Statistical inference in a spiked population model

Jian-feng Yao

Joint work with Weiming Li (Beijing), Damien Passemier (Rennes)
1 Spiked eigenvalues: an example

2 Inference on spikes: determination of their number $q_0$
   Known results on spiked population
   Estimator of $q_0$
   Discussions on the estimator $\hat{q}_0$
   Application to S&P stocks data

3 Inference of the bulk spectrum
   The problem and existing methods
   A generalized expectation based method
   Asymptotic properties of the GEE estimator
   Application to S&P 500 stocks data
1) Spiked eigenvalues: an example

- SP 500 daily stock prices; \( p = 488 \) stocks;

- \( n = 1000 \) daily returns \( r_t(i) = \log p_t(i)/p_{t-1}(i) \) from 2007-09-24 to 2011-09-12;
The sample correlation matrix

Let the SCM $(488 \times 488)$

\[ S_n = \frac{1}{n} \sum_{t=1}^{n} (r_t - \bar{r})(r_t - \bar{r})^T. \]

We consider the sample correlation matrix $R_n$ with

\[ R_n(i, j) = \frac{S_n(i, j)}{[S_n(i, i)S_n(j, j)]^{1/2}}. \]

The 10 largest and 10 smallest eigenvalues of $R_n$ are:

- 237.95801 4.8568703 ...
- 17.762811 4.394394 ...
- 14.002838 3.4999069 ...
- 8.7633113 3.0880089 ...
- 5.2995321 2.7146658 ...
- 0.0212137 0.0178129
- 0.0205001 0.0173591
- 0.0198287 0.0164425
- 0.0194216 0.0154849
- 0.0190959 0.0147696
Plots of sample eigenvalues

Left: $488 - 1 = 487$ eigenvalues

Right: $488 - 10 = 478$ eigenvalues

⇒ the point: sample eigenvalues = bulk + spikes

⇒ Analysis and estimation of spikes + bulk
A generic model

Random factor model

\[ x_t = \sum_{k=1}^{q_0} a_k s_t(t) + \varepsilon_t = As_t + \varepsilon_t, \]

- \( s_t = (s_t(1), \ldots, s_t(q_0)) \in \mathbb{R}^{q_0} \) are \( q_0 < p \) standardised random signals/factors,
- \( A = (a_1, \ldots, a_{q_0}), p \times q_0 \) deterministic matrix of factor loadings
- \( \varepsilon_t \) is an independent \( p \)-dimensional noise sequence, with a diagonal covariance matrix: \( \Psi = \text{cov}(\varepsilon_t) = \text{diag}\{\sigma_1^2, \ldots, \sigma_p^2\} \).

Therefore,

\[ \Sigma = \text{cov}(x_t) = AA^* + \Psi. \]

- this model is very old; has wide range of application fields: psychology, chemometrics, signal processing, economics, etc.
2). Inference on spikes

a). Known results

**Spiked population model**

Population covariance matrix:

\[
\Sigma = \text{Cov}[x_t] = AA^* + \sigma^2 I_p ,
\]

with eigenvalues

\[
\text{spec}(\Sigma) = (\sigma^2 + \alpha_1', \ldots, \sigma^2 + \alpha_{q_0}', \underbrace{\sigma^2, \ldots, \sigma^2}_{p-q_0}) ,
\]

where

- \( \alpha_1' \geq \alpha_2' \geq \cdots \geq \alpha_{q_0}' > 0 \) are non null eigenvalues of \( AA^* \),

or equivalently

\[
\text{spec}(\Sigma) = \sigma^2 \times (\alpha_1, \ldots, \alpha_{q_0}, \underbrace{1, \ldots, 1}_{p-q_0}) ,
\]

with

\[
\alpha_i = 1 + \alpha_i'/\sigma^2 .
\]
Asymptotic framework and assumptions

1. \( p, n \to +\infty \) such that \( p/n \to c \);

