Bernoulli **24**(1), 2018, 449–464
DOI: 10.3150/16-BEJ884

Jackknife empirical likelihood goodness-of-fit tests for U-statistics based general estimating equations

HANXIANG PENG* and FEI TAN**

*Department of Mathematical Sciences, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202-3267, USA. E-mail: *hanxpeng@iupui.edu; **feitan@iupui.edu*

Motivated by applications to goodness of fit U-statistic testing, the jackknife empirical likelihood (JEL) for vector U-statistics is justified with two approaches and the Wilks theorems are proved. This generalizes empirical likelihood (EL) for general estimating equations (GEE's) to U-statistics based GEE's. The results are extended to allow for the use of estimated constraints and for the number of constraints to grow with the sample size. It is demonstrated that the JEL can be used to construct EL tests for moment based distribution characteristics (e.g., skewness, coefficient of variation) with less computational burden and more flexibility than the usual EL. This can be done in the U-statistic representation approach and the vector U-statistic approach which were illustrated with several examples including JEL tests for Pearson's correlation, Goodman–Kruskal's Gamma, overdisperson, U-quantiles, variance components, and the simplicial depth function. The JEL tests are asymptotically distribution free. Simulations were run to exhibit power improvement of the JEL tests with incorporation of side information.

Keywords: empirical likelihood; infinitely many constraints; Kendall's tau; linear mixed effects model; overdisperson; side information; simplicial depth; U-statistics

1. Introduction

To construct tests and confidence sets in a nonparametric setting, Owen [12] introduced the empirical likelihood approach. It enjoys many desirable properties and has been extended to various areas of statistics with tremendous accomplishments. In this article, we shall develop the theory for U-statistics based general estimating equations (UGEE's) and apply it to construct JEL tests for several important common cases.

Vector U-statistics are useful and each of many frequently used test statistics can be written as a function of vector U-statistics. UGEE's provide flexible ways to describe parameters and their corresponding statistics. See, for example, Kowalski and Tu [7], Lee [8] and Serfling [17]. Recently, Jing *et al.* [6] developed an EL theory for univariate U-statistics by exploiting jackknife pseudo values. The usual EL for a U-statistic of order 2 or higher involves nonlinearality of the probability weights in the defining maximization for the EL. This leads to unavailability of the usual explicit solutions for the weights. The technique of jackknife pseudo values for U-statistics circumvents the nonlinearality. Meanwhile, it correctly estimates the variance, so that Wilks' theorems still hold. See also Yuan *et al.* [19]. As in the case of EL for time series in Nordman and Lahiri [11], independence which justifies the definition of EL as a product of probabilities is not directly available for a U-statistic of which the summands are not independent but correlated. Jing *et al.* noticed the asymptotic independence of the jackknife pseudo values of a univariate U-statistic and defined the JEL for it.

In justifying the asymptotic independence, Jing *et al.* cited a theorem from Shi [18], who proved the asymptotic independence by an application of the zero-one law for a sequence of exchangeable random variables. Since Shi's result is not readily available as it was published in Chinese, we here present two justifications of the asymptotic independence. The first justification is based on the Hoeffding decompositions for univariate U-statistics. This naturally, in view of the Hoeffding decompositions for vector U-statistics, leads to defining the JEL for vector U-statistics. The second justification is to view the jackknife pseudo values as estimates of certain constraint functions based on which the EL is well defined, see Section 2 for details. This approach is more general than the first one and is in the spirit of EL with estimated constraints of Hjort *et al.* [5] and Peng and Schick [14,15].

After presenting the two justifications, we give Wilks' theorems for vector U-statistics and for U-statistics with a fixed or growing number of known or estimated constraints. We demonstrate that the EL tests for moment based distribution characteristics can be constructed using JEL for vector U-statistics and provide two approaches to constructing such tests. We study the use of side information to improve power of JEL tests.

Peng and Schick [14] developed the theory of the usual EL for constraints to use estimated functions and for the number of constraints to grow with sample sizes. Viewed jackknife pseudo values as estimated constraints, the JEL nicely fits into the setup of their theory. As a result, we proved our theorems by applying their theory, in particular, we proved our main Theorem 1 by applying their Theorem 6.1 while we obtained Theorem 3 by generalizing their Theorem 7.4.

The rest of the paper is structured as follows: In Section 2, the JEL is introduced with two justifications. We demonstrate in Section 3 that moment based distribution characteristics can be expressed as vector U-statistics. In Section 4, the Wilks theorems for vector U-statistics and for U-statistics with a fixed or growing number of constraints are proved. Examples are given. The asymptotic behaviors of the JEL with a growing number of estimated constraints are studied in Section 5 with an illustrative example. Section 6 reports simulations. The notation is introduced and the theorems and examples are proved in the Supplement found in Peng and Tan [16].

