Biometrika (2019), **103**, 1, *pp.* 1–5
Printed in Great Britain Advance Access publication on 31 July 2019

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 $x_i, i = 1, \cdots, n$ is a random sample from model (2.2) with $E(z_j^6) < \infty, j = 1, \cdots, p$. Further assume that there exists a positive constant b_1 such that $1 - tr(\Sigma \circ \Sigma)/tr(\Sigma^2) \geq b_1^2$. Then it follows that*

$$\frac{n\|\bar{x} - \mu\|^2 - tr(\Sigma)}{\{2tr(\Sigma^2)\}^{1/2}} \to N(0, 1),$$

in distribution.

S.2. PROOF OF LEMMA 2

Define

$$H_1^R = \Big[\frac{\{2\text{tr}(\Sigma^2)\}^{1/2}}{\{2\widehat{\text{tr}(\Sigma^2)}\}^{1/2}} - 1\Big] \frac{\frac{2nl_n^2}{(n+2)^2} W(\mu, k_n) - \text{tr}(\Sigma)}{\{2\text{tr}(\Sigma^2)\}^{1/2}},$$

$$H_2^R = -\frac{\widehat{\text{tr}(\Sigma)} - \text{tr}(\Sigma)}{\{2\text{tr}(\Sigma^2)\}^{1/2}},$$

$$H_3^R = -\Big[\frac{1}{\{2\widehat{\text{tr}(\Sigma^2)}\}^{1/2}} - \frac{1}{\{2\text{tr}(\Sigma^2)\}^{1/2}}\Big] \Big\{\widehat{\text{tr}(\Sigma)} - \text{tr}(\Sigma)\Big\}.$$

Then

$$
\begin{aligned}
&\{2\,\widehat{\mathrm{tr}(\Sigma^2)}\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}W(\mu_0,k_n)-\widehat{\mathrm{tr}(\Sigma)}\right\}\\
&=\{2\,\mathrm{tr}(\Sigma^2)\}^{-1/2}\left\{\frac{2nl_n^2}{(n+2)^2}W(\mu_0,k_n)-\mathrm{tr}(\Sigma)\right\}+\sum_{i=1}^{3}H_i^R.
\end{aligned}
$$

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

$$
\begin{aligned}
\widehat{\mathrm{tr}(\Sigma)}-\mathrm{tr}(\Sigma)&=O_p\Big[\{n^{-1}\mathrm{tr}(\Sigma^2)\}^{1/2}+\{n^{-1}\mathrm{tr}(\Sigma\circ\Sigma)\}^{1/2}\Big],\\
\widehat{\mathrm{tr}(\Sigma^2)}-\mathrm{tr}(\Sigma^2)&=O_p\Big[\{n^{-2}\mathrm{tr}^2(\Sigma^2)\}^{1/2}+\{n^{-1}\mathrm{tr}(\Sigma^4)\}^{1/2}+\{n^{-1}\mathrm{tr}(\Sigma^2\circ\Sigma^2)\}^{1/2}\Big].
\end{aligned}
$$

By the assumption that all eigenvalues of Σ lie between two positive constants c_0 and C_0, $\mathrm{tr}(\Sigma\circ\Sigma)\le\mathrm{tr}(\Sigma^2)=O(p)$, $\mathrm{tr}(\Sigma^4)=O(p)$, $\mathrm{tr}^2(\Sigma^2)=O(p^2)$ and $\mathrm{tr}(\Sigma^2\circ\Sigma^2)\le\mathrm{tr}(\Sigma^4)=O(p)$. Thus, under the condition that $p_n/n=c_n\to c\in[1,\infty)$, we have that $\widehat{\mathrm{tr}(\Sigma)}-\mathrm{tr}(\Sigma)=O_p(1)$, $\widehat{\mathrm{tr}(\Sigma^2)}-\mathrm{tr}(\Sigma^2)=O_p(1)$. Now we deal with H_1^R:

