Empirical Likelihood Tests for High-dimensional Data

Ruodu Wang

Department of Statistics and Actuarial Science
University of Waterloo, Canada

ICSA - Canada Chapter 2013 Symposium
Toronto, August 2 - 3, 2013

Based on joint work with Liang Peng and Rongmao Zhang
Contents

1 Introduction

2 Testing Covariance Matrices

3 Testing Bandedness

4 Power Analysis

5 Simulation

Ruodu Wang Empirical Likelihood Tests for HD Data
In this talk we discuss empirical likelihood ratio tests for high-dimensional data.

Let $X_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq p}$. Here $p$ may depend on $n$.

- If $p$ is fixed, then it is a traditional statistical setting.
- If $p \to \infty$, then it is high-dimensional setting.
Typical testing questions:

- Testing $\mu = \mu_0$ (one sample mean test).
- Two sample means testing.
- Testing $\Sigma = \Sigma_0$ (covariance matrix test)
- Two sample covariance matrices testing.

We focus on testing covariance matrices.
Problem setup. Let $X_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq p}$. 
Problem setup. Let $X_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq p}$.

- Testing covariance matrix

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0.$$  \hspace{1cm} (1)
Problem setup. Let $X_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq p}$.

- Testing covariance matrix

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0. \quad (1)$$

- Testing bandedness

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau. \quad (2)$$
Problem setup. Let $X_i = (X_{i1}, \ldots, X_{ip}), \ i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq p}$.

- Testing covariance matrix

\[
H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0. \tag{1}
\]

- Testing bandedness

\[
H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau. \tag{2}
\]

Non-parametric. No information about sparsity.


*Literature.*

- Testing (1) for fixed $p$: traditional likelihood ratio test; scaled distance measure test (John (1971, 1972) and Nagao (1973)).
Literature.

- Testing (1) for fixed $p$: traditional likelihood ratio test; scaled distance measure test (John (1971, 1972) and Nagao (1973)).
  - Specific models are imposed.
  - Restrictions are put on $p$.
  - $p$ has to go to infinity as $n$ approaches infinity.
Testing (2) for divergent $p$: Cai and Jiang (2011).

- The test statistic: the coherence converges slowly.
- Normality are assumed.
- Restrictions are put on $p$ and $\tau$. 
Testing (2) for divergent $p$: Cai and Jiang (2011).
- The test statistic: the coherence converges slowly.
- Normality are assumed.
- Restrictions are put on $p$ and $\tau$.

Testing (2) for divergent $p$: Qiu and Chen (2012).
- Similar to Chen, Zhang and Zhong (2010), specific models; restrictions.
Our goal: build up a test statistic that works for both (1) and (2); loose the condition on \( p \); get rid of specific models.
Our goal: build up a test statistic that works for both (1) and (2); loose the condition on $p$; get rid of specific models.

First we assume $\mu$ is known. The case when $\mu$ is unknown is very similar.
Basic observations.

- $\Sigma = \Sigma_0$ is equivalent to

$$D^2 := ||\Sigma - \Sigma_0||_F^2 = \text{tr}((\Sigma - \Sigma_0)^2) = 0.$$
Basic observations.

- $\Sigma = \Sigma_0$ is equivalent to

$$D^2 := ||\Sigma - \Sigma_0||_F^2 = tr((\Sigma - \Sigma_0)^2) = 0.$$ 

- We can construct our test based on an estimator of $D^2$. 

Ruodu Wang

Empirical Likelihood Tests for HD Data
A natural estimator.

- For $i = 1, \ldots, n$, define the $p \times p$ matrix

$$Y_i = (X_i - \mu)(X_i - \mu)^T,$$

and estimator

$$e(\Gamma) = \text{tr}((Y_1 - \Gamma)(Y_2 - \Gamma)).$$

- $\mathbb{E}[Y_1] = \Sigma$ and $\mathbb{E}[e(\Sigma_0)] = D^2$. $\mathbb{E}[e(\Sigma_0)] = 0$ is equivalent to $\Sigma = \Sigma_0$. 
We need independent copies of \((Y_1, Y_2)\).
We need independent copies of \((Y_1, Y_2)\).