2. The population covariance matrix has \( K \) spikes \( \alpha_1 > \cdots > \alpha_K \) with respective multiplicity numbers \( n_i \), i.e.

\[
\text{spec}(\Sigma) = \sigma^2 \left( \underbrace{\alpha_1, \ldots, \alpha_1}_{n_1}, \underbrace{\alpha_2, \ldots, \alpha_2}_{n_2}, \ldots, \underbrace{\alpha_K, \cdots, \alpha_K}_{n_K}, \underbrace{1, \cdots, 1}_{p-q_0} \right);
\]

\[
[ n_1 + \cdots + n_K = q_0 ];
\]

3. \( \alpha_K > 1 + \sqrt{c} \) (detection level).

4. \( \mathbb{E}(|x_{ij}^4|) < +\infty \).
Convergence of spike eigenvalues

Consider the sample covariance matrix \( S_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^* \), with sample eigenvalues: \( \lambda_{n,1} \geq \lambda_{n,2} \geq \cdots \geq \lambda_{n,p} \).

Proposition (Baik and Silverstein - 2006)

Let \( s_i = n_1 + \cdots + n_i \) for \( 1 \leq i \leq K \). Then

- For each \( k \in \{1, \ldots, K\} \) and \( s_{k-1} < j \leq s_k \) almost surely,
  \[
  \lambda_{n,j} \to \psi(\alpha_k) = \alpha_k + \frac{c\alpha_k}{\alpha_k - 1};
  \]

- For all \( 1 \leq i \leq L \) with a prefixed range \( L \) almost surely,
  \[
  \lambda_{n,q_0+i} \to b = (1 + \sqrt{c})^2.
  \]

Note. This result has been extended for more general spikes by Bai & Y., Benaych-Georges & Nadakuditi.
b) Estimator of $q_0$ (number of spikes)

Based on these results, we observe that when all the spikes are simple, i.e. $n_j \equiv 1$, the spacings

$$\delta_{n,j} = \lambda_{n,j} - \lambda_{n,j+1} \rightarrow \begin{cases} r > 0 & \forall j \leq q_0 \\ 0 & \forall j > q_0 \end{cases}$$

it is possible to detect $q_0$ from index-number $j$ where $\delta_{n,j}$ becomes small (case of simple spikes). Our estimator is defined by

$$\hat{q}_n = \min\{j \in \{1, \ldots, s\} : \delta_{n,j+1} < d_n\}, \quad (1)$$

where $(d_n)_n$ is a sequence to be defined and $s > q_0$ is a fixed number.
Consistency of $\hat{q}_n$: case of simple spikes

Assume

- All spikes are different (simple spike case);
- $\sigma^2 = 1$ (if not, take $\delta_{n,j}/\sigma^2$);

and

Entries have sub-Gaussian tails: for some positive $D, D'$ we have for all $t \geq D'$,

$$\mathbb{P}(|x_{ij}| \geq t^D) \leq e^{-t}.$$ 

Theorem [Passemer & Y. 2011]

Under Assumptions (1)-(5) and in the simple spikes case, if $d_n \to 0$ such that $n^{2/3}d_n \to +\infty$ then

$$\mathbb{P}(\hat{q}_n = q_0) \to 1.$$
Proof (idea)

\[ \Pr(\hat{q}_n = q_0) = 1 - \Pr \left( \bigcup_{1 \leq j \leq q_0} \{ \delta_{n,j} < d_n \} \cup \{ \delta_{n,q_0+1} \geq d_n \} \right) \]

\[ \geq 1 - \sum_{j=1}^{q_0} \Pr(\delta_{n,j} < d_n) - \Pr(\delta_{n,q_0+1} \geq d_n). \]