$$
\begin{aligned}
|H_1^R|&\le\left|\frac{\{2\mathrm{tr}(\Sigma^2)\}^{1/2}}{\{2\widehat{\mathrm{tr}(\Sigma^2)}\}^{1/2}}-1\right|\cdot\left|\frac{\frac{2nl_n^2}{(n+2)^2}W(\mu_0,k_n)-tr(\Sigma)}{\{2\mathrm{tr}(\Sigma^2)\}^{1/2}}\right|\\
&\le\{\widehat{\mathrm{tr}(\Sigma^2)}\}^{-1/2}[\{\widehat{\mathrm{tr}(\Sigma^2)}\}^{1/2}+\{\mathrm{tr}(\Sigma^2)\}^{1/2}]^{-1}\cdot\left|\mathrm{tr}(\Sigma^2)-\widehat{\mathrm{tr}(\Sigma^2)}\right|\cdot\left|\frac{\frac{2nl_n^2}{(n+2)^2}W(\mu_0,k_n)-\mathrm{tr}(\Sigma)}{\{2\mathrm{tr}(\Sigma^2)\}^{1/2}}\right|\\
&=O_p(1/p)O_p(1)=o_p(1).
\end{aligned}
$$

Similarly, we have

$$
|H_2^R|\le\frac{1}{\{2\mathrm{tr}(\Sigma^2)\}^{1/2}}\left|\widehat{\mathrm{tr}(\Sigma)}-\mathrm{tr}(\Sigma)\right|=O_p(1/p)O_p(1)=o_p(1).
$$

For H_3^R, we have

$$
\begin{aligned}
|H_3^R|&\le\left|\{2\widehat{\mathrm{tr}(\Sigma^2)}\}^{-1/2}-\{2\mathrm{tr}(\Sigma^2)\}^{-1/2}\right|\cdot\left|\widehat{\mathrm{tr}(\Sigma)}-\mathrm{tr}(\Sigma)\right|\\
&\le C\{\mathrm{tr}(\Sigma^2)\cdot\widehat{\mathrm{tr}(\Sigma^2)}\}^{-1/2}\cdot[\{\widehat{\mathrm{tr}(\Sigma^2)}\}^{1/2}+\{\mathrm{tr}(\Sigma^2)\}^{1/2}]^{-1}\cdot\left|\mathrm{tr}(\Sigma^2)-\widehat{\mathrm{tr}(\Sigma^2)}\right|\cdot\left|\widehat{\mathrm{tr}(\Sigma)}-\mathrm{tr}(\Sigma)\right|\\
&=O_p(p^{-3/2})O_p(1)=o_p(1).
\end{aligned}
$$

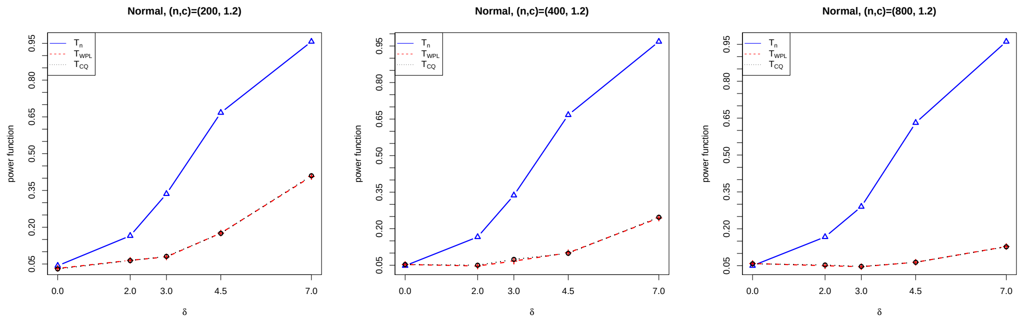
The proof of Lemma 2 is completed.

S.3. ADDITIONAL NUMERICAL RESULTS

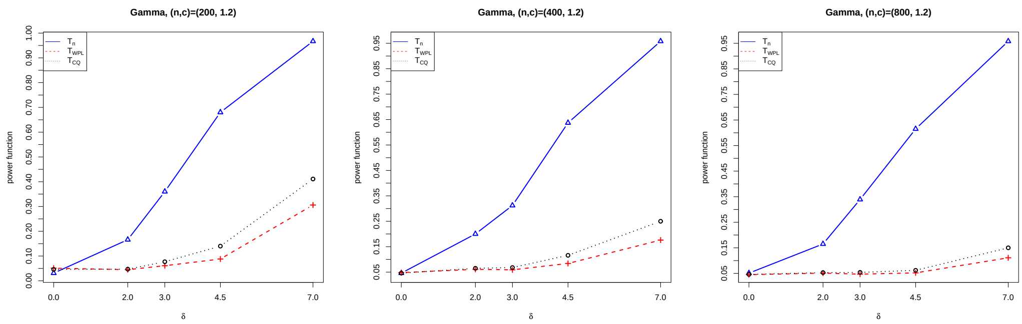
S.3.1. *Simulation results*

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.

