Ann Inst Stat Math (2017) 69:249–269
DOI 10.1007/s10463-015-0550-9

A consistent jackknife empirical likelihood test for distribution functions

Xiaohui Liu[1,2] · **Qihua Wang**[1,3] · **Yi Liu**[1]

Received: 2 March 2015 / Revised: 24 August 2015 / Published online: 1 February 2016

Abstract In this paper, a jackknife empirical likelihood based approach is developed to test whether the underlying distribution is equal to a specified one. The limiting distribution of the proposed testing statistic is derived under some mild conditions. It turns out that the proposed test is consistent and easy to be implemented. Some simulation studies are conducted to evaluate the finite sample behaviors by comparing the proposed method with the existing one. A real data example is also analyzed to illustrate the proposed test approach.

Keywords Jackknife empirical likelihood · Estimating equations · Cramér–von Mises test

1 Introduction

A canonical testing problem in statistics is that of testing whether an independent random p-dimensional ($p \geq 1$) sample $\{X_1, X_2, \ldots, X_n\}$ comes from a specified distribution $F_0(x)$, i.e., testing the following hypothesis

$$\mathcal{H}_0 : F(\cdot) \equiv F_0(\cdot) \quad \text{versus} \quad \mathcal{H}_1 : F(\cdot) \not\equiv F_0(\cdot), \tag{1}$$

where $F(x)$ denotes the underlying distribution of X_i $(1 \leq i \leq n)$.

✉ Qihua Wang
qhwang@amss.ac.cn

1 Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100190, China

2 School of Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China

3 Institute of Statistical Science, Shenzhen University, Shenzhen 518006, China

This testing problem arises in many scientific applications, such as econometrics and signal processing, and has been intensively studied in history. Many famous test statistics have been proposed. To name but a few, Kolmogorov test statistic, Cramér–von Mises test statistics, Anderson–Darling test statistics and Watson test statistic. A detailed account of these kinds of tests can be found in the monograph of D'Agostino and Stephens (1986).

It is known that some tests above can be reasonable powerful. Nevertheless, since most of their limiting distributions are generally not standard, the computation of the related critical values uses the bootstrap method in the practical applications. As a result, these tests are quite computationally intensive. Also, most of these tests consider only the case of $p = 1$. Furthermore, it is also inconvenient for such tests to utilize some auxiliary information, which may be available in practice, for improving their performances.

As an alternative, many authors have tried to consider this testing problem in the context of empirical likelihood, which offers several key benefits, e.g., the ease of using the auxiliary information to improve inference by adding constrains. Typical is the work of Einmahl and McKeague (2003) for a local empirical likelihood method. Unfortunately, their test is developed only for $p = 1$, and still very time-consuming for calculation because of processing a non-standard limiting distribution.

Recently, Feng and Peng (2012) proposed to consider this testing problem by using jackknife empirical likelihood (JEL). It seems that the JEL-based method appears to be very favorable due to its standard limiting distributions. However, from a theoretical point of view, the hypotheses that their method is based is not equivalent to the original hypothesis (1) they expected to test. The power values of their tests may decrease and hence may not converge to 1 as the sample size increases. That is, the test is inconsistent. On the other hand, their procedure is computationally intensive because it involves numerical integrations, and hence is hardly practical for application as the dimension of X is high.

To solve these problems, we first suggest a hypothesis which is equivalent to (1), and then construct a JEL test based on it. It is proved that the proposed test is consistent. Furthermore, we derive the limiting distribution of the proposed empirical log-likelihood ratio under some mild conditions. The proposed procedure is easy to be implemented and runs very fast even in high dimensions since the numerical integration is avoided for the proposed method.

JEL was first studied by Jing et al. (2009). This method has been proved useful in the applications involving nonlinear statistics such as U-statistics. The main benefit from JEL is the capability of substantially lessening the calculative burden of the ordinary empirical likelihood by introducing the so-called 'jackknife pseudo-values'. The ordinary empirical method was first proposed by Owen (1988). In the past decades, empirical likelihood has emerged as a powerful nonparametric method in statistics. See, for example, Owen (1990), Qin and Lawless (1994), Wang and Rao (2002), etc. For an excellent summary about the earlier developments of empirical likelihood, we refer the readers to Owen (2001). The updated results concerning the large dimensional empirical likelihood can be found in Chen et al. (2015) and reference therein.

The rest of this paper is organized as follows. In Sect. 2, we introduce the methodology and present the main results. In Sect. 3, we conduct some simulation studies to compare the proposed method with the JEL method due to Feng and Peng (2012). In Sect. 4, a real data example is analyzed to illustrate the proposed method. In Sect. 5, we make some brief discussions. The detailed proofs of the main results are provided in the Appendix.

2 Methodology and main results

Feng and Peng (2012) conducted a jackknife empirical likelihood based test for (1) relying on the fact that: if $F(x) \equiv F_0(x)$, then

$$\begin{cases} \int F^2(x)dF_0(x) = \beta_1 \\ \int F(x)F_0(x)dF_0(x) + \sum_{l=1}^{d} \int \sqrt{1 - F_{0l}^2(x_l)}dF_l(x_l) = \beta_2 \end{cases} \tag{2}$$

where $\beta_1 = \int F_0^2(x)dF_0(x)$, and $\beta_2 = \beta_1 + \frac{d\pi}{4}$. However, the above hypotheses that they considered are not equivalent to the original hypothesis (1), since for some $F(x) \not\equiv F_0(x)$, the equations in (2) may still hold. This implies that the test may be inconsistent. On the other hand, such a method concerns computing some quantities such as

$$\int \mathbb{I}(X \le x)F_0(x)dF_0(x),$$

which may be very time-consuming for calculation especially when dimension of X is greater than 1, although the fast computation seems possible in few cases, for example, when $F_0(x)$ is a normal distribution. This would consequently make the corresponding JEL-based tests not so practical.

This motivates us to suggest a new test which improves the JEL in following two points:

(1) Suggest a hypothesis which is equivalent to (1) such that the hypothesis-based jackknife empirical likelihood ratio test is consistent.
(2) Avoiding the numerical integration calculation on $F_0(\cdot)$.

Denote

$$\begin{aligned} \gamma &= \int (F(x) - F_0(x))^2 \omega_0(x)dF(x) \\ &= \int (F(x) - F_0(x))F(x)\omega_0(x)dF(x) - \int (F(x) - F_0(x))F_0(x)\omega_0(x)dF(x), \end{aligned}$$

where $\omega_0(x) > 0$ is a known bounded weight function for increasing flexility. When $\omega_0(x) \equiv 1$, γ is closely related to the Cramér–von Mises type test.

Note that $F(x) \equiv F_0(x)$ is equivalent to $\gamma = 0$, which further implies that

$$\begin{cases} \int (F(x) - F_0(x))F(x)\omega_0(x)dF(x) = 0 \\ \int (F(x) - F_0(x))F_0(x)\omega_0(x)dF(x) = 0. \end{cases} \tag{3}$$

Therefore, we suggest the hypothesis:

$$\widetilde{\mathcal{H}}_0 : F(\cdot) \in \mathcal{F}_{\text{dist}} \quad \text{versus} \quad \widetilde{\mathcal{H}}_1 : F(\cdot) \notin \mathcal{F}_{\text{dist}}, \tag{4}$$

where $\mathcal{F}_{\text{dist}} = \{F(\cdot) : F(\cdot)$ is the distribution function satisfying $(T_1(F), T_2(F), T_3(F)) = \Theta_0\}$, with

$$\begin{aligned}
T_1(F) &= \int (F(x) - F_0(x)) F(x) \omega_0(x) dF(x), \\
T_2(F) &= \int (F(x) - F_0(x)) F_0(x) \omega_0(x) dF(x) + \sum_{l=1}^{p} \int m_0(x_l) dF_l(x_l), \\
T_3(F) &= \sum_{l=1}^{p} \int m_0(x_l) dF_l(x_l), \\
\Theta_0 &= (0,\ \Theta_0,\ \Theta_0)^\top, \\
\Theta_0 &= \sum_{l=1}^{p} \int m_0(x_l) dF_{0l}(x_l),
\end{aligned}$$

$F_l(\cdot)$ and $F_{0l}(\cdot)$ denote the marginal distributions of $F(\cdot)$ and $F_0(\cdot)$, respectively, and $m_0(\cdot)$ a known bounded function.