2. JEL for vector U-statistics

In this section, two justifications for JEL for vector U-statistics are given.

Let $(\Omega, \mathscr{A})$ be a measurable space on which P is a probability measure. Let $\mathbf{Z}$ be a random element taking values in a measurable space $(\mathcal{Z}, \mathscr{S})$ with distribution Q under P. Let $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ be independent and identically distributed (i.i.d.) as $\mathbf{Z}$. Let $h : \mathcal{Z}^m \mapsto \mathcal{R}$ be a square integrable function which is argument-symmetric. A U-statistic with kernel h of order m is defined as

$$U_n =: U_{nm}(h) = \frac{m!(n-m)!}{n!} \sum_{1 \le i_1 < \cdots < i_m \le n} h(\mathbf{Z}_{i_1}, \dots, \mathbf{Z}_{i_m}), \qquad n \ge 2.$$

Let $\delta_{\mathbf{z}}$ be the point mass at $\mathbf{z} \in \mathcal{Z}$. As in Arcones [1], we define

$$h_c^*(\mathbf{z}_1, \dots, \mathbf{z}_c) = (\delta_{\mathbf{z}_1} - P) \cdots (\delta_{\mathbf{z}_c} - P) P^{m-c} h, \qquad c = 0, 1, \dots, m,$$

where $P_1 \cdots P_c f = \int \cdots \int f(\mathbf{z}_1, \ldots, \mathbf{z}_c)\, dQ_1(\mathbf{z}_1) \cdots dQ_c(\mathbf{z}_c)$. The Hoeffding decomposition can now be stated as

$$U_n - \theta = \sum_{c=1}^{m} \binom{m}{c} U_{nc}\big(h_c^*\big), \tag{2.1}$$

where $\theta = E(h) := E(h(\mathbf{Z}_1, \ldots, \mathbf{Z}_m))$. Let $U_{n-1}^{(-j)}$ denote the U-statistic of order m based on the $n-1$ observations $\mathbf{Z}_1, \ldots, \mathbf{Z}_{j-1}, \mathbf{Z}_{j+1}, \ldots, \mathbf{Z}_n$. The jackknife pseudo values of the U-statistic $U_n(h)$ with kernel h are defined as

$$V_{nj}(h) = nU_n(h) - (n-1)U_{n-1}^{(-j)}(h), \qquad j = 1, \ldots, n.$$

For brevity, we drop h and write $V_{nj} = V_{nj}(h)$ when there is no ambiguity. Let $\tilde{f} = f - E(f)$ for integrable f. Obviously $h_1^* = \tilde{h}_1$. From (2.1), it follows

$$V_{nj} = \theta + m\tilde{h}_1(\mathbf{Z}_j) + R_{nj}, \qquad j = 1, \ldots, n, \tag{2.2}$$

where R_{nj} is the remainder given by

$$R_{nj} = \sum_{c=2}^{m} \binom{m}{c} \Big(nU_{nc}\big(h_c^*\big) - (n-1)U_{(n-1)c}^{(-j)}\big(h_c^*\big)\Big), \qquad j = 1, \ldots, n. \tag{2.3}$$

Using (2.1) and the orthogonality of $U_{nc}(h_c^*)$'s, we can prove the following with the proof given in the Supplement at Peng and Tan [16].

Lemma 1. *The jackknife pseudo values V_{nj} of $U_n(h)$ satisfy*

$$E\Big(\big(V_{nj} - \theta - m\tilde{h}_1(\mathbf{Z}_j)\big)^2\Big) = O\big(n^{-1}\big), \qquad j = 1, \ldots, n. \tag{2.4}$$

HOEFFDING DECOMPOSITION APPROACH. From (2.4), it readily follows

$$V_{nj} = \theta + m\tilde{h}_1(\mathbf{Z}_j) + O_p\big(n^{-1/2}\big), \qquad j = 1, \ldots, n. \tag{2.5}$$

This shows that each jackknife pseudo value V_{nj} depends asymptotically on $\mathbf{Z}_j$ so that V_{nj}'s are approximately *independent* for large values of n. Another nice property of the jackknife pseudo values is that V_{nj}'s satisfy

$$\frac{1}{n} \sum_{j=1}^{n} V_{nj}(h) = U_n(h). \tag{2.6}$$

Suppose now that there is available side information about $\mathbf{Z}$ given by

$$E\big(\mathbf{g}(\mathbf{Z})\big) = 0, \tag{2.7}$$

where $\mathbf{g} : \mathcal{Z} \mapsto \mathcal{R}^r$ is square-integrable. This is referred to as (r) *constraints*. In view of $E(U_n) = \theta$ and (2.6), we are now justified to introduce the JEL for the U-statistic $U_n(h)$ with side information given by (2.7) as follows:

$$\mathscr{R}_n(h, \mathbf{g}) = \sup\left\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^{n} \pi_j \operatorname{Vec}\big(\tilde{V}_{nj}(h), \mathbf{g}(\mathbf{Z}_j)\big) = 0 \right\}, \tag{2.8}$$

where $\mathscr{P}_n$ denotes the closed probability simplex in dimension n, that is,

$$\mathscr{P}_n = \big\{ \boldsymbol{\pi} = (\pi_1, \dots, \pi_n)^\top \in [0,1]^n : \pi_1 + \cdots + \pi_n = 1 \big\}.$$