The terms in the sum converge to zero as \( d_n \to 0 \) and \( \delta_{n,j} \to r > 0 \). For the last term

\[ 1 - (*) = \Pr \left( n^{2/3}(\lambda_{n,q_0+1} - \lambda_{n,q_0+2}) \leq n^{2/3} d_n \right) \]

\[ \geq \Pr \left( \left\{ |Y_{n,1}| \leq n^{2/3} \frac{d_n}{2\beta} \right\} \cap \left\{ |Y_{n,2}| \leq n^{2/3} \frac{d_n}{2\beta} \right\} \right) \]

where \( Y \) is a tight sequence by the next proposition, and \( n^{2/3} d_n / 2\beta \to +\infty \), so \( 1 - (*) \to 1 \).
Proof (an additional important ingredient)

An (partial) extension of Tracy-Widom law in presence of spikes:

**Theorem (Benaych-Georges, Guionnet, Maida - 2010)**

*Under the above assumptions, for all $1 \leq i \leq L$ with a prefixed range $L$*

$$Y_{n,i} = \frac{n^{\frac{2}{3}}}{\beta} (\lambda_{n, q_0 + i} - b) = O_P(1)$$

*where $\beta = (1 + \sqrt{c})(1 + \sqrt{c^{-1}})^{\frac{1}{3}}$.***
Case of multiple spikes

- spacings $\delta_{n,j} \to 0$ from a same spike can also tend to 0;

- Confusion may be possible between these spacings and those from the bulk eigenvalues;

- Hopefully, fluctuations of both type of spacings have different rates:

$$n^{-1/2} \text{ v.s. } \sim n^{-2/3}.$$ 

Theorem (Bai and Y. (2008))

Under Assumptions (1)-(4) (2), the $n_k$-dimensional real vector

$$\sqrt{n}\{\lambda_{n,j} - \phi(\alpha_k), j \in \{s_{k-1} + 1, \ldots, s_k\}\}$$

converges weakly to the distribution of the $n_k$ eigenvalues of a Gaussian random matrix whose covariance depend of $\alpha_k$ and $c$.

[ related works are from Baik-Ben-Arous-Pêché, Paul ]
Consistency of $\hat{q}_n$: case of multiple spikes

The previous theorem of Bai and Y. implies:

- If $\alpha_j = \alpha_{j+1}$, convergence in $O_P(n^{-1/2})$;
- For unit eigenvalues, faster convergence in $O_P(n^{-2/3})$.

This allows us to use the same estimator provided we use a new threshold $d_n$.

Theorem (Passemier & Y. (2011))

*Under the above assumptions, if*

\[
d_n = o(n^{-1/2}), \quad \text{and} \quad n^{2/3} d_n \to +\infty,
\]

*then*

\[
P(\hat{q}_n = q_0) \to 1.
\]
Simulation experiments

We decided to use another version of our estimator which performs better

\[ \hat{q}_n^* = \min \{ j \in \{1, \ldots, s\} : \delta_{n,j+1} < d_n \text{ and } \delta_{n,j+2} < d_n \} \]

Threshold sequence: \[ d_n = Cn^{-2/3} \sqrt{2 \log \log n} \], where \( C \) is a constant to be adjusted for each case (Idea: law of the iterated logarithm for \( \lambda_{n,j}, j \leq q_0 \)).
Simulation experiments

- **Performance measure:** empirical false detection rates over 500 independent replications
  \[ P(\tilde{q}_n \neq q_0) \]