Clearly, (4) is equivalent to (1). (4) is not equivalent to (1) if one removes $T_3 = \Theta_0$ from (4). The main purpose of using $\sum_{l=1}^{p} \int m_0(x_l) dF_l(x_l)$ in $T_2(F)$ and $T_3(F)$ is to avoid the degenerate matrix problem in deriving the empirical likelihood-type test as in Feng and Peng (2012). In practice, $m_0(\cdot)$ is usually selected as a functional depending on $F_{0l}(\cdot)$. For example, one may take $m_0(r) = (1 - F_{0l}^2(r))^{1/2}$ as in Feng and Peng (2012). Clearly, it satisfies Assumption **A5** given in the Appendix. For convenience, we drop the argument F from $T_1(F)$, $T_2(F)$ and $T_3(F)$ in what follows if no confusion arises.

Note that both T_1 and T_2 are nonlinear. Therefore, we suggest testing this hypothesis by using the jackknife empirical likelihood method. Given $\{X_1,\ X_2,\ \ldots,\ X_n\}$, denote the jackknife pseudo-sample of $(T_1,\ T_2,\ T_3)^\top$ to be $\mathbf{V}_i = (V_{1i},\ V_{2i},\ V_{3i})^\top$, $i = 1,\ 2,\ \ldots,\ n$, where

$$\begin{aligned}
V_{1i} &= n\widehat{T}_1 - (n-1)\widehat{T}_{1,i}, \\
V_{2i} &= n\widehat{T}_2 - (n-1)\widehat{T}_{2,i}, \\
V_{3i} &= n\widehat{T}_3 - (n-1)\widehat{T}_{3,i},
\end{aligned}$$

where $\widehat{T}_1$, $\widehat{T}_2$ and $\widehat{T}_3$ denote the estimators of T_1, T_2 and T_3, which are given later, and $\widehat{T}_{1,i}$, $\widehat{T}_{2,i}$ and $\widehat{T}_{3,i}$ the "leave ith sample out" versions of $\widehat{T}_1$, $\widehat{T}_2$, and $\widehat{T}_3$, respectively.

Note that both T_2 and T_3 concern the same unknown quantity

$$S = \sum_{l=1}^{p} \int m_0(x_l) dF_l(x_l). \tag{5}$$

Hence, we face the problem that it is hard to derive a consistent test based on (4) if one estimates the unknown quantity in both T_2 and T_3 by the same estimator. In this case, some matrices, which are required for establishing consistent test and deriving some asymptotic properties, may be degenerate. For example, the following matrix

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{V}_i - \Theta_0)(\mathbf{V}_i - \Theta_0)^\top$$

converges to a degenerate matrix with rank two rather than three. To avoid this problem, we estimate (5) in T_2 by

$$\widehat{S}_n = \frac{1}{n}\sum_{j=1}^{n} Z_j,$$

and in T_3 by

$$\widetilde{S}_n = \sum_{j=1}^{n}\frac{a_j}{\sum_{j=1}^{n} a_j} Z_j,$$

respectively, where $Z_j = \sum_{l=1}^{p} m_0(X_{jl})$, $\{a_1, a_2, \ldots, a_n\}$ denotes a sequence of non-negative real numbers. Assume that $\{a_1, a_2, \ldots, a_n\}$ satisfies all the assumptions given in the Appendix, and let $w_{ni} = a_i / \sum_{j=1}^{n} a_j$ $(1 \le i \le n)$. Then, Lemma 1 in the Appendix guarantees that $\widetilde{S}_n$ is also a consistent estimator of S as $\widehat{S}_n$, but with a different asymptotic variance from that of $\widehat{S}_n$ as long as $a_1, a_2, \ldots, a_n$ are chosen appropriately, which in turn guarantees the positive definiteness of the following matrix

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{V}_i - \Theta_0)(\mathbf{V}_i - \Theta_0)^\top$$

and some other related matrices.

This is a significant and novel idea or technique to use $T_3 = \Theta_0$ in (4) and two different estimators for S in T_2 and T_3 for defining a consistent JEL test with some asymptotic properties. In what follows, we take

$$\widehat{T}_1 = \int (\widehat{F}_n(x) - F_0(x))\widehat{F}_n(x)\omega_0(x)d\widehat{F}_n(x),$$
$$\widehat{T}_2 = \int (\widehat{F}_n(x) - F_0(x))F_0(x)\omega_0(x)d\widehat{F}_n(x) + \widehat{S}_n,$$
$$\widehat{T}_3 = \widetilde{S}_n,$$

where $\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{I}(X_j \le x)$ with $\mathbb{I}(\cdot)$ being the indicator function.

According to Jiang et al. (2011), based on these jackknife pseudo-values, a jackknife empirical likelihood function can be defined by

$$\ell_n(\Theta_0) = \sup_{(p_1, p_2, \ldots, p_n)} \left\{ \prod_{i=1}^{n} (np_i) : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \mathbf{V}_i = \Theta_0 \right\}.$$

Its limiting distribution is stated in the following theorem.

Theorem 1 *Suppose Assumptions A1–A5 given in the Appendix hold. Then under the null hypothesis $\mathcal{H}_0$ (or equivalently $\widetilde{\mathcal{H}}_0$), we have*

$$-2\log \ell_n(\Theta_0) \xrightarrow{\mathcal{L}} Q^\top \Sigma_2^{-1} Q,$$

as $n \to \infty$, where $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution, $Q \sim N(0, \Sigma_1)$ and Σ_1 and Σ_2 are defined in (7) and (13) in the Appendix, respectively.

It is worth mentioning that, for a general sequence of $\{a_1, a_2, \ldots, a_n\}$, the matrix Σ_1 does not necessarily equal to Σ_2. When $\Sigma_1 \neq \Sigma_2$, the result of Theorem 1 can not be directly utilized in practice, because the limiting distribution of $-2\log \ell_n(\Theta_0)$ is non-standard chi-square and involves some unknown quantities, namely, unknown eigenvalues of $\Sigma_2^{-1}\Sigma_1$. For this case, we suggest adjusting $-2\log \ell_n(\Theta_0)$, as did in Wang and Rao (2002), by multiplying the factor:

$$\widehat{r}(\Theta_0) = \mathrm{tr}\left(\widehat{\Sigma}_1^{-1}\widehat{\Omega}\right)/\mathrm{tr}\left(\widehat{\Sigma}_2^{-1}\widehat{\Omega}\right),$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$ are defined in (16) in the Appendix, and

$$\widehat{\Omega} = \left(\frac{1}{n}\sum_{i=1}^{n}(\mathbf{V}_i - \Theta_0)\right)\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbf{V}_i - \Theta_0)\right)^\top.$$

For this adjusted jackknife empirical log-likelihood ratio, we have the following result.

Theorem 2 *Suppose the same assumptions of Theorem 1, we have*

$$-2\widehat{r}(\Theta_0)\log \ell_n(\Theta_0) \xrightarrow{\mathcal{L}} \chi_3^2,$$

as $n \to \infty$, where χ_3^2 denotes a standard chi-square-distributed variable with degrees of freedom three.

However, for a small sample size, it is hard to estimate Σ_1 and Σ_2 well, which in turn leads to poor performances of the proposed jackknife empirical likelihood based test, while for a large sample size, multiplying such an adjusted factor $\widehat{r}(\Theta_0)$ requires much more computation time as can be seen from the construction of $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$. Therefore, to facilitate the applications, special attention needs to be paid to the choice of $\{a_1, a_2, \ldots, a_n\}$ so that $\Sigma_1 = \Sigma_2$. Fortunately, desirable sequences do exist in practice. The regular assumptions that the desirable sequences satisfying $\Sigma_1 = \Sigma_2$

are given in the Appendix. The sequence provided in the following Theorem 3 is a special example of the sequences.

Theorem 3 *When A5 holds and $a_i = 1$ if $1 \le i \le \lfloor n\tau \rfloor$, $a_i = 0$ otherwise. Then under the null hypothesis $\mathcal{H}_0$ (or equivalently $\widetilde{\widetilde{\mathcal{H}}}_0$), we have that*

$$-2\log \ell_n(\Theta_0) \xrightarrow{\mathcal{L}} \chi_3^2,$$

as $n \to \infty$, where τ is a constant satisfying $0 < \tau < 1$, $\lfloor \cdot \rfloor$ denotes the floor function.

