\mu^\star = \mathbb{E}Z_i$. Let $l_\alpha(x) = 2\alpha|x| - \alpha^2$ when $|x| \geq \alpha$ and x^2 when $|x| \leq \alpha$.

$$\widehat{\mu} = \mathrm{argmin}_\mu \sum_{t=1}^{n} l_\alpha(Z_t - \mu)$$

Theorem (Fan, Li, and Wang [2014])

Suppose $\varepsilon \in (0,1)$ and $n \geq 8\log(\varepsilon^{-1})$. Choose $\alpha = \sqrt{(nv^2)/\log(\varepsilon^{-1})}$, where v^2 is an upper bound of $\mathrm{cov}(Z_t)$. Then,

$$P\left(|\widehat{\mu} - \mu^\star| \leq 4v\sqrt{\frac{\log(\varepsilon^{-1})}{n}}\right) \geq 1 - 2\varepsilon. \quad (1)$$

- Similar result: Catoni [2012]

Robust Covariance Estimation: Factors Known

- Cure of high dimensionality: entry-wise estimation.

- Now we have $\widehat{\Sigma}$ ✓, then:

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Need to exploit the structure and have a refined estimate.

- **Question:** How to denoise? How to disentangle?

- Blessing of pervarsive assumption: the top r eigenvalues of Σ grows linearly with d; and the elements of B are uniformly bounded.

Robust Covariance Estimation: Factors Known

- Cure of high dimensionality: entry-wise estimation.
- Now we have $\widehat{\Sigma}$ ✓, then:

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Need to exploit the structure and have a refined estimate.
- **Question:** How to denoise? How to disentangle?
- Blessing of pervarsive assumption: the top r eigenvalues of Σ grows linearly with d; and the elements of B are uniformly bounded.

- Cure of high dimensionality: entry-wise estimation.
- Now we have $\widehat{\Sigma}$ ✓, then:

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Need to exploit the structure and have a refined estimate.
- **Question:** How to denoise? How to disentangle?
- Blessing of pervarsive assumption: the top r eigenvalues of Σ grows linearly with d; and the elements of B are uniformly bounded.

Robust Covariance Estimation: Factors Known

- Cure of high dimensionality: entry-wise estimation.
- Now we have $\widehat{\Sigma}$ ✓, then:

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Need to exploit the structure and have a refined estimate.
- **Question:** How to denoise? How to disentangle?
- Blessing of pervarsive assumption: the top r eigenvalues of Σ grows linearly with d; and the elements of B are uniformly bounded.

Robust Covariance Estimation: Factors Known

- Cure of high dimensionality: entry-wise estimation.
- Now we have $\widehat{\Sigma}$ ✓, then:

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Need to exploit the structure and have a refined estimate.
- **Question:** How to denoise? How to disentangle?
- Blessing of pervarsive assumption: the top r eigenvalues of Σ grows linearly with d; and the elements of B are uniformly bounded.

Robust Covariance Estimation: Factors Known

- When factors $f_1, \ldots, f_n$ are observed, a simple trick works.
- Let $z_i^T = (y_i^T, f_i^T)$, and estimate $\mathrm{Cov}(z_i)$ robustly.
- Since $\Sigma_u = \Sigma - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{uy}$, we can do

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}$$

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

- Finally,

$$\widehat{\Sigma}^{\mathcal{T}} = \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy} + \mathcal{T}(\widehat{\Sigma}_u)$$

Robust Covariance Estimation: Factors Known

- When factors $f_1, \ldots, f_n$ are observed, a simple trick works.
- Let $z_i^T = (y_i^T, f_i^T)$, and estimate $\mathrm{Cov}(z_i)$ robustly.
- Since $\Sigma_u = \Sigma - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{uy}$, we can do

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}$$

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

- Finally,

$$\widehat{\Sigma}^{\mathcal{T}} = \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy} + \mathcal{T}(\widehat{\Sigma}_u)$$

- When factors $f_1, \ldots, f_n$ are observed, a simple trick works.
- Let $z_i^T = (y_i^T, f_i^T)$, and estimate $\text{Cov}(z_i)$ robustly.
- Since $\Sigma_u = \Sigma - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{uy}$, we can do