- **Simulation design:**
  - \( q_0 \): number of spikes;
  - \((\alpha_i)_{1 \leq i \leq q_0}\): spikes;
  - \( p \): dimension of the vectors;
  - \( n \): sample size;
  - \( c = p/n \);
  - \( \sigma^2 = 1 \) given or to be estimated;
  - \( C \): constant in \( d_n \).
<table>
<thead>
<tr>
<th>Fig. No.</th>
<th>Factors</th>
<th>Mod. No.</th>
<th>Factor values</th>
<th>Fixed parameters</th>
<th>Var. ( p, n )</th>
<th>( c )</th>
<th>( \sigma^2 )</th>
<th>( C )</th>
<th>( \text{par.} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Different</td>
<td>(( \alpha ))</td>
<td>(200, 800)</td>
<td>1/4</td>
<td>Given</td>
<td>5.5</td>
<td>9</td>
<td>( \alpha )</td>
<td></td>
</tr>
<tr>
<td>2L</td>
<td>Different</td>
<td>A (6, 5)</td>
<td>(2000, 500)</td>
<td>4</td>
<td>10</td>
<td>Given</td>
<td>11</td>
<td>( n )</td>
<td>Estimated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (10, 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (10, 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2R</td>
<td>Different</td>
<td>C (1.5)</td>
<td></td>
<td></td>
<td>1</td>
<td>Given</td>
<td>5</td>
<td>( n )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D (2.5, 1.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Possibly equal</td>
<td>E (( \alpha, \alpha, 5 ))</td>
<td>(200, 800)</td>
<td>1/4</td>
<td>Given</td>
<td>6</td>
<td>9.9</td>
<td>( \alpha )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F (( \alpha, \alpha, 15 ))</td>
<td>(2000, 500)</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4L</td>
<td>Possibly equal</td>
<td>G (6, 5, 5)</td>
<td></td>
<td>10</td>
<td>Given</td>
<td>9.9</td>
<td>( n )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H (10, 5, 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H (10, 5, 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4R</td>
<td>Possibly equal</td>
<td>I (1.5, 1.5)</td>
<td></td>
<td>1</td>
<td>Given</td>
<td>5</td>
<td>( n )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J (2.5, 1.5, 1.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Models A and D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Models G and J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>No factor</td>
<td>K</td>
<td>No factor</td>
<td>1</td>
<td>10</td>
<td>Given</td>
<td>8</td>
<td>( n )</td>
<td></td>
</tr>
<tr>
<td>8L</td>
<td>Models A and G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8R</td>
<td>Models B and H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9L</td>
<td>Models C and I, with ( C ) automatically chosen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9R</td>
<td>Models D and J, with ( C ) automatically chosen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 1. Misestimation rates as a function of factor strength for $(p, n) = (200, 800)$ and $(p, n) = (2000, 500)$. 
FIGURE 2. Misestimation rates as a function of $n$ for Models A, B (left) and Model C, D (right).
c) Discussions
- Comparison with an estimator by Kritchman and Nadler

In the non-spikes case \((q_0 = 0)\), \(nS_n \sim W_p(I, n)\). In this case

**Proposition (Johnstone - 2001)**

\[ P \left( \lambda_{n,1} < \sigma^2 \frac{\beta_{n,p}}{n^{2/3}} s + b \right) \to F_1(s) \]

where \(F_1\) is the Tracy-Widom distribution of order 1 and
\[ \beta_{n,p} = (1 + \sqrt{p/n})(1 + \sqrt{n/p})^{\frac{1}{3}}. \]

To distinguish a spike eigenvalue \(\lambda_{n,k}\) from a non-spike one at an asymptotic significance level \(\gamma\), their idea is to check whether

\[ \lambda_{n,k} > \sigma^2 \left( \frac{\beta_{n,p-k}}{n^{2/3}} s(\gamma) + b \right) \]

where \(s(\gamma)\) verifies \(F_1(s(\gamma)) = 1 - \gamma\). Their estimator is

\[ \tilde{q}_n = \arg\min_k \left( \lambda_{n,k} < \tilde{\sigma}^2 \left( \frac{\beta_{n,p-k}}{n^{2/3}} s(\gamma) + b \right) \right) - 1. \]
Figure 5. Misestimation rates as a function of $n$ for Model A (left) and Model D (right).
c) Discussions
- on the tuning parameter $C$

- $C$ has been tuned manually in each case;
- For real applications, need a procedure to choose this constant;
- Idea: use Wishart distributions as a benchmark to calibrate $C$;
- consider the gap between two largest eigenvalues: $\tilde{\lambda}_1 - \tilde{\lambda}_2$
By simulation to get empirical distribution of $\tilde{\lambda}_1 - \tilde{\lambda}_2$;  
500 independent replications.