Remark 1 In Theorem 3, τ actually denotes the proportion of the data used in estimating S in T_3. According to Theorem 3, τ should be smaller than 1. On the other hand, a too small τ may result in a very inefficient estimator of S in T_3. Hence, as a tradeoff, we recommend empirically to choose a moderate τ, i.e., 0.8, in practical applications.

Theorem 3 implies that, with a proper choice of $\{a_1, a_2, \ldots, a_n\}$, an α-level JEL test rejects $\mathcal{H}_0 : F(\cdot) \equiv F_0(\cdot)$ if $-2\log \ell_n(\Theta_0) \ge \chi^2_{3,1-\alpha}$, where $\chi^2_{3,1-\alpha}$ is the α-quantile of χ_3^2. The power of the proposed test comes from the fact that $\Theta_* \ne 0$ if $\mathcal{H}_0 : F(\cdot) \equiv F_0(\cdot)$ is violated and hence the limiting of $\frac{1}{n}\sum_{i=1}^n(\mathbf{V}_i - \Theta_0)$ is not zero, where

$$\Theta_* = \begin{pmatrix} T_1(F) \\ T_2(F) - \Theta_0 \\ T_3(F) - \Theta_0 \end{pmatrix}.$$

Consequently, there exists a positive constant ϵ_0 such that

$$-2\log \ell_n(\Theta_0) \ge \epsilon_0 \cdot \lambda_{\max}^{-1} \cdot \sqrt{n} \cdot \left\| \frac{1}{n}\sum_{i=1}^n (\mathbf{V}_i - \Theta_0) \right\|^2 \to \infty$$

with probability tending to 1 as $n \to \infty$, where $\|\cdot\|$ denotes the Euclidean distance and $\lambda_{\max}$ the maximum eigenvalue of Σ_2. That is,

$$P\left(-2\log \ell_n(\Theta_0) \ge \chi^2_{3,1-\alpha} | \mathcal{H}_1\right) \to 1, \quad \text{as } n \to \infty,$$

namely, the proposed test is consistent.

3 Simulation studies

We conducted some simulation studies to evaluate the proposed testing method by comparing it with that due to Feng and Peng (2012) in terms of size, power and time consumption for calculation. We do not report the results of some other tests here since Feng and Peng (2012) made a thorough discussion and comparison between their method and other testing methods, including the Cramér–von Mises test and the test statistic in Einmahl and McKeague (2003) (a type of Anderson–Darling test), and displayed some advantages of their method.

The data were, respectively, generated from the following distributions

$$\text{(D1)} \quad (1-\delta)\, N(\mu,\, \sigma^2) + \delta t_1(\zeta),$$
$$\text{(D2)} \quad (1-\delta)\, N\left(\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 0.5\sigma \\ 0.5\sigma & \sigma^2 \end{pmatrix}\right) + \delta t_2(\zeta),$$

where $t_k(\zeta)$ $(k = 1,\, 2)$ denote the k-dimensional t-distribution having independent components with degrees of freedom ζ. The null hypothesis is true when $\delta = 0$. Various combinations of $(\mu,\, \sigma,\, \delta,\, n)$ are investigated with $\zeta = 1$. The nominal significance level is set to be 0.05. In these simulations, $a_i = 1$ if $1 \le i \le \lfloor n\tau \rfloor$ with $\tau = 0.80$, otherwise $a_i = 0$, and $m_0(r) = \arctan(r)$.

The empirical size and power values based on 1000 repeated computations for the proposed method and that due to Feng and Peng (2012) were reported in Tables 1 and 2. For scenario (D1), Table 1 shows that the proposed method performs better than that of Feng and Peng (2012) in terms of size. Especially, the empirical sizes of the testing method due to Feng and Peng (2012) are about one and even two times larger than the nominal level 0.05 when $n = 100$ and 200. In most cases, the power values of the proposed method are larger than that of Feng and Peng (2012) and even much larger for some cases such as $(\mu, \sigma, \delta) = (1, 0.3, 0.10)$ and $(1, 0.3, 0.15)$ in Table 1. Also, it is noted that the power values of the testing method of Feng and Peng (2012) are larger than the proposed method for some cases, but it is generally slight, and the power value of their methods does not increase as the sample size increases for some cases such as $(\mu, \sigma, \delta) = (-1, 3, 0.10)$ and $(-1, 3, 0.15)$ in Table 1. This may be caused by the inconsistency of their test. For scenario (D2), the proposed method outperforms the method of Feng and Peng (2012) in terms of both size and power in most combinations of $(\mu,\, \sigma,\, \delta,\, n)$. The power values of the proposed method are far larger or even 3 times larger than that of Feng and Peng (2012) for some cases such as $(\mu, \sigma, \delta) = (1, 0.3, 0.10)$, $(1, 0.3, 0.15)$, $(1, 0.3, 0.10)$ and $(1, 0.3, 0.10)$ in Table 2.

For the proposed method, both Tables 1, 2 suggest that the power value of the proposed method increases when the sample size n increases. This confirms the theoretical results of this paper.

Note that the computing issue is of great interest for the practical applications of a statistical procedure. Hence we also recorded the computation times (in s) of these two methods. It is found that the computation time does obviously not depend on the choice of the combination of $(\mu,\, \sigma,\, \delta)$ for a fixed n. Therefore, in the sequel we only present the computation time for scenario (D2) when $(\mu,\, \sigma,\, \delta) = (-1,\, 0.3,\, 0.00)$ with $100 \le n \le 800$, respectively. Figure 1 plotted the time curves for calculation of the two methods for *one run*. Figure 1 suggests that the proposed method runs far faster than that of Feng and Peng (2012). Especially, it should be pointed out that the calculation for the Feng and Peng (2012) testing method is hardly practical when the dimension of X is large. The numerical integrations involved in the test of Feng and Peng (2012) are computed by utilizing *dblquad.m*. All results are obtained on a HP

Table 1 Empirical size and power values of the proposed method (New) and that of Feng and Peng (2012) (FP2012) for Scenario (D1)

$(\mu,\ \sigma,\ \delta)$	$n = 100$		$n = 200$		$n = 300$	
	New	FP2012	New	FP2012	New	FP2012
(1, 0.3, 0.00)	0.0680	0.1630	0.0570	0.1010	0.0540	0.0740
(1, 0.3, 0.10)	0.4260	0.1870	0.6410	0.2820	0.8400	0.4910
(1, 0.3, 0.15)	0.7040	0.2370	0.9290	0.6250	0.9850	0.8680
(1, 0.3, 0.30)	0.9950	0.8460	1.0000	0.9990	1.0000	1.0000
(1, 0.3, 0.50)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(1, 0.3, 0.70)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(1, 3, 0.00)	0.0560	0.1500	0.0520	0.1010	0.0630	0.0810
(1, 3, 0.10)	0.0680	0.0910	0.0900	0.0710	0.1250	0.0710
(1, 3, 0.15)	0.0970	0.0890	0.1560	0.0920	0.1720	0.1290
(1, 3, 0.30)	0.2190	0.1290	0.4510	0.3260	0.5870	0.4930
(1, 3, 0.50)	0.5690	0.3970	0.8760	0.8350	0.9690	0.9730
(1, 3, 0.70)	0.8850	0.8070	0.9970	0.9990	1.0000	1.0000
(−1, 0.3, 0.00)	0.0600	0.1210	0.0590	0.0990	0.0510	0.0660
(−1, 0.3, 0.10)	0.3990	0.5470	0.5690	0.9160	0.7430	0.9950
(−1, 0.3, 0.15)	0.6400	0.8770	0.8580	0.9990	0.9970	1.0000
(−1, 0.3, 0.30)	0.9920	0.9990	1.0000	1.0000	1.0000	1.0000
(−1, 0.3, 0.50)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(−1, 0.3, 0.70)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(−1, 3, 0.00)	0.0560	0.1450	0.0530	0.0910	0.0450	0.0900
(−1, 3, 0.10)	0.0730	0.1910	0.1080	0.1540	0.1240	0.1240
(−1, 3, 0.15)	0.0910	0.2220	0.1720	0.1740	0.2190	0.1640
(−1, 3, 0.30)	0.2560	0.3220	0.4810	0.3680	0.7110	0.3980
(−1, 3, 0.50)	0.6320	0.4690	0.9450	0.6320	0.9910	0.7860
(−1, 3, 0.70)	0.9130	0.6840	0.9980	0.8740	1.0000	0.9600

Here the null hypothesis is true when $\delta = 0.00$

Pavilion dv7 Notebook PC with Intel(R) Core(TM) i7-2670QM CPU @ 2.20GHz, RAM 6.00GB, Windows 7 Home Premium and Matlab 7.8.