Estimated constraints approach. If we replace the jackknife pseudo values $\tilde{V}_{nj}(h)$ by $m\tilde{h}_1(\mathbf{Z}_j)$ in (2.8), then the resulting supremum is $\mathscr{R}_n(mh_1, \mathbf{g})$, which is the usual well-defined empirical likelihood. Now consider estimating $m\tilde{h}_1(\mathbf{Z}_j)$ by $\tilde{V}_{nj}(h)$ and work with the estimated constraints. The resulting supremum is then (2.8). In fact, applications of Theorem 1 with $\mathbf{h} = \operatorname{Vec}(h, \mathbf{g})$ and $\mathbf{h} = \operatorname{Vec}(mh_1, \mathbf{g})$ respectively yield that the JEL $\mathscr{R}_n(h, \mathbf{g})$ and the EL $\mathscr{R}_n(mh_1, \mathbf{g})$ have the same asymptotic distribution.

Let us mention that, using the Hoeffding decompositions for vector U-statistics, the preceding definition of JEL for U-statistics with side information can be extended to the JEL for vector U-statistics. Specifically, let $h^{(k)} : \mathcal{Z}^{m_k} \mapsto \mathcal{R}$ be a kernel for $k = 1, \dots, r$. Let $E(U_{nm_k}(h^{(k)})) = \theta_k$ and $\tilde{V}_{nj}(h^{(k)}) = V_{nj}(h^{(k)}) - \theta_k$. Let $\mathbf{h} = (h^{(1)}, \dots, h^{(m_k)})^\top$, $\mathbf{U}_n(\mathbf{h}) = (U_{nm_1}(h^{(1)}), \dots, U_{nm_r}(h^{(r)}))^\top$ and $\tilde{\mathbf{V}}_{nj}(\mathbf{h}) = (\tilde{V}_{nj}(h^{(1)}), \dots, \tilde{V}_{nj}(h^{(m_k)}))^\top$. The JEL for the vector U-statistic $\mathbf{U}_n(\mathbf{h})$ is now justified to be defined by

$$\mathscr{R}_n(\mathbf{h}) = \sup\left\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^{n} \pi_j \tilde{\mathbf{V}}_{nj}(\mathbf{h}) = 0 \right\}. \tag{2.9}$$

One verifies $\tilde{\mathbf{V}}_{nj}(\mathbf{h}) = \mathbf{V}_{nj}(\tilde{\mathbf{h}})$ so that $\mathscr{R}_n(\mathbf{h}) = \mathscr{R}_n(\tilde{\mathbf{h}})$. In the below sections, we shall study the asymptotic behaviors of the preceding JEL's.

3. Two U-statistic approaches

In this section, we give two approaches to constructing JEL tests for moment based distribution characteristics. We shall illustrate the idea using two examples.

The U-statistic representation approach. Let $\mathbf{Z} = (X, Y)$ be a r.v. with finite second moment. Consider testing the null hypothesis that Pearson's correlation coefficient is equal to some specified value ρ_0. In this case, the constraint for constructing the EL is

$$\operatorname{Cov}^2(X, Y) - \rho_0^2 \operatorname{Var}(X) \operatorname{Var}(Y) = 0. \tag{3.1}$$

As this equation contains the quadratic terms $E^2(X)$, $E^2(Y)$, etc., there are no explicit formulas available for the probability weights in the EL. Thanks to the JEL, this obstacle can be overcome as explained below. Let us first mention that U-statistics are quite general. Heffernan [4] showed that a statistical functional $\theta = \theta(Q)$ of a distribution Q admits an unbiased estimator if and only if there is a function ψ of m variables such that $\theta(Q) = \int \cdots \int \psi \, dQ^m$, and derived the U-statistic as the unique MVUE of a central moment. Moment based distribution characteristics (e.g., Pearson's correlation) are functions of central moments, so that the sample versions as test statistics can be expressed as functions of U-statistics.