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}$$

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

- Finally,

$$\widehat{\Sigma}^{\mathcal{T}} = \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy} + \mathcal{T}(\widehat{\Sigma}_u)$$

- When factors $f_1, \ldots, f_n$ are observed, a simple trick works.
- Let $z_i^T = (y_i^T, f_i^T)$, and estimate $\mathrm{Cov}(z_i)$ robustly.
- Since $\Sigma_u = \Sigma - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{uy}$, we can do

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}$$

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\widehat{\Sigma}_u}$$

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

- Finally,

Robust Covariance Estimation: Factors Known

- When factors $f_1, \ldots, f_n$ are observed, a simple trick works.
- Let $z_i^T = (y_i^T, f_i^T)$, and estimate $\mathrm{Cov}(z_i)$ robustly.
- Since $\Sigma_u = \Sigma - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{uy}$, we can do

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}$$

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

- Finally,

$$\widehat{\Sigma}^{\mathcal{T}} = \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy} + \mathcal{T}(\widehat{\Sigma}_u)$$

Robust Covariance Estimation: Factors Known

- When factors $f_1, \ldots, f_n$ are observed, a simple trick works.
- Let $z_i^T = (y_i^T, f_i^T)$, and estimate $\text{Cov}(z_i)$ robustly.
- Since $\Sigma_u = \Sigma - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{uy}$, we can do

$$\widehat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}$$

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

- Finally,

$$\widehat{\Sigma}^{\mathcal{T}} = \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy} + \mathcal{T}(\widehat{\Sigma}_u)$$

Robust Covariance Estimation: Factors Known

- **Step 1:** Obtain a robust estimate of $\mathrm{Cov}(z_i)$ in an entry-wise way.

- **Step 2:** Compute

$$\widehat{\Sigma}_u = \widehat{\Sigma}_{yy} - \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy},$$

and apply adaptive thresholding

$$\mathcal{T}(\widehat{\Sigma}_u)_{ij} = \begin{cases} (\widehat{\Sigma}_u)_{ij}, & i = j \\ s_{ij}((\widehat{\Sigma}_u)_{ij})\mathbf{1}((\widehat{\Sigma}_u)_{ij} \geq \tau_{ij}), & i \neq j, \end{cases}$$

where $\tau_{ij} = \tau((\widehat{\Sigma}_u)_{ii}(\widehat{\Sigma}_u)_{jj})^{1/2}, \tau \asymp \sqrt{\log p/n}$.

- **Step 3:** Final estimator

$$\widehat{\Sigma}^{\mathcal{T}} = \widehat{\Sigma}_{yu}\widehat{\Sigma}_{uu}^{-1}\widehat{\Sigma}_{uy} + \mathcal{T}(\widehat{\Sigma}_u).$$

Robust Covariance Estimation: Factors Known

- Let $m_q = \max_{i \le p} \sum_{j \le p} (\Sigma_u)_{ij}^q$ for some $q \in [0, 1]$. Assume pervasiveness, bounded fourth moments and $\|\Sigma_u\|, \|\Sigma_f\|$ bounded above and below from 0.

Theorem (Fan, Wang, and Zhong [2016a])

If $m_q(\log d/n)^{(1-q)/2} = o(1)$, then

$$\|\widehat{\Sigma}^{\mathcal{T}} - \Sigma\|_\infty = O_P\left(\sqrt{\frac{\log d}{n}}\right),$$

$$\|\widehat{\Sigma}^{\mathcal{T}} - \Sigma\|_\Sigma = O_P\left(\frac{\sqrt{d}\log d}{n} + m_q\left(\frac{\log d}{n}\right)^{(1-q)/2}\right),$$

$$\|(\widehat{\Sigma}^{\mathcal{T}})^{-1} - \Sigma^{-1}\| = O_P\left(m_q\left(\frac{\log d}{n}\right)^{(1-q)/2}\right),$$

where $\|A\|_\Sigma = d^{-1/2}\|\Sigma^{-1/2}A\Sigma^{-1/2}\|_F$ is the relative Frobenius norm.

Question:

What if the factors are unobserved?

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\text{how to estimate?}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

Need:

a sharp low rank estimate directly from $\widehat{\Sigma}$.

Question:

What if the factors are unobserved?

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\text{how to estimate?}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

Need:

a sharp low rank estimate directly from $\widehat{\Sigma}$.