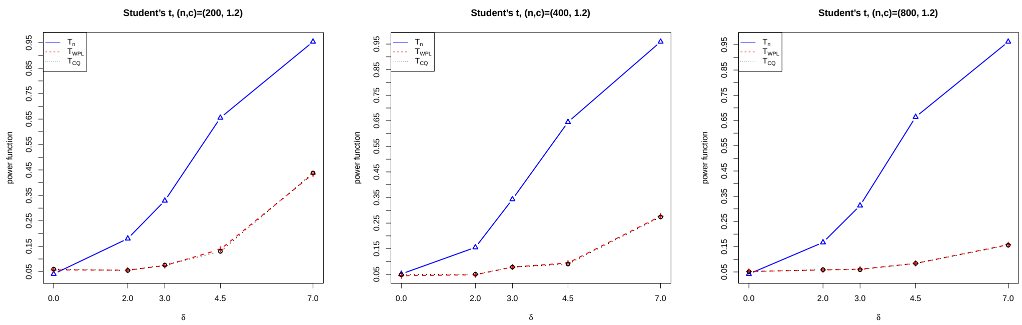
Linear hypothesis. We now examine the performance of T_n^F for the linear hypothesis in Section 2.3. We take $\mu_0=0$, $\mu=\delta(2,1,\cdots,1)^T/\sqrt{n}$ with $\delta=0$, 4, 7, 8 and 9, and $F=(f_{ij})$, a $q\times p$ matrix with $f_{jj}=1$, $f_{j,j+1}=-1$ for $j=1,\cdots,q$, and all other elements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



(a) $z_j \sim N(0,1)$



(b) $z_j \sim Gamma(4,2) - 2$



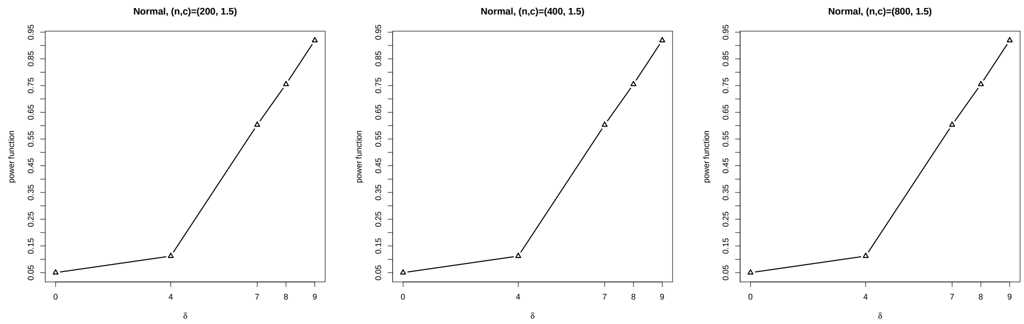
(c) $z_j \sim (3/5)^{1/2} t(5)$

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 $z_j \sim N(0,1), Gamma(4,2) - 2$ 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.