As $\mathbf{Z}_1, \mathbf{Z}_2$ are i.i.d. $\mathbf{Z}$,

$$E\big(X^a Y^b\big) E\big(X^c Y^d\big) = E\big(X_1^a Y_1^b X_2^c Y_2^d\big), \tag{3.2}$$

where a, b, c, d are reals for which the above expected values are defined. Using (3.2) repeatedly, we can write (3.1) as $E(\psi(\mathbf{Z}_1, \ldots, \mathbf{Z}_4)) = 0$, where $\psi(\mathbf{z}_1, \ldots, \mathbf{z}_4) = x_1 x_2 y_1 y_2 - x_1 y_2 x_3 y_3 - y_1 x_2 x_3 y_3 - \rho_0^2 (x_1^2 y_2^2 - x_1^2 y_2 y_3 - y_1^2 x_2 x_3) + (1 - \rho_0^2) x_1 x_2 y_3 y_4$. Denote by κ the symmetrized version of ψ. Then (3.1) can further be written as the U-statistic equation $E(U_{n4}(\kappa)) = 0$.

In his Theorem 3.3, Owen [12] gave a method to construct confidence regions for smooth functions of means. While Owen's method requires to solve five equations in this case, our approach only needs to solve one equation, though the jackknife pseudo values must be computed. It appears that our approach has less computational burden than Owen's. In fact, it was pursued in Li *et al.* [9] that the JEL technique can be used to reduce computational cost of EL. Moreover, as the jackknife technique turns a constraint into an equation of pseudo values, it is convenient to use our approach when there are multiple constraints. See Example 1 for more details.

The vector U-statistic approach. Consider Goodman and Kruskal's Gamma: $\gamma = (\theta_1 - \theta_2)/(\theta_1 + \theta_2)$, where $\theta_1 = P((X_1 - X_2)(Y_1 - Y_2) > 0)$ and $\theta_2 = P((X_1 - X_2)(Y_1 - Y_2) < 0)$. Associated with it can a vector U-statistic $\mathbf{U}_{n2}(\mathbf{h})$ of order 2 be constructed with the kernel equal to

$$\mathbf{h}(\mathbf{z}_1, \mathbf{z}_2) = \mathrm{Vec}\big(\mathbf{1}\big[(x_1 - x_2)(y_1 - y_2) > 0\big], \mathbf{1}\big[(x_1 - x_2)(y_1 - y_2) < 0\big]\big). \tag{3.3}$$

See Example 2 below for the construction of confidence set for γ.

A more general side information than (2.7) is given by

$$E\big(\mathbf{g}(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)\big) = 0, \tag{3.4}$$

where $\mathbf{g} : \mathcal{R}^m \mapsto \mathcal{R}^r$ is argument-symmetric and square-integrable. Using (3.2)-like identities and symmetrization, we can express sample moment based tests as U-statistics with kernel $\mathbf{g}$ in (3.4), see Examples 1 and 4.

4. The Wilks theorems and examples

In this section, we present the theorems for vector U-statistics and for U-statistics with side information given by a growing number of constraints and several examples.

Our first main result generalizes Owen's vector EL and Jing *et al.* [6] JEL for univariate U-statistics to vector U-statistics.

Theorem 1. *Suppose the variance–covariance matrix* $\mathrm{Var}(\mathbf{m}\mathbf{h}_1(\mathbf{Z}))$ *is finite and nonsingular. Then the JEL* $\mathscr{R}_n(\mathbf{h})$ *for a r-dimensional vector U-statistic* $\mathbf{U}_n(\mathbf{h})$ *defined in* (2.9) *converges in distribution to the chi-square distribution with r degrees of freedom, that is,* $-2\log\mathscr{R}_n(\mathbf{h}) \Rightarrow \chi^2(r)$.

Often the kernel $\mathbf{h}$ depends on a parameter $\boldsymbol{\theta} \in \Theta \subset \mathcal{R}^s$, $\mathbf{h} = \mathbf{h}(\cdot;\boldsymbol{\theta}) \in \mathcal{R}^r$. An application of Theorem 1 yields the following result.

Corollary 1. *Suppose the variance–covariance matrix* $\mathrm{Var}(\mathbf{m}\mathbf{h}_1(\mathbf{Z};\boldsymbol{\theta}_0))$ *is finite and nonsingular. If* $\boldsymbol{\theta}_0$ *satisfies* $E(\mathbf{h}(\mathbf{Z}_1,\dots,\mathbf{Z}_m;\boldsymbol{\theta}_0)) = 0$, *then the JEL* $\mathscr{R}_n(\mathbf{h}(;\boldsymbol{\theta}_0))$ *satisfies* $-2\log\mathscr{R}_n(\mathbf{h}(;\boldsymbol{\theta}_0)) \Rightarrow \chi^2(r)$.