4 A real data example

To gain more insight to the proposed testing method, we also provide a real data example. The data set $\mathcal{X}^n = \{X_i\}_{i=1}^n$ is a part of the daily simple returns of IBM stock from 2006 January 03 to 2008 December 31. It consists of 755 observations, i.e., $n = 755$. These data can be downloaded from the teaching page of Tsay (2010): http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/d-ibm3dx7008.txt. A thorough investigation on this data set is beyond the scope of this paper, but we utilize two columns under the titles of rtn and vwretd as an illustration of how to use the proposed method.

Table 2 Empirical size and power values of the proposed method (New) and that of Feng and Peng (2012) (FP2012) for Scenario (D2)

$(\mu,\ \sigma,\ \delta)$	$n = 100$		$n = 200$		$n = 300$	
	New	FP2012	New	FP2012	New	FP2012
(1, 0.3, 0.00)	0.0570	0.0760	0.0610	0.0570	0.0590	0.0480
(1, 0.3, 0.10)	0.2720	0.1250	0.4500	0.1650	0.6200	0.1710
(1, 0.3, 0.15)	0.4580	0.1760	0.7270	0.2500	0.8700	0.2850
(1, 0.3, 0.30)	0.8490	0.3260	0.9850	0.4770	0.9960	0.6270
(1, 0.3, 0.50)	0.9590	0.5050	0.9890	0.7360	0.9970	0.8830
(1, 0.3, 0.70)	0.9400	0.6210	0.9980	0.8580	0.9990	0.9600
(0, 0.3, 0.00)	0.0500	0.0690	0.0430	0.0600	0.0700	0.0530
(0, 0.3, 0.10)	0.2210	0.1450	0.3690	0.1820	0.5020	0.2450
(0, 0.3, 0.15)	0.3640	0.2170	0.6250	0.3110	0.7820	0.4530
(0, 0.3, 0.30)	0.7900	0.5080	0.9680	0.7800	0.9960	0.9000
(0, 0.3, 0.50)	0.9560	0.8750	0.9870	0.9880	0.9800	1.0000
(0, 0.3, 0.70)	0.9710	0.9880	0.9670	1.0000	0.9630	1.0000
(−1, 0.3, 0.00)	0.0470	0.0720	0.0440	0.0600	0.0540	0.0490
(−1, 0.3, 0.10)	0.2490	0.2350	0.3790	0.3680	0.5160	0.5260
(−1, 0.3, 0.15)	0.4030	0.4080	0.6610	0.6770	0.8400	0.8410
(−1, 0.3, 0.30)	0.8730	0.8980	0.9930	0.9960	1.0000	1.0000
(−1, 0.3, 0.50)	0.9940	1.0000	1.0000	1.0000	1.0000	1.0000
(−1, 0.3, 0.70)	0.9970	1.0000	1.0000	1.0000	1.0000	1.0000
(1, 3, 0.00)	0.0550	0.0750	0.0460	0.0600	0.0540	0.0650
(1, 3, 0.10)	0.1680	0.0800	0.2400	0.1010	0.3240	0.1280
(1, 3, 0.15)	0.2630	0.1220	0.4680	0.1630	0.6030	0.1880
(1, 3, 0.30)	0.6590	0.2550	0.9030	0.3730	0.9830	0.4660
(1, 3, 0.50)	0.9160	0.4550	0.9840	0.6780	0.9900	0.8440
(1, 3, 0.70)	0.9180	0.6350	0.9630	0.8980	0.9980	0.9700
(0, 3, 0.00)	0.0540	0.0640	0.0610	0.0520	0.0530	0.0590
(0, 3, 0.10)	0.1160	0.1120	0.1740	0.1220	0.2180	0.1370
(0, 3, 0.15)	0.1900	0.1520	0.2760	0.1750	0.3510	0.1900
(0, 3, 0.30)	0.4430	0.3090	0.6830	0.4350	0.8480	0.5960
(0, 3, 0.50)	0.7590	0.5850	0.9480	0.8410	0.9920	0.9340
(0, 3, 0.70)	0.9140	0.8260	0.9690	0.9740	0.9700	0.9990
(−1, 3, 0.00)	0.0470	0.0740	0.0540	0.0410	0.0570	0.0460
(−1, 3, 0.10)	0.1150	0.1420	0.1250	0.1410	0.1680	0.1750
(−1, 3, 0.15)	0.1580	0.1720	0.2140	0.2410	0.2940	0.3210
(−1, 3, 0.30)	0.3650	0.4150	0.6060	0.6660	0.7990	0.8030
(−1, 3, 0.50)	0.7130	0.8150	0.9410	0.9710	0.9880	0.9960
(−1, 3, 0.70)	0.9170	0.9770	0.9980	0.9990	1.0000	1.0000

Here the null hypothesis is true when $\delta = 0.00$

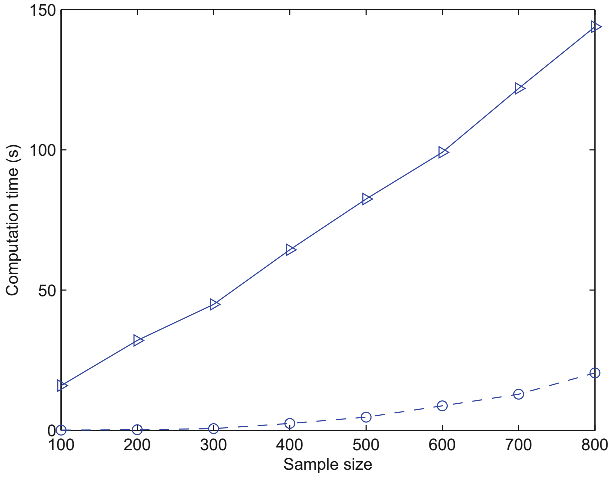

Fig. 1 The computation times (in s) of these two methods for scenario (D2) when $(\mu, \sigma, \delta) = (-1, 0.3, 0.00)$ with $n = 100, 200, \ldots, 800$. Here the curves with circle and triangle are the computation time curves of the proposed method and that of Feng and Peng (2012), respectively

The issue of interest is to test whether or not the simple returns are independently and identically distributed as normal with fixed mean and variance. The assumption of normal distribution makes statistical properties of returns tractable in financial study. Before performing the test, the following transformation is taken: $Y_i = \widehat{\Sigma}^{-1/2}(X_i - \widehat{\mu})$ on $\mathcal{X}^n$, where $\widehat{\mu}$, $\widehat{\Sigma}$ denote the sample mean and covariance-matrix of $\mathcal{X}^n$, respectively; see Fig. 2a for the scatter plot of Y_is. The null hypothesis is that Y_is are generated from the bivariate standard normal distribution.

In descriptive statistics, there are already some useful graphical methods for visualizing whether or not the observations are generated from an underlying normal distribution. For univariate data, the QQ-plot (quantile versus quantile plot) usually serves for this purpose. In higher dimensions, a similar tool is the *DD*-plot (depth versus depth plot) introduced by Liu et al. (1999). We plot the QQ-plot for both of the components of Y_is (see Fig. 2c, d), and the DD-plot of Y_is (see Fig. 2b) relying on the halfspace depth. The halfspace depth is capable to characterize the underlying empirical distribution and elliptically symmetric distributions; see Kong and Zuo (2010) and references therein for details.

Ideally, if the observations were generated from the underlying distribution, the points in both the QQ-plot and DD-plot should approximately lie on the line $y = x$. However, this is not the case for these points in Fig. 2b, c. These figures suggest rejecting the normal assumption.