- compute the upper 5% quantile $s$:

$$P(\tilde{\lambda}_1 - \tilde{\lambda}_2 \leq s) \approx 0.95.$$

- Define a value

$$\tilde{C} = sn^{2/3} / \sqrt{2 \times \log \log(n)}.$$ 

Results:

<table>
<thead>
<tr>
<th>Table 4. Approximation of the threshold $s$ such that $P(\tilde{\lambda}_1 - \tilde{\lambda}_2 \leq s) = 0.98$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p,n)</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Value of $s$</td>
</tr>
</tbody>
</table>
Assessment of the automated value $\tilde{C}$ with $c = 10$

- $\tilde{C} >$ tuned $C$ slightly ;
- Using $\tilde{C}$ $\rightarrow$ only a small drop of performance ;
- higher error rates in the case of equal factors for moderate sample sizes.

Figure 8. Misestimation rates as a function of $n$ for Models A, G (left) and Models B, H (right).
Application to S&P stocks data

- Estimated number of factors: $\hat{q}_0 = 17$;
- Residual variance: $\hat{\sigma}^2 = 0.3616$. 
3) Inference of the bulk spectrum

Estimation of population spectral distribution

**Population**
- $X$, mean-zero, $p$-dim
- $\text{Cov}(X) = \Sigma_p$

**Sample**
- $x_1, \ldots, x_n$, i.i.d, size $n$
- $S_n = \sum_{i=1}^n x_i x_i^*/n$

Large dimensional situations

$$\lim_{n \to \infty} \frac{p}{n} = c > 0$$

**PSD $H_p$**
- the empirical spectral distribution of $\Sigma_p$

**ESD $F_n$**
- the empirical spectral distribution of $S_n$.

**Problem:** Estimate $H_p$ from $F_n$. 
The Marčenko-Pastur equation

- Suppose that
  \[ \frac{p}{n} \to c > 0, \quad H_p \xrightarrow{w} H, \]
then under suitable conditions, cf. Marčenko-Pastur '68, Silverstein '95,
  \[ F_n \xrightarrow{w} F, \quad n \to \infty. \]

- Let \( s(z) = -(1 - c)/z + c \int \frac{1}{x - z} dF(x), \)
be the Stieltjes transform of (the companion distribution of) \( F, \)
then
  \[ z = -\frac{1}{s(z)} + c \int \frac{t}{1 + ts(z)} dH(t), \quad z \in \mathbb{C}^+, \]
which is called Marčenko-Pastur (MP) equation.

- This gives the inverse map of \( s(z) \) on \( \mathbb{C} \setminus \mathbb{R}. \)

  Almost all statistical tools for inference of \( H \) are based on this equation!!
a). Existing methods for estimation of PSD $H$

- Inversion of the MP equation:
  1. [El Karoui (2008)], nonparametric, complex field;
  2. [Li et al. (2012)], parametric, real field.

- Methods based on moments of $F$:
  1. [Rao et al. (2008)], quasi-likelihood;
  2. [Bai et al. (2010)], complete moment method.

- Methods based on moments and contour-integrals:
  1. [Mestre (2008)], eigenvalue splitting condition;
  2. [Yao et al. (2012)], global moment of $H$;
  3. [Li and Yao (2012)], local moment of $H$. 
Still needs new methods!

However,

- global inversion methods in [El Karoui (2008)] and [Li et al. (2012)] have some implementation issues that are non trivial to overcome;

- other methods are based on moments, but there are situations where these moments can not help to identify model parameters.

Example of a PSD $H$ not identifiable by moments

- $H$ has an inverse cubic density function ([Bouchaud and Potters (2009)])

$$h(t|\alpha) = \frac{b}{(t - a)^3}, \quad t \geq \alpha,$$

where the parameter is $0 \leq \alpha < 1$ is the parameter to be estimated and $a = 2\alpha - 1, \quad b = 2(1 - \alpha)^2$.