Let $\boldsymbol{\tau}$ be a measurable function on $\boldsymbol{\Theta}$. Under the assumptions of Corollary 1 a $100(1-\alpha)\%$ confidence set for $\boldsymbol{\tau}(\boldsymbol{\theta})$ is

$$\left\{\boldsymbol{\tau}(\boldsymbol{\theta}) : -2\log\mathscr{R}_n\big(\mathbf{h}(;\boldsymbol{\theta})\big) \le \chi^2_{1-\alpha}(r), \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}. \tag{4.1}$$

A special case of Theorem 1 is when side information is given by the usual equation (2.7). This is the JEL $\mathscr{R}_n(\mathbf{h},\mathbf{g})$ which generalizes (2.8) from a scalar kernel h to a vector kernel $\mathbf{h}$. Recall the definition of the EL ratio $\mathscr{R}_n(\mathbf{h},\mathbf{g})$. In the presence of side information (2.7), we naturally look at the EL ratio, $\mathcal{R}_n(\mathbf{h},\mathbf{g}) = \mathscr{R}_n(\mathbf{h},\mathbf{g})/\mathscr{R}_n(\mathbf{g})$, as $\mathcal{R}_n(\mathbf{h},\mathbf{g}) \ge \mathscr{R}_n(\mathbf{h},\mathbf{g})$. Using Theorem 1, Cochran's theorem and the standard proof for parametric likelihood ratios, one shows the following result.

Corollary 2. *Let* $\mathbf{h}$ *be a vector kernel and* $\mathbf{m}$ *be a vector of positive integers, both in* $\mathcal{R}^s$. *Assume* $\mathbf{g} : \mathcal{Z} \mapsto \mathcal{R}^r$ *satisfies* (2.7). *Suppose* $\mathrm{Cov}(\mathrm{Vec}(\mathbf{m}\mathbf{h}_1,\mathbf{g})(\mathbf{Z}))$ *is finite and nonsingular. Then*

$$-2\log\mathscr{R}_n(\mathbf{h},\mathbf{g}) \Rightarrow \chi^2(r+s). \tag{4.2}$$

Hence,

$$-2\log\mathcal{R}_n(\mathbf{h},\mathbf{g}) = -2\log\mathscr{R}_n(\mathbf{h},\mathbf{g}) + 2\log\mathscr{R}_n(\mathbf{g}) \Rightarrow \chi^2(s). \tag{4.3}$$

We now study the JEL when the dimension $r = r_n \to \infty$ of a constraint function $\mathbf{g} = \mathbf{g}_n$ as $n \to \infty$. With (2.7) as side information, the JEL for U-statistic $U_n(h)$ is $\mathscr{R}_n(h,\mathbf{g}_n)$. We would establish under suitable conditions,

$$\frac{-2\log\mathscr{R}_n(h,\mathbf{g}_n) - (r_n+1)}{\sqrt{2(r_n+1)}} \Rightarrow \mathscr{N}(0,1), \quad r_n \to \infty. \tag{4.4}$$

One may interpret (4.4) as approximately $-2\log\mathscr{R}_n(h,\mathbf{g}_n)$ distributed as $\chi^2(r_n+1)$ whence $-2\log\mathcal{R}_n(h,\mathbf{g}_n) = -2\log\mathscr{R}_n(h,\mathbf{g}_n) + 2\log\mathscr{R}_n(\mathbf{g}_n)$ is distributed as $\chi^2(1)$. See related work in Chen *et al.* [3], Hjort *et al.* [5] and Peng and Schick [14,15].

To establish (4.4), introduce $\lambda_{\min}(\mathbb{M}_n)$ ($\lambda_{\max}(\mathbb{M}_n)$) the smallest (largest) eigenvalue of a $r_n \times r_n$ symmetric matrix $\mathbb{M}_n$. Following Peng and Schick [14], a sequence of $r_n \times r_n$ dispersion matrices $\mathbb{M}_n$ is *regular* if

$$0 < \inf_n \lambda_{\min}(\mathbb{M}_n) \le \sup_n \lambda_{\max}(\mathbb{M}_n) < \infty. \tag{R}$$

A sequence of vector functions $\{\mathbf{v}_n\}$ is *Lindeberg* if for every $\epsilon > 0$,

$$\int \|\mathbf{v}_n\|^2 \mathbf{1}\big[\|\mathbf{v}_n\| > \epsilon\sqrt{n}\big]\, dQ \to 0. \tag{L}$$

Useful properties for Lindeberg sequences can be found in Peng and Schick [14,15]. For matrices $\mathbb{A}$, $\mathbb{C}$ and $\mathbb{M}$ of compatible dimensions, let

$$\mathscr{W}(\mathbb{A}, \mathbb{C}, \mathbb{M}) = \begin{pmatrix} \mathbb{A} & \mathbb{C}^\top \\ \mathbb{C} & \mathbb{M} \end{pmatrix} \quad \text{and} \tag{4.5}$$

$$\mathbf{C}_n = E\big(mh_1(\mathbf{Z})\mathbf{g}_n(\mathbf{Z})\big), \qquad \mathbb{W}_n = E\big(\mathbf{g}_n^{\otimes 2}(\mathbf{Z})\big). \tag{4.6}$$

As a special case of Theorem 3 below, the distribution of $-2\log\mathscr{R}_n(h, \mathbf{g}_n)$ is approximately $\chi^2(r_n + 1)$ as stated next. This generalizes Theorem 1 from a fixed number of constraints to a growing number.