We test the hypothesis by the proposed method under the same setting as that for the simulated data. The value of the empirical log-likelihood ratio multiplying -2 is 23.0732 (that of Feng and Peng (2012) is 25.6169), which is larger than 7.8147, the 0.05 quantile of χ_3^2. This suggests us to reject the null hypothesis under the nominal signification level 0.05. This coincides with the descriptive results provided by both the QQ-plot and DD-plot.

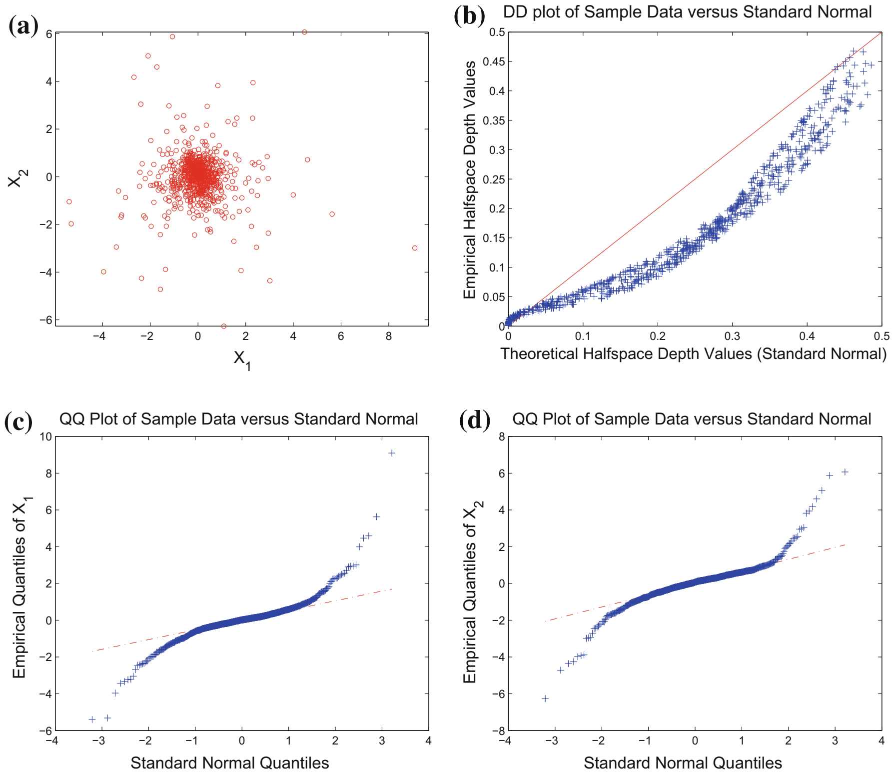

Fig. 2 The transformed daily simple returns Y_is of IBM stock from 2006 January 03 to 2008 December 31

5 Concluding discussions

It is of great interest to test whether the underlying distribution is equal to a specified one before making the data analysis. There is a long history to consider this issue in statistics. Many useful tools have been proposed in the literature. In this paper, we further enriched the toolkits for such tests by providing a new approach based on the jackknife empirical likelihood. One advantage of the proposed method is its capability of utilizing some auxiliary information if available. Compared to the methods developed by Feng and Peng (2012), the proposed testing statistics can be shown to be consistent. Its implementation is also very easy to be achieved, especially when the underlying distribution is not normally distributed in high dimensions. Similar methods may easily be extended to the case of composite null hypothesis, we hence did not pursue it in this paper. In the presence of missing values, most of the existing procedures are invalid for such a testing issue, how to construct the corresponding testing statistics has not been considered and is still worthy of further study.

Acknowledgements The authors greatly appreciate the thoughtful and constructive remarks of the Co-Editors and three anonymous referees, which led to many improvements in this paper. Xiaohui Liu's research was supported by the NNSF of China (No. 11461029), the NSF of Jiangxi Province (No. 20142BAB211014) and the YSFP of Jiangxi provincial education department (No. GJJ14350). Wang's research was supported by the National Science Fund for Distinguished Young Scholars in China (10725106), the National Natural Science Foundation of China (General program 11171331 and Key program 11331011), a grant from the Key Lab of Random Complex Structure and Data Science, CAS and Natural Science Foundation of SZU.

6 Appendix: proofs of the main results

To prove the main results, we need to assume that the sequence of $\{a_1, a_2, \ldots, a_n\}$ satisfies the following assumptions.

A1. $\max_{1\le i\le n}\left(\frac{a_i}{\sum_{l=1}^n a_l}\right)^2 \to 0$, as $n\to\infty$.

A2. $\lim_{n\to\infty}\sum_{i=1}^n \frac{na_i^2}{(\sum_{l=1}^n a_l)^2} = \kappa_0 < +\infty$.

A3. $\left(\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n b_j\right)^2 \ne \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n b_j^2$, where

$$b_j = a_j\left(\frac{n^2}{\sum_{l=1}^n a_l} - (n-1)\sum_{i=1,i\ne j}^n \frac{1}{\sum_{l=1,l\ne i}^n a_l}\right). \tag{6}$$

A4. $\left(\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n c_j\right)^2 \ne \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n c_j^2$, where $c_j = \frac{(n-1)a_j}{\sum_{l=1,l\ne j}^n a_l}$.

A5. $m_0(\cdot)$ must be chosen so that $\sigma_{11}\sigma_{22} - \sigma_{12}^2 > 0$, where σ_{11}, σ_{22} and σ_{12} are specified in (12).

We now briefly make some comments on these assumptions. **A1**–**A2** are assumed to guarantee the convergence and asymptotic normality of the estimator $\widetilde{S}_n$; see Lemma 1 and Jiang et al. (2011). **A3**–**A5** are technically used to avoid the degenerate problem of the proposed JEL ratio. The existence of the limits, namely, $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n b_j$, $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n b_j^2$, $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n c_j$ and $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n c_j^2$, can be derived by Assumptions **A1**–**A2**; see the proofs of Lemmas 2–3 for details. These assumptions are mild, and trivially hold in practice. A particular example is given in Theorem 3.

We need the following lemmas to derive the main results.

Lemma 1 *Assume that $Y_1, Y_2, \ldots, Y_n$ are independently and identically distributed with a finite mean μ_y. Let $\bar{Y}_n = \sum_{i=1}^n w_{ni}Y_i$, where $\{w_{n1}, w_{n2}, \ldots, w_{nn}\}$ is a sequence of real numbers, and satisfies the following conditions: (i) for any $1\le i\le n$, it holds $\lim_{n\to\infty} w_{ni} = 0$; (ii) $\lim_{n\to\infty}\sum_{i=1}^n w_{ni} = 1$; (iii) there exists a constant $M_0 > 0$ such that for every n, $\sum_{i=1}^n |w_{ni}| \le M_0$. Then,*

$$\bar{Y}_n \xrightarrow{p} \mu_y \quad\Leftrightarrow\quad \max_{1\le i\le n}|w_{ni}| \to 0,$$

as $n\to\infty$, where $\xrightarrow{p}$ denotes the convergence in probability.

Proof The proof of this lemma can be found in Pruitt (1966). □

Lemma 2 *Under the same assumptions of Theorem 1, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{V}_i-\Theta_0)\xrightarrow{\mathcal{L}} N(0,\ \Sigma_1),$$

where

$$\Sigma_1=\begin{pmatrix}\sigma_{11} & \sigma_{11}+\sigma_{12} & \kappa_1\sigma_{12}\\ \sigma_{11}+\sigma_{12} & \sigma_{11}+2\sigma_{12}+\sigma_{22} & \kappa_1(\sigma_{12}+\sigma_{22})\\ \kappa_1\sigma_{12} & \kappa_1(\sigma_{12}+\sigma_{22}) & \kappa_2\sigma_{22}\end{pmatrix}, \tag{7}$$

where $\kappa_1=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}b_i$, $\kappa_2=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}b_i^2$, σ_{11}, σ_{12} *and* σ_{22} *are specified in* (12).

