Supplementary Material for “Empirical likelihood test for a large dimensional mean vector”

BY XIA CUI
School of Economics and Statistics, Guangzhou University, Guangzhou, China, 510006
cuixia@gzhu.edu.cn

RUNZE LI
Department of Statistics, The Pennsylvania State University, University Park, PA 16802
rzli@psu.edu

GUANGREN YANG
Department of Statistics, School of Economics, Jinan University, Guangzhou, China, 510632
tygr@jnu.edu.cn

AND WANG ZHOU
Department of Statistics and Applied Probability, National University of Singapore
Singapore, 117546
stazw@nus.edu.sg

S.1. A LEMMA

To prove Lemma 1, we need the following Lemma S.1, which is a direct result of Götze & Tikhomirov (2002). The proof of Lemma S.1 is similar to Bai & Sarandasa (1996) and is omitted.

LEMMA S.1. Suppose that \( x_i, i = 1, \ldots, n \) is a random sample from model (2.2) with \( E(z_j^6) < \infty, j = 1, \ldots, p \). Further assume that there exists a positive constant \( b_1 \) such that

\[
1 - tr(\Sigma) / tr(\Sigma^2) \geq b_1^2,
\]

Then it follows that

\[
\frac{n||\bar{x} - \mu||^2 - tr(\Sigma)}{2tr(\Sigma^2)} \rightarrow N(0, 1),
\]

in distribution.

S.2. PROOF OF LEMMA 2

Define

\[
H_1^R = \left\{ \frac{2tr(\Sigma^2)}{2tr(\Sigma^2)} \right\}^{1/2} - 1 \left[ \frac{2n^2}{(n+2)^2} W(\mu, k_n) - tr(\Sigma) \right],
\]

\[
H_2^R = \frac{\widehat{tr}(\Sigma) - tr(\Sigma)}{2tr(\Sigma^2)} \right\}^{1/2},
\]

\[
H_3^R = - \left[ \frac{1}{\{2tr(\Sigma^2)\}} - \frac{1}{\{2tr(\Sigma^2)\}} \right] \left\{ \widehat{tr}(\Sigma) - tr(\Sigma) \right\}.
\]

© 2018 Biometrika Trust
Similarly, we have 
\[
\begin{align*}
&\{2 \text{tr}(\Sigma^2)\}^{-1/2} \left\{ \frac{2n_0^2}{(n+2)^2} W(\mu_0, k_n) - \text{tr}(\Sigma) \right\}
\end{align*}
\]
\[
= \{2 \text{tr}(\Sigma^2)\}^{-1/2} \left\{ \frac{2n_0^2}{(n+2)^2} W(\mu_0, k_n) - \text{tr}(\Sigma) \right\} + \sum_{i=1}^{3} H_i^R.
\]

By Lemma 1, it suffices to show that \( H_i^R = o_p(1), i = 1, 2, 3 \), in order to show Lemma 2. According to Proposition A.2 of Chen et al. (2010), it follows that
\[
\begin{align*}
\text{tr}(\Sigma) - \text{tr}(\Sigma) &= O_p \left\{ \{n^{-1}\text{tr}(\Sigma^2)\}^{1/2} + \{n^{-1}\text{tr}(\Sigma o \Sigma)\}^{1/2} \right\}, \\
\text{tr}(\Sigma^2) - \text{tr}(\Sigma^2) &= O_p \left\{ \{n^{-2}\text{tr}(\Sigma^2)\}^{1/2} + \{n^{-1}\text{tr}(\Sigma^4)\}^{1/2} + \{n^{-1}\text{tr}(\Sigma^2 o \Sigma^2)\}^{1/2} \right\}.
\end{align*}
\]

By the assumption that all eigenvalues of \( \Sigma \) lie between two positive constants \( c_0 \) and \( C_0 \), \( \text{tr}(\Sigma \circ \Sigma) \leq \text{tr}(\Sigma^2) = O(p), \text{tr}(\Sigma^4) = O(p) \), \( \text{tr}(\Sigma^2) = O(p^2) \) and \( \text{tr}(\Sigma^2 o \Sigma^2) \leq \text{tr}(\Sigma^4) = O(p) \). Thus, under the condition that \( p_n/n = c_n \) for \( c \in [1, \infty) \), we have that \( \text{tr}(\Sigma) - \text{tr}(\Sigma) = O_p(1), \text{tr}(\Sigma^2) - \text{tr}(\Sigma^2) = O_p(1) \). Now we deal with \( H^R \):
\[
|H^R| \leq \left| \left( \frac{2\text{tr}(\Sigma^2)\{1/2}{2\text{tr}(\Sigma^2)}\{1/2} - 1 \right) \cdot \left( \frac{2n_0^2}{(n+2)^2} W(\mu_0, k_n) - \text{tr}(\Sigma) \right) \right|
\]
\[
\leq \{\text{tr}(\Sigma^2)\}^{-1/2} \cdot \left| \text{tr}(\Sigma^2)\right|^{1/2} \left\{ \{\text{tr}(\Sigma^2)\}^{1/2} + \{\text{tr}(\Sigma^2)\}^{1/2} \right\} \cdot \left| \text{tr}(\Sigma^2) - \text{tr}(\Sigma^2) \right| \cdot \left( \frac{2n_0^2}{(n+2)^2} W(\mu_0, k_n) - \text{tr}(\Sigma) \right)
\]
\[
= O_p(1/p)O_p(1) = o_p(1).
\]