**Splitting the sample.**

Let \(N = [n/2]\). For \(i = 1, 2, \ldots, N\), we define

\[
e_i(\Sigma) = \text{tr}\left((Y_i - \Sigma)(Y_{N+i} - \Sigma)\right).
\]
We need independent copies of \((Y_1, Y_2)\).

**Splitting the sample.**

Let \(N = [n/2]\). For \(i = 1, 2, \ldots, N\), we define

\[
e_i(\Sigma) = \text{tr}((Y_i - \Sigma)(Y_{N+i} - \Sigma)).
\]

- Very difficult to estimate the variance of \(e_i\).
We need independent copies of \((Y_1, Y_2)\).

**Splitting the sample.**

Let \(N = \lfloor n/2 \rfloor\). For \(i = 1, 2, \ldots, N\), we define

\[
e_i(\Sigma) = \text{tr}((Y_i - \Sigma)(Y_{N+i} - \Sigma)).
\]

- Very difficult to estimate the variance of \(e_i\).
- Empirical likelihood method automatically catches the asymptotic variance.
Define the empirical likelihood ratio function with constraint (estimating equation) $E[e_1(\Sigma)] = 0$:

$$L_0(\Sigma) = \sup\left\{ \prod_{i=1}^{N} (Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i e_i(\Sigma) = 0, p_i \geq 0 \right\}.$$
Define the empirical likelihood ratio function with constraint (estimating equation) $E[e_1(\Sigma)] = 0$:

$$L_0(\Sigma) = \sup \{ \prod_{i=1}^{N} (Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i e_i(\Sigma) = 0, p_i \geq 0 \}.$$ 

Under some regularity conditions and $H_0$, $-2 \log L_0(\Sigma_0)$ converges weakly to $\chi^2_1$. This seems good but....
Define the empirical likelihood ratio function with constraint (estimating equation) $\mathbb{E}[e_1(\Sigma)] = 0$:

$$L_0(\Sigma) = \sup\left\{ \prod_{i=1}^{N} (Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i e_i(\Sigma) = 0, p_i \geq 0 \right\}.$$ 

Under some regularity conditions and $H_0$, $-2 \log L_0(\Sigma_0)$ converges weakly to $\chi^2_1$. This seems good but....

**Shortfall of the test based on $L_0$.**

When $\|\Sigma - \Sigma_0\|_F^2$ is small, $\mathbb{E}[e_1(\Sigma_0)]$ will be very close to 0 (in a rate of $\|\Sigma - \Sigma_0\|_F^2$). In this case the test based on $L_0$ has a poor power. Later we will see this in power analysis.
Secondary constraint.

- We add one more constraint which is easier to break under $H_1$. 

\[ v_i(\Sigma) = 1^T p (Y_i - Y_N + i - 2\Sigma 1) \] can be used in a constraint \( E[v_1(\Sigma_0)] = 0. \)
Secondary constraint.

- We add one more constraint which is easier to break under $H_1$.
- The choice of the second linear constraint can be arbitrary.
Secondary constraint.

- We add one more constraint which is easier to break under $H_1$.
- The choice of the second linear constraint can be arbitrary.
- With no prior information, the following statistics $\nu_i(\Sigma)$:

$$
\nu_i(\Sigma) = \mathbf{1}_p^T (Y_i - Y_{N+i} - 2\Sigma) \mathbf{1}_p
$$

can be used in a constraint $\mathbb{E}[\nu_1(\Sigma_0)] = 0$. 