Theorem 2. *Suppose $\mathbf{g}_n : \mathcal{Z} \mapsto \mathcal{R}^{r_n}$ satisfies (2.7). Suppose further the sequences $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ satisfy (L) such that the sequence of matrices $\mathscr{W}(\mathrm{Var}(mh_1(Z_1)), \mathbf{C}_n, \mathbb{W}_n)$ satisfies (R). Then (4.4) holds as both r_n and n tend to infinity such that $r_n = o(n^{1/2})$.*

We now apply Theorems 1–2 to derive the JEL tests and confidence sets for a number of frequently used moment based distribution characteristics.

Example 1. TESTING THE EQUALITY OF TWO MEANS. Consider testing the null hypothesis of the equality $E(X) = E(Y)$ of a r.v. $\mathbf{Z} = (X, Y)$ in the presence of side information of $\mathrm{Var}(X) = \mathrm{Var}(Y)$ and $\mathrm{Cov}(X, Y) = 0$. Let $h^{(1)}(\mathbf{z}) = x - y$. Using the U-STATISTICS REPRESENTATION APPROACH given in Section 3, the equality of variances can be expressed by the UGEE $E(U_n(h^{(2)})) = 0$ with kernel $h^{(2)}(\mathbf{z}_1, \mathbf{z}_2) = 2^{-1}(x_1 - x_2)^2 - 2^{-1}(y_1 - y_2)^2$. Let $h^{(3)}$ be the kernel κ given in Section 3 and $\mathbf{h} = \mathrm{Vec}(h^{(1)}, h^{(2)}, h^{(3)})$. Then the test can be formulated by the JEL $\mathscr{R}_n(\mathbf{h})$ for the vector U-statistic $\mathbf{U}_{n4}(\mathbf{h})$ of order 4. By Corollary 2 and Theorem 1, two JEL tests of asymptotic size α are given by

$$T_1 = \mathbf{1}\big[-2\log\mathcal{R}_n(\mathbf{h}) > \chi^2_{1-\alpha}(1)\big], \qquad T_2 = \mathbf{1}\big[-2\log\mathscr{R}_n(\mathbf{h}) > \chi^2_{1-\alpha}(3)\big],$$

provided that $\mathrm{Cov}(\mathbf{h}_1(\mathbf{Z}))$ is nonsingular. A simulation study was conducted based on this example, see Table 1.

Example 2. TESTING GOODMAN AND KRUSKAL'S GAMMA. The Gamma induces the vector U-statistic $\mathbf{U}_n(\mathbf{h})$ with kernel $\mathbf{h}$ given in (3.3) with $E(\mathbf{h}(\mathbf{Z}_1, \mathbf{Z}_2)) = \boldsymbol{\theta} = (\theta_1, \theta_2)^\top$. The JEL with side information given by (2.7) is $\mathscr{R}_n(\mathbf{h}(;\boldsymbol{\theta}), \mathbf{g})$, where $\mathbf{h}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = \mathbf{h}(\mathbf{z}_1, \mathbf{z}_2) - \boldsymbol{\theta}$. Let $\mathbf{h}_1(\mathbf{z}) = E(\mathbf{h}(\mathbf{Z}_1, \mathbf{Z}_2)|\mathbf{Z}_1 = \mathbf{z})$. By Corollary 2, $\mathbf{1}[-2\log\mathcal{R}_n(\boldsymbol{\theta}_0, \mathbf{g}) > \chi^2_{1-\alpha}(2)]$ is an asymptotic test of size α for testing the null $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ provided that the matrix $\mathscr{W}(\mathrm{Var}(2\mathbf{h}_1(\mathbf{Z})), \mathbf{C}, \mathrm{Var}(\mathbf{g}(\mathbf{Z})))$ is nonsingular, where $\mathbf{C} = E(\mathbf{h}_1(\mathbf{Z}) \otimes \mathbf{g}(\mathbf{Z}))$. Thus the rejection of the null at the α level of significance implies the null $\gamma = \gamma_0 = (\theta_{10} - \theta_{20})/(\theta_{10} + \theta_{20})$ must be rejected at the same level.