- Then

$$\int_{\alpha} xh(x)dx \equiv 1, \quad \int_{\alpha} x^k h(x)dx = \infty, \quad \text{for} \quad k \geq 2.$$

Moments of $H$ are independent from the parameter $\alpha$!
### b). A generalized expectation based method

**Main idea**

- Use of general test functions $f$ instead of monomials $x^k$ (moments);
- These test functions are usually smaller than the monomials $x^k$ so that $T(f) = \int f(x) dH(x)$ are finite.

In the example above of inverse cubic density, $f(x) = \sin(x)$ has a finite integral:

$$T(f) = b \int_{\alpha}^{\infty} \frac{\sin(x)}{(x - a)^3} dx .$$
Generalized expectations and their estimates

Let $f$ be a analytic function on an open $U \supset S_F$, support of $F$;

Define a generalized expectation $T(f) := \int f(t) dH(t)$;

It will be shown that

$$T(f) = K(c, f) + \frac{1}{2\pi i c} \oint_C z s'(z) f(-1/s(z)) dz,$$

where $K(c, f)$ is a constant, independent from $H$ and $C$ is a contour enclosing $S_F$.

With sample eigenvalues, $s(z)$ has an empirical estimate

$$s_n(z) = -(1 - p/n)/z + (p/n) \int 1/(x - z) dF_n(x),$$

Therefore, the above generalized expectation can be estimated by

$$\hat{T}(f) = K(p/n, f) + \frac{n}{p} \frac{1}{2\pi i} \oint_C z s'_n(z) f(-1/s_n(z)) dz. \quad (1)$$
Generalized expectation based estimator of $H$

- Suppose that $H$ belongs to a parametric family:
  \[ \mathcal{H} = \{ H_\theta : \theta \in \Theta \subset \mathbb{R}^q \} . \]

- Construct a $q$-dim vector of generalized expectations,
  \[ \gamma = (T(f_j))_{1 \leq j \leq q} = \left( \int f_j dH_\theta \right) ; \]
  such that $g : \theta \mapsto \gamma$ is an one-to-one map on $\Theta$;

- The \textit{generalized expectation estimator} (GEE) of $\theta$ is defined to be
  \[ \hat{\theta}_n = g^{-1}(\hat{\gamma}_n) , \]
  where $\hat{\gamma}_n = (\hat{T}(f_j))_{1 \leq j \leq L_i}$ with elements defined by (1).
c). Asymptotic properties of the GEE estimator

Assumptions:

Assumption (a). \( n, p \to \infty \) with \( p/n \to c \in (0, \infty) \).

Assumption (b). The sample covariance takes form

\[
S_n = \Sigma_p^{1/2} W_n W_n^* \Sigma_p^{1/2}/n,
\]

where the entries of \( W_n(p \times n) \) are i.i.d. standard real or complex normal variables, and \( \Sigma_p^{1/2} \) stands for any Hermitian square root of \( \Sigma_p \).

Assumption (c). \( H_p \overset{w}{\to} H \), a proper probability distribution on \([0, \infty)\). Moreover, the sequence of spectral norms \( (\|\Sigma_p\|) \) is bounded.
Asymptotics of \( \{ \hat{T}(f_j) \} \)'s

**Theorem (Li and Y. (2012))**

Under the assumptions (a)-(c), for each \( j = 1, \ldots, q \),

1. the generalized expectation \( T(f_j) \) can be expressed as

\[
T(f_j) = K(c, f_j) + \frac{1}{2\pi i c} \oint_C z s'(z) f_j(-1/s(z)) dz,
\]

where the constant \( K(c, f_j) = (1 - 1/c) f_j(0) \) if \( C \) encloses 0, and zero otherwise;

2. its empirical counterpart \( \hat{T}(f_j) \) based on \( s_n(z) \) converges almost surely to \( T(f_j) \);