Ruodu Wang
Empirical Likelihood Tests for HD Data
We define the empirical likelihood function with two constraints as

$$L_1(\Sigma_0) = \sup \left\{ \prod_{i=1}^{N} (Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i \begin{pmatrix} e_i(\Sigma_0) \\ v_i(\Sigma_0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, p_i \geq 0 \right\}.$$
We define the empirical likelihood function with two constraints as

\[ L_1(\Sigma_0) = \sup \left\{ \prod_{i=1}^{N} (Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i \begin{pmatrix} e_i(\Sigma_0) \\ v_i(\Sigma_0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, p_i \geq 0 \right\}. \]

**Theorem 1**

Suppose \( e_1(\Sigma_0) \) and \( v_1(\Sigma_0) \) satisfy a regularity condition (P). Then under \( H_0 \), \( -2 \log L_1(\Sigma_0) \) converges in distribution to \( \chi_2^2 \) as \( n \to \infty \).
CLT condition (similar to the Lyapunov condition).

(P) We say a statistic $T$ with size $n$ satisfies condition (P) if 
\[ \mathbb{E} T^2 > 0 \] and for some $\delta > 0$,
\[
\frac{\mathbb{E}|T|^{2+\delta}}{(\mathbb{E} T^2)^{1+\delta/2}} = o(n^{\frac{\delta+\min(\delta,2)}{4}}).
\]

For example, if \( \mathbb{E}(T^4)/(\mathbb{E}(T^2))^2 = o(n) \), then $T$ satisfies (P) with $\delta = 2$. 
Remark 1

In order to prove Theorem 1, it is sufficient to prove condition (P) guarantees that the sample $t_i = \left( \frac{e_i(\Sigma_0)}{\sqrt{\text{Var}(e_i(\Sigma_0))}}, \frac{v_i(\Sigma_0)}{\sqrt{\text{Var}(v_i(\Sigma_0))}} \right)^T$ satisfies CLT and $t_i^2$ satisfies LLN, with a controlled maximum.
Remark 1

In order to prove Theorem 1, it is sufficient to prove condition (P) guarantees that the sample $t_i = \left( \frac{e_i(\Sigma_0)}{\sqrt{\text{Var}(e_i(\Sigma_0))}}, \frac{v_i(\Sigma_0)}{\sqrt{\text{Var}(v_i(\Sigma_0))}} \right)^T$ satisfies CLT and $t_i^2$ satisfies LLN, with a controlled maximum.

Remark 2

When $\mu$ is unknown, just replace $\mu$ in $Y_i$ by the sample means and the theorem still holds with one extra moment condition.
Remark 3

In the factor model considered by Chen, Zhang and Zhong (2010), $e_1(\Sigma_0)$ and $v_1(\Sigma_0)$ satisfy (P). With this model, our test allows $p$ to diverge arbitrarily fast or stay finite.
The problem is testing

\[ H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau. \]  

(3)

Here we consider \( \mu \) is known.
We are interested in the information of the black squares in $\Sigma$ and we will ignore the stars.

$$
\begin{pmatrix}
  \ast & \ast & \blacksquare & \blacksquare & \blacksquare \\
  \ast & \ast & \ast & \blacksquare & \blacksquare \\
  \ast & \ast & \ast & \ast & \blacksquare \\
  \ast & \ast & \ast & \ast & \ast \\
  \ast & \ast & \ast & \ast & \ast \\
\end{pmatrix}
$$

**Basic observation.**

$H_0$ is equivalent to the black squares of $\Sigma$ being 0.
Define the $\tau$-off-diagonal upper triangular matrix $M^{(\tau)}$ of a matrix $M$:

$$(M^{(\tau)})_{ij} = \begin{cases} 
M_{ij} & j \geq i + \tau; \\
0 & j < i + \tau.
\end{cases}$$

$H_0$ is equivalent to $\text{tr}((\Sigma^{(\tau)})^T \Sigma^{(\tau)}) = 0$. 
• For \( i = 1, \ldots, N \), Let

\[
e'_i = \text{tr} \left( (Y_i^{(\tau)})^T Y_{N+i}^{(\tau)} \right),
\]

\[
v'_i = 1_p^T (Y_i^{(\tau)} + Y_{N+i}^{(\tau)}) 1_p.
\]

• We define the empirical likelihood function as

\[
L_2 = \sup \left\{ \prod_{i=1}^N (Np_i) : \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \begin{pmatrix} e'_i \\ v'_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, p_i \geq 0 \right\}.
\]

Here we omit the \( \Sigma_0 \) in \( e'_i \) and \( v'_i \).
Theorem 2

Suppose that $e'_1$ and $v'_1$ satisfy (P). Then under $H_0$ in (3), $-2 \log L_2$ converges in distribution to $\chi^2_2$ as $n \to \infty$. 