Outline

1. Matrix Eigenvector Perturbation
   - Davis-Kahan Theorem
   - Our result: ℓ^∞ Eigenvector Perturbation

2. Robust Covariance Estimation via Factor Models
   - Known factors
   - Factors unknown

3. Simulations

4. Real Data Analysis: Portfolio Risk Estimation

Robust Covariance Estimation: Factors Unknown

- Recall factor model:

$$y_i = Bf_i + u_i$$

where $y_i, u_i \in \mathbf{R}^d$, $f \in \mathbf{R}^r$ and $B \in \mathbf{R}^{d \times r}$.

- Identifiability: suppose $\Sigma_f = I_r$.

- What if factors $f_1, \ldots, f_n$ are unobserved? How to estimate the low-rank part?

$$\widehat{\Sigma} = \underbrace{BB^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Pervasiveness $\Leftrightarrow$ B spiked eigenvalues + matrix incoherence!

Robust Covariance Estimation: Factors Unknown

- Recall factor model:

$$y_i = Bf_i + u_i$$

where $y_i, u_i \in \mathbf{R}^d$, $f \in \mathbf{R}^r$ and $B \in \mathbf{R}^{d\times r}$.

- Identifiability: suppose $\Sigma_f = I_r$.

- What if factors $f_1, \ldots, f_n$ are unobserved? How to estimate the low-rank part?

$$\widehat{\Sigma} = \underbrace{BB^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Pervasiveness $\Leftrightarrow$ B spiked eigenvalues + matrix incoherence!

Robust Covariance Estimation: Factors Unknown

- Recall factor model:

$$y_i = Bf_i + u_i$$

where $y_i, u_i \in \mathbf{R}^d$, $f \in \mathbf{R}^r$ and $B \in \mathbf{R}^{d\times r}$.

- Identifiability: suppose $\Sigma_f = I_r$.

- What if factors $f_1, \ldots, f_n$ are unobserved? How to estimate the low-rank part?

$$\widehat{\Sigma} = \underbrace{BB^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\widehat{\Sigma} - \Sigma)}_{\text{Noise}}.$$

- Pervasiveness $\Leftrightarrow$ B spiked eigenvalues + matrix incoherence!

Robust Covariance Estimation: Factors Unknown

- **Solution**: directly apply PCA to $\widehat{\Sigma}$.
- Low rank $A := BB^T$ and perturbation $E := \Sigma_u + (\widehat{\Sigma} - \Sigma)$. A sharp and uniform bound:

$$\|\widehat{v}_i - v_i\|_\infty = O_P\left(\sqrt{\frac{\log d}{nd}} + \frac{1}{\sqrt{d}}\right), \quad i = 1, \ldots, r$$

- So we have sharp uniform bound.

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\text{use PCA}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

- **Solution**: directly apply PCA to $\widehat{\Sigma}$.
- Low rank $A := BB^T$ and perturbation $E := \Sigma_u + (\widehat{\Sigma} - \Sigma)$.

  A sharp and uniform bound:

$$\|\widehat{v}_i - v_i\|_\infty = O_P\Big(\sqrt{\frac{\log d}{nd}} + \frac{1}{\sqrt{d}}\Big), \quad i = 1, \ldots, r$$

- So we have sharp uniform bound.

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\text{use PCA}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

Robust Covariance Estimation: Factors Unknown

- **Solution**: directly apply PCA to $\widehat{\Sigma}$.

- Low rank $A := BB^T$ and perturbation $E := \Sigma_u + (\widehat{\Sigma} - \Sigma)$.

  A sharp and uniform bound:

$$\|\widehat{v}_i - v_i\|_\infty = O_P\left(\sqrt{\frac{\log d}{nd}} + \frac{1}{\sqrt{d}}\right), \quad i = 1, \ldots, r$$

- So we have sharp uniform bound.

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\text{use PCA}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

Robust Covariance Estimation: Factors Unknown

- **Solution**: directly apply PCA to $\widehat{\Sigma}$.

- Low rank $A := BB^T$ and perturbation $E := \Sigma_u + (\widehat{\Sigma} - \Sigma)$.
  A sharp and uniform bound:

$$\|\widehat{v}_i - v_i\|_\infty = O_P\Big(\sqrt{\frac{\log d}{nd}} + \frac{1}{\sqrt{d}}\Big), \quad i = 1, \ldots, r$$

- So we have sharp uniform bound.