Similarly, we have
\[
|H^R| \leq \frac{1}{\{2\text{tr}(\Sigma^2)\}^{1/2}} \left| \text{tr}(\Sigma) - \text{tr}(\Sigma) \right| = O_p(1/p)O_p(1) = o_p(1).
\]

For \( H^R_3 \), we have
\[
|H^R_3| \leq \left| \left( \frac{2\text{tr}(\Sigma^2)\{1/2}{2\text{tr}(\Sigma^2)}\{1/2} - 1 \right) \cdot \left( \frac{2n_0^2}{(n+2)^2} W(\mu_0, k_n) - \text{tr}(\Sigma) \right) \right|
\]
\[
\leq C \{\text{tr}(\Sigma^2)\}^{-1/2} \cdot \left| \text{tr}(\Sigma^2)\right|^{1/2} \left\{ \{\text{tr}(\Sigma^2)\}^{1/2} + \{\text{tr}(\Sigma^2)\}^{1/2} \right\} \cdot \left| \text{tr}(\Sigma^2) - \text{tr}(\Sigma^2) \right| \cdot \left( \frac{2n_0^2}{(n+2)^2} W(\mu_0, k_n) - \text{tr}(\Sigma) \right)
\]
\[
= O_p(p^{-3/2})O_p(1) = o_p(1).
\]

The proof of Lemma 2 is completed.

**S.3. Additional numerical results**

**S.3.1 Simulation results**

Figure S1 presents the plot of Type I error rate and local power when \( p/n = 1.2 \). The overall pattern of Figure S1 is similar to that of Figure 2 in the main text.

*Linear hypothesis.* We now examine the performance of \( T^F_n \) for the linear hypothesis in Section 2.3. We take \( \mu_0 = 0, \mu = \delta(2, 1, \cdots, 1)^T / \sqrt{n} \) with \( \delta = 0, 4, 7, 8 \) and 9, and \( F = (f_{ij}) \), a \( p \times p \) matrix with \( f_{jj} = 1, f_{jj+1} = -1 \) for \( j = 1, \cdots, q \), and all other elements being 0. This is equivalent to testing \( H_0 : \mu_1 = \mu_2 = \cdots = \mu_q \), where \( q = 1.1n \). We set \( l_n = n^{5/4} \log n, k_n = \)
Test for high-dimensional mean vector

Fig. S1. Empirical power functions of $T_n$, $T_{CQ}$ and $T_{WPL}$ with $p/n = 1.2$. Top, middle and bottom panels are for $z_j \sim N(0, 1)$, $\text{Gamma}(4, 2) - 2$ and $z_j \sim (3/5)^{1/2}t(5)$, respectively. The solid, dotted and dashed curves are the empirical power curves of $T_n$, $T_{CQ}$ and $T_{WPL}$, respectively.

\[
(q/\log q)^{1/2}, \gamma = (1, \ldots, 1)^T/\sqrt{p}\text{ in } T^F_n.\]

Figure S2 depicts the empirical powers of $T^F_n$ based on 1000 simulations. From Figure S2, we can see that the empirical rejection probabilities under $H_0$ are very close to 0.05 across all cases, when $\delta = 0$. This indicates that the limiting null distribution provides correct critical values. Also from Figure S2, the power functions increase rapidly and approach one as the value of $\delta$ increases.

S.3.2. Real data analysis

In this section we illustrate the proposed test procedure by an empirical analysis of stock data, which consist of nine sectors (consumer discretionary (CD), consumer staples (CS), en-
Fig. S2. Power functions for assessing the performance of $T^F_n$ under local alternatives $H_0$

Let $p_k$ denote the number of stocks contained in the sector $k$ with $k = 1, 2, \ldots, 9$. The detailed value of $p_k$ is listed in the second column of Table S1. The price records of the stocks in sector $k$ at time $t$ are denoted by $\{h_{ji}^{(k)}, j = 1, \ldots, p_k, i = 1, \ldots, 25\}$. The $i$-th, $i = 1, \ldots, 24, \log-$
Test for high-dimensional mean vector

Table S1. The performance of $T_n$, $T_{CQ}$ and $T_{WPL}$ for large dimension data in S&P 500.