Suppose that $e_1'$ and $v_1'$ satisfy (P). Then under $H_0$ in (3),
$-2 \log L_2$ converges in distribution to $\chi_2^2$ as $n \to \infty$.

- The method can be used to test some other structures. One interesting application is to test the assumption or estimation of the sparsity.
Remark 4

(1) In the Gaussian model used by Cai and Jiang (2011), $e_1'$ and $\nu_1'$ satisfy (P) provided that $\tau = o \left( \frac{\sum_{1 \leq i, j \leq p} \sigma_{ij}}{(\sum_{1 \leq i, j \leq p} |\sigma_{ij}|)^{1/2}} \right)$. 

(2) With moment or boundedness conditions, the Gaussian assumption can be removed.
Remark 4

(1) In the Gaussian model used by Cai and Jiang (2011), $e_1'$ and $v_1'$ satisfy $(P)$ provided that $\tau = o \left( \frac{\sum_{1 \leq i, j \leq p} \sigma_{ij}}{(\sum_{1 \leq i, j \leq p} |\sigma_{ij}|)^{1/2}} \right)$.

(2) With moment or boundedness conditions, the Gaussian assumption can be removed.
Power Analysis

Denote $\pi_{11} = E(e_1(\Sigma)^2)$, $\pi_{22} = E(v_1(\Sigma)^2)$,

$$\zeta_1 = \text{tr}((\Sigma - \Sigma_0)^2)/\sqrt{\pi_{11}}$$

and

$$\zeta_2 = 21_p^T(\Sigma - \Sigma_0)1_p/\sqrt{\pi_{22}}.$$

For most models we discuss,

$$\zeta_1 = O\left(\frac{\text{tr}((\Sigma - \Sigma_0)^2)}{\text{tr}(\Sigma^2)}\right)$$

and

$$\zeta_2 = O\left(\frac{1_p^T(\Sigma - \Sigma_0)1_p}{1_p^T \Sigma^2 1_p}\right).$$
In addition to the conditions of Theorem 1, if $H_1 : \Sigma \neq \Sigma_0$ holds, then

$$P\{-2 \log L_1(\Sigma_0) > \xi_{1-\alpha}\} = P\{\chi_{2,\nu}^2 > \xi_{1-\alpha}\} + o(1)$$

for any level $\alpha$ as $n \to \infty$, where $\chi_{2,\nu}^2$ is a noncentral chi-square distribution with two degrees of freedom and noncentrality parameter $\nu = N(\zeta_1^2 + \zeta_2^2)$,
Remark 5

From the above power analysis, the new test rejects the null hypothesis with probability tending to one when
\[
\max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \to \infty.
\]
Remark 5

- From the above power analysis, the new test rejects the null hypothesis with probability tending to one when \( \max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \to \infty \).

- Note that the test given in Chen, Zhang and Zhong (2010) requires \( n\zeta_1 \to \infty \).
Remark 5

- From the above power analysis, the new test rejects the null hypothesis with probability tending to one when
  \[ \max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \rightarrow \infty. \]

- Note that the test given in Chen, Zhang and Zhong (2010) requires \( n\zeta_1 \rightarrow \infty. \)

- Our test may have a better power or a worse power in different settings.
Remark 5

- From the above power analysis, the new test rejects the null hypothesis with probability tending to one when
  \[ \max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \to \infty. \]

- Note that the test given in Chen, Zhang and Zhong (2010) requires \( n\zeta_1 \to \infty \).

- Our test may have a better power or a worse power in different settings.