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\text{use PCA}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

Robust Covariance Estimation: Factors Unknown

- **Solution**: directly apply PCA to $\widehat{\Sigma}$.
- Low rank $A := BB^T$ and perturbation $E := \Sigma_u + (\widehat{\Sigma} - \Sigma)$.
  A sharp and uniform bound:

$$\|\widehat{v}_i - v_i\|_\infty = O_P\Big(\sqrt{\frac{\log d}{nd}} + \frac{1}{\sqrt{d}}\Big), \quad i = 1, \ldots, r$$

- So we have sharp uniform bound.

$$\widehat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\text{use PCA}} = \underbrace{\Sigma_u + (\widehat{\Sigma} - \Sigma)}_{\mathcal{T}(\widehat{\Sigma}_u)\text{:thresholding}}$$

Robust Covariance Estimation: Factors Unknown

- **Step 1:** Estimate of $\Sigma = \mathrm{Cov}(y_i)$ robustly in an entry-wise way, and denote it as $\widehat{\Sigma}$.

- **Step 2:** Compute

$$\widehat{\Sigma}_u = \widehat{\Sigma}_y - \widehat{U}\widehat{\Lambda}\widehat{U}^T,$$

where $\widehat{U}$ and $\widehat{\Lambda}$ are given by top r eigen-decomposition. Then apply adaptive thresholding on $\widehat{\Sigma}_u$.

- **Step 3:** Final estimator

$$\widehat{\Sigma}^{\mathcal{T}} = \mathcal{T}(\widehat{\Sigma}_u) + \widehat{U}\widehat{\Lambda}\widehat{U}^T.$$

Theorem (Fan, Wang, and Zhong [2016b])

Let $w_n = \sqrt{\log d / n} + 1/\sqrt{d}$. Under the same assumptions of Theorem 4, with the choice of parameter $\tau \asymp w_n$ and if $m_q w_n^{1-q} = o(1)$ we have

$$\|\widehat{\Sigma}^{\top} - \Sigma\|_{\max} = O_P\left(w_n\right),$$

$$\|\widehat{\Sigma}^{\top} - \Sigma\|_{\Sigma} = O_P\left(\frac{\sqrt{d}\log d}{n} + m_d w_n^{1-q}\right),$$

$$\|(\widehat{\Sigma}^{\top})^{-1} - \Sigma^{-1}\|_2 = O_P\left(m_d w_n^{1-q}\right).$$

Take-home message:

Power of ℓ^{∞} perturbation bound:

Uniform and sharp bound when estimating low rank structure.

Take-home message:

Power of ℓ^∞ perturbation bound:

Uniform and sharp bound when estimating low rank structure.

1. Matrix Eigenvector Perturbation
   - Davis-Kahan Theorem
   - Our result: ℓ^∞ Eigenvector Perturbation
2. Robust Covariance Estimation via Factor Models
   - Known factors
   - Factors unknown
3. **Simulations**
4. Real Data Analysis: Portfolio Risk Estimation

Simulation: Robust Covariance Estimation

- Setting:
  - Simulate n samples of (f_t^T, u_t^T) from (i) a multivariate t-distribution (ii) an i.i.d. t-distribution, with covariance diag$\{I_r, 5I_d\}$.
  - Each element of B is an i.i.d. standard normal. Compute $y_t = Bf_t + u_t$.
  - Set $n = d/2$; degree of freedom $\nu = 3, 5$ and ∞.
- Compare our robust estimator $\widehat{\Sigma}^R$ with (a) $\widehat{\Sigma}^S$: sample covariance based method (b) $\widehat{\Sigma}^K$: Kendall's tau based method.
- Calcuate relative errors $\widehat{\Sigma} - \Sigma$ under various norms, e.g. $\|\widehat{\Sigma}^R - \Sigma\| / \|\widehat{\Sigma}^S - \Sigma\|$.
- The median errors and their IQR over 100 simulations are reported.