3. if in addition, the entries of \( W_n \) \( (p \times n) \) are complex normal, the random vector

\[
n \left[ \hat{T}(f_j) - H_p(f_j) \right]_{1 \leq j \leq q} \overset{D}{\longrightarrow} N_q(0, \Phi),
\]

where the centralization term \( H_p(f_j) \) stands for the expectation of \( f_j \) with respect to \( H_p \), where the asymptotic covariances \( \Phi = (\phi_{ij})_{q \times q} \) are

\[
\phi_{ij} = \frac{-1}{4\pi^2 c^2} \oint_C \oint_{C'} f_i(-1/s(z_1)) f_j(-1/s(z_2)) k(z_1, z_2) dz_1 dz_2,
\]

where \( k(z_1, z_2) = s'(z_1)s'(z_2)/(s(z_1) - s(z_2))^2 - 1/(z_1 - z_2)^2 \).
Asymptotics of the GEE estimator $\hat{\theta}_n$

**Theorem (Li and Y. (2012))**

In addition to the assumptions (a)-(c), suppose that the true value of the parameter $\theta_0$ is an inner point of $\Theta$. Also, suppose that the function $g(\theta)$ is differentiable in a neighborhood of $\theta_0$ and the Jacobian matrix $J(\theta) = \partial g / \partial \theta$ is invertible at $\theta_0$. Then,

1. the GEE $\hat{\theta}_n$ is strongly consistent, i.e.
   $$\hat{\theta}_n \to \theta_0, \quad \text{a.s.},$$

2. moreover, if in addition, the entries of $W_n (p \times n)$ are complex normal, then
   $$n(\hat{\theta}_n - g^{-1}(\gamma_p)) \xrightarrow{D} N_q(0, \Gamma(\theta_0)),$$
   where $\gamma_p = (H_p(f_j))_{1 \leq j \leq q}$, and $\Gamma(\theta_0) = J^{-1}(\theta_0)\Phi(\theta_0)(J^{-1}(\theta_0))'$ with $\Phi$ being defined in Theorem 1.
d). Application: PSD of S&P 500 stocks covariances

Data analysis:

- Removed the 6 largest eigenvalues (deemed as spike eigenvalues);
- Assume an inverse cubic density for PSD $H$ associated to the 482 bulk eigenvalues, that is,

$$
h(t|\alpha) = \frac{b}{(t - a)^3}, \quad t \geq \alpha,
$$

where $0 < \alpha < 1$, $b = 2(1 - \alpha)^2$ and $a = 2\alpha - 1$;
- Moments-based methods fail, LEE may work!
Consider

\[ f(z) = \sin(z), \quad T(f, \alpha) = \int \sin(t)h(t|\alpha)dt; \]

\[ T(f, \alpha) \] is increasing with respect to \( \alpha \),

\[ \frac{\partial T(f, \alpha)}{\partial \alpha} \]

Figure: Curves of \( T(f, \alpha) \) (left) and \( \frac{\partial T(f, \alpha)}{\partial \alpha} \) (right).
Results on S&P 500 stocks data

- GEE: $\hat{T}(f, \alpha) = 0.5546, \hat{\alpha} = 0.3205$;
- LSE: $\hat{\alpha}' = 0.4384$ (see [Li et al. (2012)]);
- Denote by $f_\alpha$ the density function of LSD $F$ with respect to $H(\alpha)$. Compute a kernel density estimate $\hat{f}_{\text{ker}}$ from the 482 bulk eigenvalues (Gaussian kernel, bandwidth $h = 0.01$).
- Consider $d(\alpha) = L^2(f_\alpha, \hat{f}_{\text{ker}})$, then $d(\hat{\alpha}) = 0.2047$, $d(\hat{\alpha}') = 0.2863$.

Figure: $\hat{f}_{\text{ker}}$ (plain black), $f_\alpha$ (left, blue), and $f_{\alpha}'$ (right, blue).

- GEE yields a significantly better fit to the density of bulk eigenvalues.
Thank you!