- Same results for the test in Theorem 2.
It is clear that our tests is powerful when $\Sigma - \Sigma_0$ is dense, and not powerful when $\Sigma - \Sigma_0$ is sparse.
It is clear that our tests is powerful when $\Sigma - \Sigma_0$ is dense, and not powerful when $\Sigma - \Sigma_0$ is sparse.

With only the first constraint $\mathbb{E}(e_1(\Sigma_0)) = 0$, the test power (requires $\sqrt{n}\zeta_1 \to \infty$) is worse than the test in Chen, Zhang and Zhong (2010).
It is clear that our tests is powerful when $\Sigma - \Sigma_0$ is dense, and not powerful when $\Sigma - \Sigma_0$ is sparse.

With only the first constraint $\mathbb{E}(e_1(\Sigma_0)) = 0$, the test power (requires $\sqrt{n\zeta_1} \to \infty$) is worse than the test in Chen, Zhang and Zhong (2010).

The test in Cai and Jiang (2011) is good when $\Sigma - \Sigma_0$ is sparse but is powerless when $\Sigma - \Sigma_0$ is dense, since their test power depends on $||\Sigma - \Sigma_0||_{\text{max}}$. 
Testing covariance matrices

- We assume a dense model and a local alternative.
- We compare with Chen, Zhang and Zhong (2010) for testing covariance matrices and Cai and Jiang (2011) for testing bandedness.
- The ELT has biased size for small $n$, so we also give a bootstrap calibrated version of ELT.
Table: Testing covariance matrices

<table>
<thead>
<tr>
<th>$(n, p)$</th>
<th>$EL(0.05)$</th>
<th>$BCEL(0.05)$</th>
<th>$CZZ(0.05)$</th>
<th>$EL(0.05)$</th>
<th>$BCEL(0.05)$</th>
<th>$CZZ(0.05)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\delta = 0$</td>
<td>$\delta = 0$</td>
<td>$\delta = 0$</td>
<td>$\delta = 1$</td>
<td>$\delta = 1$</td>
<td>$\delta = 1$</td>
</tr>
<tr>
<td>(50, 25)</td>
<td>0.127</td>
<td>0.054</td>
<td>0.053</td>
<td>0.296</td>
<td>0.118</td>
<td>0.219</td>
</tr>
<tr>
<td>(50, 50)</td>
<td>0.148</td>
<td>0.065</td>
<td>0.067</td>
<td>0.324</td>
<td>0.136</td>
<td>0.216</td>
</tr>
<tr>
<td>(50, 100)</td>
<td>0.138</td>
<td>0.068</td>
<td>0.038</td>
<td>0.317</td>
<td>0.125</td>
<td>0.212</td>
</tr>
<tr>
<td>(50, 200)</td>
<td>0.168</td>
<td>0.081</td>
<td>0.041</td>
<td>0.310</td>
<td>0.113</td>
<td>0.221</td>
</tr>
<tr>
<td>(50, 400)</td>
<td>0.151</td>
<td>0.071</td>
<td>0.045</td>
<td>0.342</td>
<td>0.145</td>
<td>0.242</td>
</tr>
<tr>
<td>(50, 800)</td>
<td>0.154</td>
<td>0.064</td>
<td>0.041</td>
<td>0.337</td>
<td>0.137</td>
<td>0.219</td>
</tr>
<tr>
<td>(200, 25)</td>
<td>0.065</td>
<td>0.048</td>
<td>0.052</td>
<td>0.348</td>
<td>0.305</td>
<td>0.179</td>
</tr>
<tr>
<td>(200, 50)</td>
<td>0.058</td>
<td>0.052</td>
<td>0.041</td>
<td>0.336</td>
<td>0.298</td>
<td>0.162</td>
</tr>
<tr>
<td>(200, 100)</td>
<td>0.068</td>
<td>0.054</td>
<td>0.059</td>
<td>0.353</td>
<td>0.319</td>
<td>0.179</td>
</tr>
<tr>
<td>(200, 200)</td>
<td>0.056</td>
<td>0.051</td>
<td>0.058</td>
<td>0.358</td>
<td>0.322</td>
<td>0.155</td>
</tr>
<tr>
<td>(200, 400)</td>
<td>0.069</td>
<td>0.064</td>
<td>0.051</td>
<td>0.374</td>
<td>0.343</td>
<td>0.180</td>
</tr>
<tr>
<td>(200, 800)</td>
<td>0.058</td>
<td>0.047</td>
<td>0.050</td>
<td>0.366</td>
<td>0.338</td>
<td>0.182</td>
</tr>
</tbody>
</table>
### Table: Testing bandedness