Table 1. Simulated power of the UJEL tests for the equality of two means with side information of equal variances and zero correlation at the nominal level 0.05. Here d, ANT and UJEr denote respectively the mean difference, asymptotic-normal test and UJEL tests with r constraints

n	d	T	ANT	UJEL0	UJEL1	UJEL2	ANT	UJEL0	UJEL1	UJEL2
				Exponential				Lognormal		
150	0	T_1	0.0535	0.0550	0.0545	0.0605	0.0470	0.0665	0.0870	0.0935
		T_2	0.0535	0.0550	0.0870	0.1050	0.0470	0.0665	0.1600	0.1980
	0.10	T_1	0.1470	0.1485	0.2815	0.2965	0.0675	0.0855	0.1630	0.1645
		T_2	0.1470	0.1485	0.2530	0.2420	0.0675	0.0855	0.2215	0.2475
	0.15	T_1	0.2520	0.2500	0.5090	0.5195	0.0995	0.1230	0.2075	0.2135
		T_2	0.2520	0.2500	0.4260	0.4215	0.0995	0.1230	0.2625	0.3055
200	0	T_1	0.0480	0.0525	0.0510	0.0580	0.0475	0.0580	0.0815	0.0890
		T_2	0.0480	0.0525	0.0810	0.1025	0.0475	0.0580	0.1380	0.1905
	0.10	T_1	0.1655	0.1690	0.3305	0.3420	0.0805	0.0950	0.1610	0.1760
		T_2	0.1655	0.1690	0.2800	0.2580	0.0805	0.0950	0.2245	0.2465
	0.15	T_1	0.3140	0.3145	0.6075	0.6250	0.1195	0.1250	0.2595	0.2790
		T_2	0.3140	0.3145	0.5395	0.5005	0.1195	0.1250	0.2855	0.3100
				Poisson				Binomial		
150	0	T_1	0.0490	0.0515	0.0525	0.0535	0.0475	0.0500	0.0440	0.0440
		T_2	0.0490	0.0515	0.0625	0.0685	0.0475	0.0500	0.0465	0.0635
	0.10	T_1	0.1510	0.1540	0.1820	0.1895	0.0920	0.0945	0.0960	0.1040
		T_2	0.1510	0.1540	0.1585	0.1495	0.0920	0.0945	0.0870	0.0845
	0.15	T_1	0.2530	0.2560	0.3705	0.3765	0.1290	0.1290	0.1360	0.1440
		T_2	0.2530	0.2560	0.3130	0.2870	0.1290	0.1290	0.1095	0.1055
200	0	T_1	0.0490	0.0500	0.0515	0.0520	0.0500	0.0500	0.0550	0.0605
		T_2	0.0490	0.0500	0.0590	0.0615	0.0500	0.0500	0.0570	0.0620
	0.10	T_1	0.1800	0.1825	0.2525	0.2535	0.1120	0.1145	0.1195	0.1230
		T_2	0.1800	0.1825	0.2035	0.1865	0.1120	0.1145	0.1020	0.0960
	0.15	T_1	0.3155	0.3220	0.4560	0.4625	0.1715	0.1720	0.1830	0.1870
		T_2	0.3155	0.3220	0.3760	0.3285	0.1715	0.1720	0.1415	0.1300

This example illustrates the VECTOR U-STATISTICS APPROACH given in Section 3. Similarly to the confidence set (4.1) for the JEL ratio $\mathcal{R}_n(\boldsymbol{\theta}, \mathbf{g})$, a $100(1-\alpha)\%$ confidence set for γ is

$$\left\{(\theta_1 - \theta_2)/(\theta_1 + \theta_2) : -2\log\mathcal{R}_n(\boldsymbol{\theta}, \mathbf{g}) \le \chi^2_{1-\alpha}(2)\right\}.$$

Example 3. JOINT CONFIDENCE SETS FOR VARIANCE COMPONENTS. JEL tests for linear mixed effects models, binomial/Poisson overdisperson and inflated Poisson models can be formulated using UGEE's (see, e.g., Kowalski and Tu [7] for the UGEE's of these models). Consider now the balanced one-way random effect model, in which the response Y_{ij}, random effect u_i and error ϵ_{ij} satisfy

$$Y_{ij} = \mu + u_i + \epsilon_{ij}, \qquad i = 1, \ldots, n,\ j = 1, \ldots, J\ (J \ge 2), \tag{4.7}$$