Proof Let $\widehat{F}_{ni}(x)$ be the "leave the ith sample out" version of $\widehat{F}_n(x)$. Then, we have

$$\begin{aligned}V_{1i}&=n\widehat{T}_1-(n-1)\widehat{T}_{1,i}\\&=\left(n\int\widehat{F}_n^2(x)\omega_0(x)d\widehat{F}_n(x)-(n-1)\int\widehat{F}_{ni}^2(x)\omega_0(x)d\widehat{F}_{ni}(x)\right)\\&\quad-\left(n\int\widehat{F}_n(x)F_0(x)\omega_0(x)d\widehat{F}_n(x)-(n-1)\int\widehat{F}_{ni}(x)F_0(x)\omega_0(x)d\widehat{F}_{ni}(x)\right)\\&:=V_{1i}^{[1]}-V_{1i}^{[2]}.\end{aligned}\tag{8}$$

By $\widehat{F}_{ni}(x)=\frac{n}{n-1}\widehat{F}_n(x)-\frac{1}{n-1}\mathbb{I}(X_i\le x)$ and

$$\widehat{F}_{ni}^2(x)=\frac{n^2}{(n-1)^2}\widehat{F}_n^2(x)-2\frac{n}{(n-1)^2}\widehat{F}_n(x)\mathbb{I}(X_i\le x)+\frac{1}{(n-1)^2}\mathbb{I}(X_i\le x),$$

a direct calculation yields

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}V_{1i}^{[1]}&=\frac{1}{n}\sum_{i=1}^{n}\widehat{F}_{ni}^2(X_i)\omega_0(X_i)+\frac{-2n+1}{(n-1)^2}\sum_{j=1}^{n}\widehat{F}_n^2(X_j)\omega_0(X_j)\\&\quad+\frac{2n}{(n-1)^2}\sum_{j=1}^{n}\widehat{F}_n^2(X_j)\omega_0(X_j)+\frac{-1}{(n-1)^2}\sum_{j=1}^{n}\widehat{F}_n(X_j)\omega_0(X_j)\\&=\frac{1}{n}\sum_{i=1}^{n}\widehat{F}_n^2(X_i)\omega_0(X_i)+O_p(n^{-1}).\end{aligned}\tag{9}$$

Similarly, we have $\frac{1}{n}\sum_{i=1}^{n}V_{1i}^{[2]}=\frac{1}{n}\sum_{i=1}^{n}\widehat{F}_n(X_i)F_0(X_i)\omega_0(X_i)+O_p(n^{-1})$. This together with (8) and (9) leads to

$$\frac{1}{n}\sum_{i=1}^{n} V_{1i} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{F}_n(X_i) - F_0(X_i)\right)\widehat{F}_n(X_i)\omega_0(X_i) + O_p(n^{-1})$$
$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} h_1(X_i,\ X_j,\ X_k) + O_p(n^{-1}),$$

where $h_1(X_i,\ X_j,\ X_k) = \frac{1}{3!}\sum_I \widetilde{h}_1(X_{I_1},\ X_{I_2},\ X_{I_3})$, $\widetilde{h}_1(X_{I_1},\ X_{I_2},\ X_{I_3}) = (\mathbb{I}(X_{I_2} \le X_{I_1}) - F_0(X_{I_1}))\mathbb{I}(X_{I_3} \le X_{I_1})\omega_0(X_{I_1})$, and the sum $\sum_I$ taken over the set of all unordered subsets I of 3 different integers, namely, $I_1,\ I_2,\ I_3$, chosen from $\{i,\ j,\ k\}$. Then, by V-statistics theory (Serfling 1980), we have

$$\frac{1}{n}\sum_{i=1}^{n} V_{1i} = \binom{n}{3}^{-1}\sum_{1\le i<j<k\le n}\widetilde{h}_1(X_i,\ X_j,\ X_k) + o_p(n^{-1/2}).$$

Then, a direct application of Theorem 12.3 of Van der Vaart (2000) leads to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_{1i} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i) + o_p(1), \tag{10}$$

where $\eta(X_i) = E\left((\mathbb{I}(X_i \le X) - F_0(X))F_0(X)\omega_0(X)|\, X_i\right)$.

Similarly, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(V_{2i} - \Theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\eta(X_i) + Z_i - \Theta_0) + o_p(1). \tag{11}$$

Recalling the definition of V_{3i}, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(V_{3i} - \Theta_0)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(n\left(\widehat{T}_3 - \Theta_0\right) - (n-1)\frac{\sum_{l=1}^{n} a_l}{\sum_{l=1,l\ne i}^{n} a_l}\left(\left(\widehat{T}_3 - \Theta_0\right) - \frac{a_i(Z_i - \Theta_0)}{\sum_{l=1}^{n} a_l}\right)\right)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} b_j(Z_j - \Theta_0),
\end{aligned}$$

where b_j is defined in Assumption **A3**. This, combined with (10) and (11), leads to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{V}_i - \Theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{A}\xi_i + o_p(1),$$

where $\mathbb{A} = \begin{pmatrix} 1 & 0 & 0\\ 1 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$, and $\xi_i = \begin{pmatrix}\eta(X_i)\\ Z_i - \Theta_0\\ b_i(Z_i - \Theta_0)\end{pmatrix}$.

Clearly,

$$
\begin{aligned}
E(\xi_i \xi_i^\top) &= \begin{pmatrix} E(\eta^2(X_i)) & E(\eta(X_i)(Z_i - \Theta_0)) & b_i E(\eta(X_i)(Z_i - \Theta_0)) \\ E(\eta(X_i)(Z_i - \Theta_0)) & E((Z_i - \Theta_0)^2) & b_i E((Z_i - \Theta_0)^2) \\ b_i E(\eta(X_i)(Z_i - \Theta_0)) & b_i E((Z_i - \Theta_0)^2) & b_i^2 E((Z_i - \Theta_0)^2) \end{pmatrix} \\
&:= \begin{pmatrix} \sigma_{11} & \sigma_{12} & b_i \sigma_{12} \\ \sigma_{12} & \sigma_{22} & b_i \sigma_{22} \\ b_i \sigma_{12} & b_i \sigma_{22} & b_i^2 \sigma_{22} \end{pmatrix}. \qquad (12)
\end{aligned}
$$

Next, we prove the existence of $\kappa_2 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n b_j^2$. Since

$$
\begin{aligned}
\frac{1}{n} \sum_{j=1}^n b_j^2 &= \sum_{j=1}^n \left(n a_j^2 \left(\frac{n}{\sum_{l=1}^n a_l} - \frac{n-1}{n} \sum_{i=1, l\neq j}^n \frac{1}{\sum_{l=1, l\neq i}^n a_l} \right)^2 \right) \\
&= \sum_{j=1}^n \left(\frac{n a_j^2}{\left(\sum_{l=1}^n a_l\right)^2} \left(\frac{2n-1}{n} - \frac{n-1}{n} \sum_{i=1, i\neq j}^n \frac{a_i}{\sum_{l=1, l\neq i}^n a_l} \right)^2 \right),
\end{aligned}
$$

then, for a sufficient large n, by noting the fact

$$
\left(\frac{2n-1}{n} - \frac{n-1}{n} \sum_{i=1, i\neq j}^n \frac{a_i}{\sum_{l=1, l\neq i}^n a_l} \right)^2 \leq 8 + \frac{2}{1 + \max_{1\leq i\leq n} \left(\frac{a_i}{\sum_{l=1} a_l} \right)^2} < 10,
$$

we obtain $\lim\limits_{n\to\infty} \frac{1}{n} \sum_{j=1}^n b_j^2 < 10\kappa_0 < \infty$. Next, by invoking Jensen's inequality, we have $\lim_{n\to\infty} \left(\frac{1}{n} \sum_{j=1}^n |b_j| \right)^2 \leq \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n b_j^2 < \infty$, which implies the existence of $\kappa_1 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n b_j$.

Finally, by Assumptions **A1**–**A3**, we obtain

$$
\operatorname{cov}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{A} \xi_i \right) \xrightarrow{p} \mathbb{A} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \kappa_1 \sigma_{12} \\ \sigma_{12} & \sigma_{22} & \kappa_1 \sigma_{22} \\ \kappa_1 \sigma_{12} & \kappa_1 \sigma_{22} & \kappa_2 \sigma_{22} \end{pmatrix} \mathbb{A}^\top = \Sigma_1 > 0.
$$

This implies the Lindeberg condition. A direct use of the Lindeberg–Feller central limit theorem proves this lemma. □

Lemma 3 *Under Assumptions **A1**, **A2**, **A4** and **A5**, we have*

$$
\frac{1}{n} \sum_{i=1}^n (\boldsymbol{V}_i - \Theta_0)(\boldsymbol{V}_i - \Theta_0)^\top \xrightarrow{p} \Sigma_2
$$

with

$$\Sigma_2 = \begin{pmatrix} \sigma_{11} & \sigma_{11}+\sigma_{12} & \kappa_3\sigma_{12} \\ \sigma_{11}+\sigma_{12} & \sigma_{11}+2\sigma_{12}+\sigma_{22} & \kappa_3(\sigma_{12}+\sigma_{22}) \\ \kappa_3\sigma_{12} & \kappa_3(\sigma_{12}+\sigma_{22}) & \kappa_4\sigma_{22} \end{pmatrix}, \tag{13}$$

where $\kappa_3 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n c_i$, *and* $\kappa_4 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n c_i^2$.