Simulation: Robust Covariance Estimation

- Setting:
  - Simulate n samples of (f_t^T, u_t^T) from (i) a multivariate t-distribution (ii) an i.i.d. t-distribution, with covariance $\mathrm{diag}\{I_r, 5I_d\}$.
  - Each element of B is an i.i.d. standard normal. Compute $y_t = Bf_t + u_t$.
  - Set $n = d/2$; degree of freedom $\nu = 3, 5$ and ∞.
- Compare our robust estimator $\widehat{\Sigma}^R$ with (a) $\widehat{\Sigma}^S$: sample covariance based method (b) $\widehat{\Sigma}^K$: Kendall's tau based method.
- Calcuate relative errors $\widehat{\Sigma} - \Sigma$ under various norms, e.g. $\|\widehat{\Sigma}^R - \Sigma\| / \|\widehat{\Sigma}^S - \Sigma\|$.
- The median errors and their IQR over 100 simulations are reported.

Simulation: Robust Covariance Estimation

- Setting:
  - Simulate n samples of (f_t^T, u_t^T) from (i) a multivariate t-distribution (ii) an i.i.d. t-distribution, with covariance $\mathrm{diag}\{I_r, 5I_d\}$.
  - Each element of B is an i.i.d. standard normal. Compute $y_t = Bf_t + u_t$.
  - Set $n = d/2$; degree of freedom $\nu = 3, 5$ and ∞.
- Compare our robust estimator $\widehat{\Sigma}^R$ with (a) $\widehat{\Sigma}^S$: sample covariance based method (b) $\widehat{\Sigma}^K$: Kendall's tau based method.
- Calcuate relative errors $\widehat{\Sigma} - \Sigma$ under various norms, e.g. $\|\widehat{\Sigma}^R - \Sigma\| / \|\widehat{\Sigma}^S - \Sigma\|$.
- The median errors and their IQR over 100 simulations are reported.

- Setting: observed factors and multivariate t-distribution.

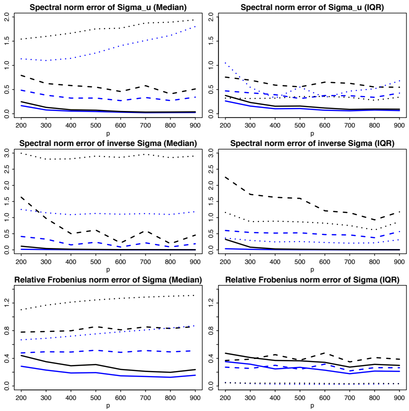

Figure: *Blue: relative error of* $\widehat{\Sigma}_z^R$; *black: relative error of* $\widehat{\Sigma}_z^K$. *Degree of freedom:* $df = 3$ *(solid), 5 (dashed) and* ∞ *(dotted).*

- Setting: observed factors and i.i.d. t-distribution.

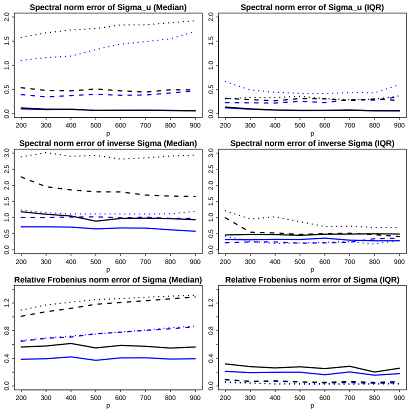

Figure: *Blue: relative error of $\widehat{\Sigma}_z^R$; black: relative error of $\widehat{\Sigma}_z^K$. Degree of freedom: $df = 3$ (solid), 5 (dashed) and ∞ (dotted).*

- Setting: unobserved factors and multivariate t-distribution.

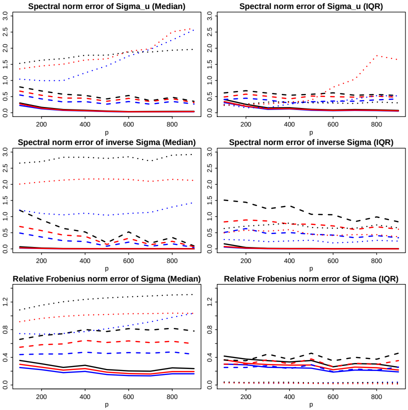

Figure: *Blue: relative error of* $\widehat{\Sigma}_z^R$*; black and red: relative error of* $\widehat{\Sigma}_z^K$*. Degree of freedom: df* $= 3$ *(solid), 5 (dashed) and* ∞ *(dotted).*

Simulation: Robust Covariance Estimation

- Setting: unobserved factors and multivariate t-distribution.