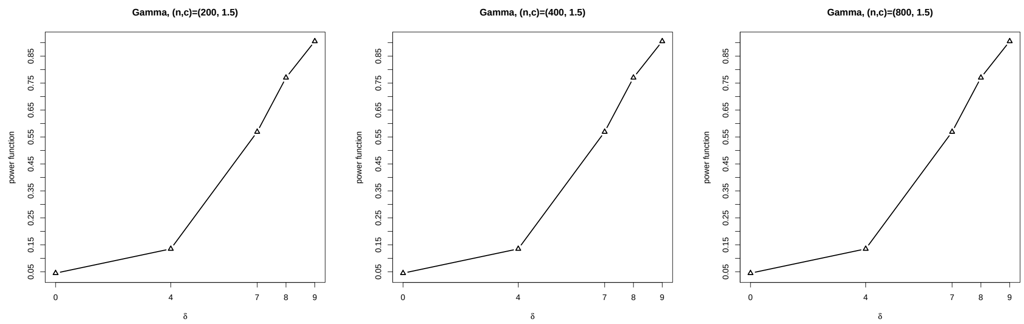
$(q/\log q)^{1/2}$, $\gamma = (1, \ldots, 1)^T/\sqrt{p}$ in T_n^F. Figure S2 depicts the empirical powers of T_n^F based on 1000 simulations. From Figure S2, we can see that the empirical rejection probabilities under H_0 are very close to 0.05 across all cases, when $\delta = 0$. This indicates that the limiting null distribution provides correct critical values. Also from Figure S2, the power functions increase rapidly and approach one as the value of δ increases.

S.3.2. *Real data analysis*

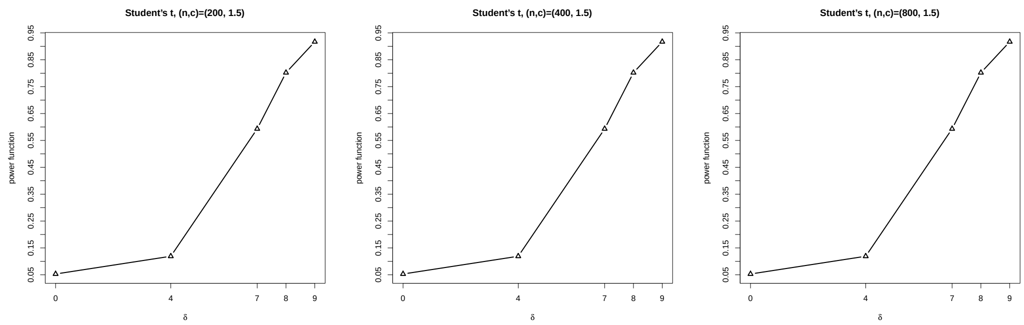
In this section we illustrate the proposed test procedure by an empirical analysis of stock data, which consist of nine sectors (consumer discretionary (CD), consumer staples (CS), en-



(a) $z_j \sim N(0,1)$



(b) $z_j \sim Gamma(4,2) - 2$



(c) $z_j \sim (3/5)^{1/2} t(5)$

Fig. S2. Power functions for assessing the performance of T_n^F under local alternatives H_a

ergy, financials, health care (HC), industrials (IND), materials, information technology (IT) and utilities) in the S&P 500 over a 25-month period from December 31, 2013. We mainly focus on testing whether the mean vector of the monthly rate of returns of stocks in each sector is equal to zero.

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 $\{h_{ji}^{(k)}, j = 1, \cdots, p_k, i = 1, \cdots, 25\}$. The i-th, $i = 1, \ldots, 24$, log-

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

Sector	p	T_n	T_{CQ}	T_{WPL}	EL-Pvalue	CQ-Pvalue	WPL-Pvalue
CD	83	0.1954	0.2448	0.4054	0.8451	0.8066	0.6852
CS	38	2.2011	1.2698	1.1877	0.0277	0.2042	0.2349
Energy	40	0.7948	0.2830	0.9466	0.2601	0.7772	0.3439
Financials	80	-0.0715	-0.0455	0.0498	0.9430	0.9637	0.9603
HC	46	-0.5791	-0.4244	-0.2561	0.5625	0.6713	0.7978
IND	60	0.7227	0.4847	1.0451	0.4698	0.6279	0.2960
IT	63	-1.6169	-1.5866	-0.9996	0.1059	0.1126	0.3175
Materials	28	2.6797	2.4471	1.9694	0.0074	0.0144	0.0489
Utilities	32	0.2111	0.2870	-0.3202	0.8328	0.7741	0.7488

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

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

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

BAI, Z. & SARANDASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Stat Sinica* **6**, 311–329.

CHEN, S. X., ZHANG, L.-X. & ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Am. Statist. Ass.* **105**, 810–819.

GÖTZE, F. & TIKHOMIROV, A. (2002). Asymptotic distribution of quadratic forms and applications. *J. Theor. Probab.* **15**, 423–475.