<table>
<thead>
<tr>
<th>Sector</th>
<th>$p$</th>
<th>$T_n$</th>
<th>$T_{CQ}$</th>
<th>$T_{WPL}$</th>
<th>EL-Pvalue</th>
<th>CQ-Pvalue</th>
<th>WPL-Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>83</td>
<td>0.1954</td>
<td>0.2448</td>
<td>0.4054</td>
<td>0.8451</td>
<td>0.8066</td>
<td>0.6852</td>
</tr>
<tr>
<td>CS</td>
<td>38</td>
<td>2.2011</td>
<td>1.2698</td>
<td>1.1877</td>
<td>0.0277</td>
<td>0.2042</td>
<td>0.2349</td>
</tr>
<tr>
<td>Energy</td>
<td>40</td>
<td>0.7948</td>
<td>0.2830</td>
<td>0.9466</td>
<td>0.5625</td>
<td>0.6713</td>
<td>0.7978</td>
</tr>
<tr>
<td>Financials</td>
<td>80</td>
<td>-0.0715</td>
<td>-0.0455</td>
<td>-0.2561</td>
<td>0.9430</td>
<td>0.9637</td>
<td>0.9603</td>
</tr>
<tr>
<td>HC</td>
<td>46</td>
<td>-0.5791</td>
<td>-0.4244</td>
<td>-0.2561</td>
<td>0.5625</td>
<td>0.6713</td>
<td>0.7978</td>
</tr>
<tr>
<td>IND</td>
<td>60</td>
<td>-1.6169</td>
<td>-1.5866</td>
<td>-0.9996</td>
<td>0.1059</td>
<td>0.1126</td>
<td>0.3175</td>
</tr>
<tr>
<td>IT</td>
<td>63</td>
<td>2.6797</td>
<td>2.4471</td>
<td>1.9694</td>
<td>0.0074</td>
<td>0.0144</td>
<td>0.0489</td>
</tr>
<tr>
<td>Materials</td>
<td>28</td>
<td>0.2111</td>
<td>0.2870</td>
<td>-0.3202</td>
<td>0.8328</td>
<td>0.7741</td>
<td>0.7488</td>
</tr>
</tbody>
</table>

returns for stocks in sector $k$ is

$$\mathbf{x}_{i}^{(k)} = \left( \log \frac{h_{1,i+1}^{(k)}}{h_{1,i}^{(k)}}, \log \frac{h_{2,i+1}^{(k)}}{h_{2,i}^{(k)}}, \ldots, \log \frac{h_{p_k,i+1}^{(k)}}{h_{p_k,i}^{(k)}} \right)^T.$$  

Denote $\mu^{(k)} = E\mathbf{x}_{1}^{(k)}$. Of interest is to test

$$H_0^{(k)} : \mu^{(k)} = 0 \text{ versus } H_1^{(k)} : \mu^{(k)} \neq 0. \quad (S.1)$$

We calculate $T_n$ for data in each sector with $l = n^{5/4} \log n$, $k_n = (p_k / \log p_k)^{1/2}$ and $\alpha = (1, 0, \ldots, 0)^T$ in (2.7) for each $k$. As a comparison, we also apply $T_{CQ}$ and $T_{WPL}$ for data in each sector. Table S1 depicts the values of $T_n$, $T_{CQ}$ and $T_{WPL}$ and their corresponding $P$-values, EL-Pvalue, CQ-Pvalue and WPL-Pvalue, respectively. The number of companies in each sector ranges from 28 to 83 and is greater than the sample size 25.

Table S1 shows that the $P$-values of $T_n$, $T_{CQ}$ and $T_{WPL}$ for consumer staples sector are 0.0277, 0.2042 and 0.2349. It suggests that $T_n$ is in favor of rejecting the null hypothesis at level 0.05, while $T_{CQ}$ and $T_{WPL}$ fail to reject the null hypothesis. As to the sector of materials, $T_n$, $T_{CQ}$ and $T_{WPL}$ all reject the null hypothesis at level 0.05. In particular, the $p$-value of $T_n$ is 0.0074, which is smaller than the $p$-values of $T_{CQ}$ and $T_{WPL}$. As to the remaining seven sectors, all three test statistics fail to reject the null hypothesis. In particular, the data from the sectors of information technology (IT) and Financials show little change among three test statistics.

References