Proof By the Glivenko–Cantelli theorem (Serfling 1980, Theorem A in page 61), we have

$$\sup_x \left|\widehat{F}_n(x) - F_0(x)\right| = o_p(1) \quad \text{and} \quad \max_{1\le i\le n}\sup_x \left|\widehat{F}_{ni}(x) - F_0(x)\right| = o_p(1). \tag{14}$$

Denote

$$\widetilde{V}_{1i} = \frac{n}{(n-1)^2}\sum_{j=1}^n \widetilde{h}_2(X_i,\, X_j) + \frac{1}{(n-1)^2}\sum_{j=1}^n (F_0(X_j)-1)\mathbb{I}(X_i \le X_j)\omega_0(X_j),$$

where $\widetilde{h}_2(X_i,\, X_j) = (\mathbb{I}(X_i \le X_j) - F_0(X_j))F_0(X_j)\omega_0(X_j)$. Decompose

$$\frac{1}{n}\sum_{i=1}^n V_{1i}^2 = \frac{1}{n}\sum_{i=1}^n \widetilde{V}_{1i}^2 + \frac{1}{n}\sum_{i=1}^n \left(V_{1i} - \widetilde{V}_{1i}\right)^2 + \frac{1}{n}\sum_{i=1}^n 2\widetilde{V}_{1i}\left(V_{1i} - \widetilde{V}_{1i}\right). \tag{15}$$

Note that the boundedness of $F_0(\cdot)$, $\mathbb{I}(\cdot)$ and $\omega_0(\cdot)$ implies that

$$\max_{1\le i\le n}\left|\widetilde{V}_{1i} - V_{1i}\right| \le \epsilon_0 \cdot \max\left\{\sup_x \left|\widehat{F}_n(x) - F_0(x)\right|,\ \max_{1\le i\le n}\sup_x \left|\widehat{F}_{ni}(x) - F_0(x)\right|\right\},$$
$$\max_{1\le i\le n}\left|\frac{1}{(n-1)^2}\sum_{j=1}^n (F_0(X_j)-1)\mathbb{I}(X_i \le X_j)\omega_0(X_j)\right| = o_p(1),$$

where ϵ_0 denotes a positive constant. This, together with (14) and (15), leads to

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n V_{1i}^2 &= \frac{1}{n}\sum_{i=1}^n \widetilde{V}_{1i}^2 + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^n \left(\frac{n}{(n-1)^2}\sum_{j=1}^n \widetilde{h}_2(X_i,\, X_j)\right)^2 + o_p(1) \end{aligned}$$

$$
\begin{aligned}
&= \frac{n^4}{(n-1)^4}\left(\frac{1}{n}\sum_{i=1}^{n}\eta^2(X_i) + \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widetilde{h}_2(X_i,\ X_j) - \eta(X_i)\right)\eta(X_i)\right.\\
&\quad \left. + \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\widetilde{h}_2(X_i,\ X_j) - \eta(X_i)\right)\left(\widetilde{h}_2(X_i,\ X_k) - \eta(X_i)\right)\right) + o_p(1)\\
&\xrightarrow{p} \sigma_{11},
\end{aligned}
$$

as $n \to \infty$, by V- and U-statistics theory (Serfling 1980) as in (10), and the following facts

$$
\begin{aligned}
&E\left(\left(\widetilde{h}_2(X_i,\ X_j) - \eta(X_i)\right)\eta(X_i)\right) = 0, \quad \text{for } i \neq j,\\
&E\left(\left(\widetilde{h}_2(X_i,\ X_j) - \eta(X_i)\right)\left(\widetilde{h}_2(X_i,\ X_k) - \eta(X_i)\right)\right) = 0, \ \text{for } i \neq k,\ i \neq j,\ j \neq k.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} V_{1i}(V_{2i} - \Theta_0) \xrightarrow{p} \sigma_{11} + \sigma_{12},\\
&\frac{1}{n}\sum_{i=1}^{n}(V_{2i} - \Theta_0)^2 \xrightarrow{p} \sigma_{11} + 2\sigma_{12} + \sigma_{22}, \quad \text{as } n \to +\infty.
\end{aligned}
$$

Note that $E\left(\widetilde{h}_2(X_i,\ X_j) - \eta(X_i)|X_i\right) = 0\ (i \neq j)$. Then, by Bernstein's inequality (Serfling 1980) and the boundedness of $F_0(\cdot)$, $\mathbb{I}(\cdot)$ and $\omega_0(\cdot)$, we have that

$$
\begin{aligned}
&P\left(\max_{1\le i\le n}\left|\frac{1}{n}\sum_{j=1, j\neq i}^{n}\widetilde{h}_2(X_i,\ X_j) - \eta(X_i)\right| \ge \epsilon_0 n^{-1/2+s/2}\right)\\
&\le \sum_{i=1}^{n} E\left(P_{X_i}\left(\left|\sum_{j=1, j\neq i}^{n}\widetilde{h}_2(X_i,\ X_j) - \eta(X_i)\right| \ge \epsilon_0 n^{1/2+s/2}\right)\right)\\
&\le \sum_{i=1}^{n} E\left(\exp\left(\frac{-\frac{1}{2}\epsilon_0^2 n^{1+s}}{\sum_{j=1, j\neq i}^{n} E_{X_i}(\widetilde{h}_2(X_i,\ X_j) - \eta(X_i))^2 + M_0\epsilon_0 n^{1/2+s/2}/3}\right)\right)\\
&\le 2n\exp\left(-\epsilon_1 n^{s/2}\right) \le \epsilon_2 n^{-2}
\end{aligned}
$$

for any $0 < s < 1$, where $M_0 = \sup_{x\in R^p}\omega_0(x)$, ϵ_0, ϵ_1, ϵ_2 denote three positive constants, and P_{X_i} and E_{X_i} denote the conditional probability and expectation given X_i, respectively. This leads to

$$
\begin{aligned}
&\max_{1\le i\le n}\left|\frac{1}{n}\sum_{j=1}^{n}\widetilde{h}_2(X_i,\ X_j)-\eta(X_i)\right| \\
&\quad\le \max_{1\le i\le n}\left|\frac{1}{n}\sum_{j=1,j\ne i}^{n}\left(\widetilde{h}_2(X_i,\ X_j)-\eta(X_i)\right)\right| + \frac{2M_0}{n} \\
&\quad= O_p(n^{-1/2+s/2}).
\end{aligned}
$$

Relying on this, a simple derivation leads to

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} V_{1i}(V_{3i}-\Theta_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{1}{n}\sum_{j=1}^{n}\widetilde{h}_2(X_i,\ X_j)-\eta(X_i)\right)\times\left(\frac{(n-1)a_i}{\sum_{l=1,l\ne i}^{n}a_l}(Z_i-\Theta_0)\right)\right\} \\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(n-1)a_i}{\sum_{l=1,l\ne i}^{n}a_l}\eta(X_i)(Z_i-\Theta_0)\right)+o_p(1) \\
&\xrightarrow{p} \kappa_3\sigma_{12},\quad \text{as } n\to+\infty.
\end{aligned}
$$

Similarly, as $n \to +\infty$, we have

$$
\frac{1}{n}\sum_{i=1}^{n}(V_{2i}-\Theta_0)(V_{3i}-\Theta_0) \xrightarrow{p} \kappa_3(\sigma_{12}+\sigma_{22}),
$$

and

$$
\frac{1}{n}\sum_{i=1}^{n}(V_{3i}-\Theta_0)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(n-1)a_i}{\sum_{l=1,l\ne i}^{n}a_l}(Z_i-\Theta_0)\right)^2 + o_p(1) \xrightarrow{p} \kappa_4\sigma_{22}.
$$