<table>
<thead>
<tr>
<th>(n, p)</th>
<th>EL(0.05)</th>
<th>BCEL(0.05)</th>
<th>CJ(0.05)</th>
<th>EL(0.05)</th>
<th>BCEL(0.05)</th>
<th>CJ(0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>δ = 0</td>
<td>δ = 0</td>
<td>δ = 0</td>
<td>δ = 1</td>
<td>δ = 1</td>
<td>δ = 1</td>
</tr>
<tr>
<td>(50, 25)</td>
<td>0.118</td>
<td>0.036</td>
<td>0.015</td>
<td>0.272</td>
<td>0.093</td>
<td>0.017</td>
</tr>
<tr>
<td>(50, 50)</td>
<td>0.124</td>
<td>0.049</td>
<td>0.010</td>
<td>0.266</td>
<td>0.097</td>
<td>0.018</td>
</tr>
<tr>
<td>(50, 100)</td>
<td>0.126</td>
<td>0.057</td>
<td>0.005</td>
<td>0.268</td>
<td>0.099</td>
<td>0.004</td>
</tr>
<tr>
<td>(50, 200)</td>
<td>0.128</td>
<td>0.058</td>
<td>0.003</td>
<td>0.268</td>
<td>0.100</td>
<td>0.001</td>
</tr>
<tr>
<td>(50, 400)</td>
<td>0.113</td>
<td>0.053</td>
<td>0.002</td>
<td>0.282</td>
<td>0.121</td>
<td>0.001</td>
</tr>
<tr>
<td>(50, 800)</td>
<td>0.128</td>
<td>0.062</td>
<td>0.001</td>
<td>0.281</td>
<td>0.109</td>
<td>0.000</td>
</tr>
<tr>
<td>(200, 25)</td>
<td>0.078</td>
<td>0.062</td>
<td>0.019</td>
<td>0.288</td>
<td>0.253</td>
<td>0.034</td>
</tr>
<tr>
<td>(200, 50)</td>
<td>0.074</td>
<td>0.059</td>
<td>0.033</td>
<td>0.323</td>
<td>0.286</td>
<td>0.020</td>
</tr>
<tr>
<td>(200, 100)</td>
<td>0.057</td>
<td>0.053</td>
<td>0.019</td>
<td>0.332</td>
<td>0.304</td>
<td>0.044</td>
</tr>
<tr>
<td>(200, 200)</td>
<td>0.066</td>
<td>0.046</td>
<td>0.024</td>
<td>0.293</td>
<td>0.263</td>
<td>0.032</td>
</tr>
<tr>
<td>(200, 400)</td>
<td>0.061</td>
<td>0.052</td>
<td>0.020</td>
<td>0.336</td>
<td>0.304</td>
<td>0.016</td>
</tr>
<tr>
<td>(200, 800)</td>
<td>0.053</td>
<td>0.046</td>
<td>0.026</td>
<td>0.317</td>
<td>0.297</td>
<td>0.025</td>
</tr>
</tbody>
</table>
Conclusion.

The new technique

- works for non-parametric models;
- allows arbitrary \( p \); requires only moment conditions;
- avoids to estimate asymptotic variance; the limiting distribution is always \( \chi^2_2 \);
- can be applied to testing sample mean, two-sample means, and two-sample covariance matrices under the HD framework.
Shortfalls:

- the number of observations is reduced by half;
- the power is good in the dense setting but not in the sparse setting.
- The optimal choice of the second constraint is unknown.


References II


Thank you!