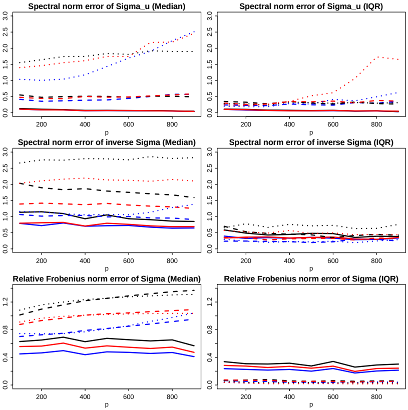

Figure: *Blue: relative error of $\widehat{\Sigma}_z^R$; black and red: relative error of $\widehat{\Sigma}_z^K$. Degree of freedom: df $= 3$ (solid), 5 (dashed) and ∞ (dotted).*

Real Data Analysis: Portfolio Risk Estimation

- Daily returns of 393 stocks from S&P 500 index from 2005 to 2013.
- Factors: Fama-French three-factor model.
- Y: 393×2013 data matrix, and F: 2013×3.
- Let Σ be the true covariance matrix of Y. Then $w^T \Sigma w$ is the porfolio risk given weights w.
- Compare risk estimation errors:

$$R^R(w) = \frac{1}{2013} \sum_{t=1}^{2013} \left| w^T \widehat{\Sigma}_t^R w - (w^T \gamma_t)^2 \right|,$$

$$R^S(w) = \frac{1}{2013} \sum_{t=1}^{2013} \left| w^T \widehat{\Sigma}_t^S w - (w^T \gamma_t)^2 \right|.$$

Real Data Analysis: Portfolio Risk Estimation

- Daily returns of 393 stocks from S&P 500 index from 2005 to 2013.
- Factors: Fama-French three-factor model.
- Y: 393×2013 data matrix, and F: 2013×3.
- Let Σ be the true covariance matrix of Y. Then $w^T \Sigma w$ is the porfolio risk given weights w.
- Compare risk estimation errors:

$$R^R(w) = \frac{1}{2013} \sum_{t=1}^{2013} \left| w^T \widehat{\Sigma}_t^R w - (w^T \gamma_t)^2 \right|,$$

$$R^S(w) = \frac{1}{2013} \sum_{t=1}^{2013} \left| w^T \widehat{\Sigma}_t^S w - (w^T \gamma_t)^2 \right|.$$

Real Data Analysis: Portfolio Risk Estimation

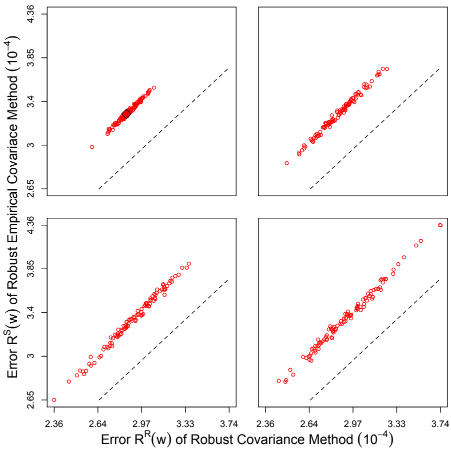

Figure: $(R^R(w), R^S(w))$ *for multiple randomly generated* w*. Error comparison with different exposure* c*. (upper left: no short selling; upper right:* $c = 1.4$*; lower left:* $c = 1.8$*; lower right:* $c = 2.2$*).*

References I

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

Olivier Catoni. Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 48, pages 1148–1185. Institut Henri Poincaré, 2012.

Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Jianqing Fan, Quefeng Li, and Yuyan Wang. Robust estimation of high-dimensional mean regression. *arXiv preprint arXiv:1410.2150*, 2014.

Jianqing Fan, Weichen Wang, and Yiqiao Zhong. Robust covariance estimation for approximate factor models. *arXiv preprint arXiv:1602.00719*, 2016a.

Jianqing Fan, Weichen Wang, and Yiqiao Zhong. An ℓ^∞ eigenvector perturbation bound and its application to robust covariance estimation. *arXiv preprint arXiv:1603.03516*, 2016b.

Thank you!