Here the existence of κ_3 and κ_4 can be proved by following a similar argument to those of κ_1 and κ_2. Thus, we omit the details. This completes the proof. □

Proof of Theorem 1 By using Lagrange multiplier method, we obtain

$$
-2\log \ell_n(\Theta_0) = 2\sum_{i=1}^{n}\log(1+\lambda^\top(\mathbf{V}_i-\Theta_0)),
$$

where λ is the solution to $\sum_{i=1}^{n}(\mathbf{V}_i-\Theta_0)/(1+\lambda^\top(\mathbf{V}_i-\Theta_0)) = 0$. Note that $\max_{1\le i\le n}\|\mathbf{V}_i-\Theta_0\|$ is bounded, a standard argument to that of Owen (1988) can lead to $\|\lambda\| = O_p(n^{-1/2})$, and in turn $-2\log \ell_n(\Theta_0) \xrightarrow{\mathcal{L}} Q^\top\Sigma_2^{-1}Q$ based on Lemmas 2-3. This completes the proof of Theorem 1. □

Proof of Theorem 2 Let

$$\widehat{\Sigma}_1 = \begin{pmatrix} \widehat{\sigma}_{11} & \widehat{\sigma}_{11} + \widehat{\sigma}_{12} & \widehat{\kappa}_1 \widehat{\sigma}_{12} \\ \widehat{\sigma}_{11} + \widehat{\sigma}_{12} & \widehat{\sigma}_{11} + 2\widehat{\sigma}_{12} + \widehat{\sigma}_{22} & \widehat{\kappa}_1 (\widehat{\sigma}_{12} + \widehat{\sigma}_{22}) \\ \widehat{\kappa}_1 \widehat{\sigma}_{12} & \widehat{\kappa}_1 (\widehat{\sigma}_{12} + \widehat{\sigma}_{22}) & \widehat{\kappa}_2 \widehat{\sigma}_{22} \end{pmatrix},$$

and

$$\widehat{\Sigma}_2 = \begin{pmatrix} \widehat{\sigma}_{11} & \widehat{\sigma}_{11} + \widehat{\sigma}_{12} & \widehat{\kappa}_3 \widehat{\sigma}_{12} \\ \widehat{\sigma}_{11} + \widehat{\sigma}_{12} & \widehat{\sigma}_{11} + 2\widehat{\sigma}_{12} + \widehat{\sigma}_{22} & \widehat{\kappa}_3 (\widehat{\sigma}_{12} + \widehat{\sigma}_{22}) \\ \widehat{\kappa}_3 \widehat{\sigma}_{12} & \widehat{\kappa}_3 (\widehat{\sigma}_{12} + \widehat{\sigma}_{22}) & \widehat{\kappa}_4 \widehat{\sigma}_{22} \end{pmatrix}, \tag{16}$$

where $\widehat{\sigma}_{11} = \frac{1}{n}\sum_{i=1}^n \left(\frac{1}{n}\sum_{j=1}^n h_2(X_i, X_j)\right)^2$, $\widehat{\sigma}_{12} = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n h_2(X_i, X_j)(Z_i - \Theta_0)$, $\widehat{\sigma}_{22} = \frac{1}{n}\sum_{i=1}^n (Z_i - \Theta_0)^2$, $\widehat{\kappa}_1 = \frac{1}{n}\sum_{i=1}^n b_i$, $\widehat{\kappa}_2 = \frac{1}{n}\sum_{i=1}^n b_i^2$, $\widehat{\kappa}_3 = \frac{1}{n}\sum_{i=1}^n c_i$ and $\widehat{\kappa}_4 = \frac{1}{n}\sum_{i=1}^n c_i^2$. Clearly, $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$ are consistent estimators of Σ_1 and Σ_2, respectively. Based on the proof of Theorem 1, a similar derivation to that of Theorem 2 in Wang and Rao (2002) leads to

$$\begin{aligned} & -2 \times \widehat{r}(\Theta_0) \times \log \ell_n(\Theta_0) \\ & = -2 \cdot \mathrm{tr}\left(\widehat{\Sigma}_1^{-1} \widehat{\Omega}\right) / \mathrm{tr}\left(\widehat{\Sigma}_2^{-1} \widehat{\Omega}\right) \cdot \log \ell_n(\Theta_0) \\ & = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{V}_i - \Theta_0)\right)^\top \widehat{\Sigma}_1^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{V}_i - \Theta_0)\right) + o_p(1) \\ & \xrightarrow{\mathcal{L}} \chi_3^2. \end{aligned}$$

This completes the proof. □

Proof of Theorem 3 For

$$a_i = \begin{cases} 1, & \text{if } i \le \lfloor n\tau \rfloor \\ 0, & \text{otherwise} \end{cases}, \quad i = 1, 2, \ldots, n,$$

we have

$$\begin{aligned} & \max_{1 \le i \le n} \left(\frac{a_i}{\sum_{l=1}^2 a_l}\right)^2 = \frac{1}{\lfloor n\tau \rfloor^2} \to 0, \quad \text{as } n \to \infty, \\ & \lim_{n \to \infty} \sum_{i=1}^n \frac{n a_i^2}{\left(\sum_{l=1}^n a_l\right)^2} = \lim_{n \to \infty} \lfloor n\tau \rfloor \times \frac{n}{\lfloor n\tau \rfloor^2} = \frac{1}{\tau} < \infty. \end{aligned}$$

That is, Assumptions **A1** and **A2** hold. Next, for this particular case, we have

$$b_i = \frac{n^2}{\lfloor n\tau \rfloor} - (n-1)\left((n - \lfloor n\tau \rfloor)\frac{1}{\lfloor n\tau \rfloor} + (\lfloor n\tau \rfloor - 1)\frac{1}{\lfloor n\tau \rfloor - 1}\right) = \frac{n}{\lfloor n\tau \rfloor},$$

and $c_i = \frac{n-1}{\lfloor n\tau \rfloor - 1}$, if $1 \leq i \leq \lfloor n\tau \rfloor$, otherwise $b_i = 0$ and $c_i = 0$. Based on these, a tedious derivation leads to $\kappa_1 = 1$, $\kappa_2 = \frac{1}{\tau} \neq \kappa_1$, $\kappa_3 = 1$, and $\kappa_4 = \frac{1}{\tau} \neq \kappa_3$. Hence, we have $\Sigma_1 = \Sigma_2 > 0$ when $0 < \tau < 1$. The rest of the proof of this theorem is similar to that of Theorem 1. $\square$

References

Chen, B., Pan, G., Yang, Q., Zhou, W. (2015). Large dimensional empirical likelihood. *Accepted by Statistica Sinica*. doi:10.5705/ss.2013.246.

D'Agostino, R.B., Stephens, M.A. (1986). *Goodness-of-fit Technicques* (Vol. 68). New York:CRC press.

Einmahl, J. H., McKeague, I. W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli*, *9*(2), 267–290.

Feng, H., Peng, L. (2012). Jackknife empirical likelihood tests for distribution functions. *Journal of Statistical Planning and Inference*, *142*(6), 1571–1585.

Jiang, Y., Wang, S., Ge, W., Wang, X. (2011). Depth-based weighted empirical likelihood and general estimating equations. *Journal of Nonparametric Statistics*, *23*(4), 1051–1062.

Jing, B. Y., Yuan, J., Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*, *104*, 1224–1232.

Kong, L., Zuo, Y. (2010). Smooth depth contours characterize the underlying distribution. *Journal of Multivariate Analysis*, *101*(9), 2222–2226.

Liu, R. Y., Parelius, J. M., Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by Liu and Singh). *The Annals of Statistics*, *27*(3), 783–858.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*(2), 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statitics*, *18*, 90–120.

Owen, A. B. (2001). *Empirical likelihood*. London: M Chapman and Hall.

Pruitt, W. E. (1966). Summability of independent random variables. *Journal of Mathematics and Mechanics*, *15*, 769–776.

Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statitics*, *22*, 300–325.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*. New York: Wiley.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). New Jersey: Wiley.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. London: Cambridge University Press.

Wang, Q. H., Rao, J. N. K. (2002). Empirical likelihood-based inference in linear errors-in-covariables models with validation data. *Biometrika*, *89*(2), 345–358.