where μ is the mean, ϵ_{ij}'s are i.i.d., u_i's are i.i.d., and ϵ_{ij}'s and u_i's are independent and have mean zero and finite fourth moments. Let $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_{ij})$ and $\sigma_u^2 = \mathrm{Var}(u_j)$. The commonly used confidence regions for the variances heavily depend on normality assumption of the model. Here we relax normality and use the JEL to give confidence sets. Following Arvesen [2], put $\mathbf{X}_i = \mathrm{Vec}(Y_{i\cdot}, (J-1)^{-1}\sum_{j=1}^J (Y_{ij} - Y_{i\cdot})^2)$, $i = 1, \dots, n$, where $A_{i\cdot} = J^{-1}\sum_{j=1}^J A_{ij}$ denotes the average of A_{ij} over j. Clearly $\mathbf{X}_1, \dots, \mathbf{X}_n$ are i.i.d. Set $\mathbf{h} = (h^{(1)}, h^{(2)})^\top$ where $h^{(1)}(\mathbf{X}_i, \mathbf{X}_{i'}) = 2^{-1}(\kappa(\mathbf{X}_i) + \kappa(\mathbf{X}_{i'}))$ and $h^{(2)}(\mathbf{X}_i, \mathbf{X}_{i'}) = 2^{-1}(Y_{i\cdot} - Y_{i'\cdot})^2$ with $\kappa(\mathbf{X}_i) = (J-1)^{-1}\sum_{j=1}^J (Y_{ij} - Y_{i\cdot})^2$. One finds

$$E\big(h^{(1)}(\mathbf{X}_1, \mathbf{X}_2)\big) = \sigma_\epsilon^2, \qquad E\big(h^{(2)}(\mathbf{X}_1, \mathbf{X}_2)\big) = \sigma^2 := \sigma_u^2 + J^{-1}\sigma_\epsilon^2. \tag{4.8}$$

Hence, the vector U-statistic $\mathbf{U}_n(\mathbf{h}) = \mathrm{Vec}(U_n(h^{(1)}), U_n(h^{(2)}))$ is an unbiased estimate of $\boldsymbol{\theta} = (\sigma_\epsilon^2, \sigma^2)^\top$. Let $\mathbf{h}(\mathbf{X}_1, \mathbf{X}_2; \boldsymbol{\theta}) = \mathbf{h}(\mathbf{X}_1, \mathbf{X}_2) - \boldsymbol{\theta}$. The JEL for the vector U-statistic $\mathbf{U}_n(\mathbf{h}(;\boldsymbol{\theta}))$ is $\mathscr{R}_n(\boldsymbol{\theta}) = \mathscr{R}_n(\mathbf{h}(;\boldsymbol{\theta})), \boldsymbol{\theta} \in [0, \infty)^2$. Let $\mathbf{h}_1(\mathbf{x}_1; \boldsymbol{\theta}) = E(\mathbf{h}(\mathbf{x}_1, \mathbf{X}_2; \boldsymbol{\theta}))$. By Theorem 1, if $\mathrm{Var}(\mathbf{h}_1(\mathbf{X}_1; \boldsymbol{\theta}_0))$ is nonsingular then a joint confidence set for $\boldsymbol{\theta}$ at the level of $1 - \alpha$ is given by $\{\boldsymbol{\theta} \in \mathcal{R}^+ \times \mathcal{R}^+ : -2\log\mathscr{R}_n(\boldsymbol{\theta}) \le \chi^2_{1-\alpha}(2)\}$. A confidence set for $\vartheta = (\sigma_\epsilon^2, \sigma_u^2)^\top$ can be obtained by the transform $\vartheta_1 = \theta_1, \vartheta_2 = \theta_2 - \theta_1/J$. A confidence set for σ_u^2 can be obtained by setting $J \to \infty$.

TESTING THE RANDOM EFFECT. Let $\psi = h^{(2)} - J^{-1}h^{(1)}$. Clearly ψ is argument-symmetric, By (4.8), $E(\psi(\mathbf{X}_1, \mathbf{X}_2)) = \sigma_u^2$. Let $\psi(\mathbf{X}_1, \mathbf{X}_2; \sigma_u^2) = \psi(\mathbf{X}_1, \mathbf{X}_2) - \sigma_u^2$. This suggests to look at the JEL $\mathscr{R}_n(\sigma_u^2) = \mathscr{R}_n(\psi(;\sigma_u^2)), \sigma_u^2 \in [0, \infty)$. Let $\psi_1(\mathbf{x}_1; \sigma_u^2) = E(\psi(\mathbf{x}_1, \mathbf{X}_2; \sigma_u^2))$. By Theorem 1, if $\mathrm{Var}(\psi_1(\mathbf{X}; 0))$ is nonsingular then an asymptotic test of size α for the null $H_0 : \sigma_u^2 = 0$ is $\mathbf{1}[-2\log\mathscr{R}_n(0) > \chi^2_{1-\alpha}(1)]$.

Example 4. TESTING U-QUANTILES. The theory of U-quantiles provides a unified treatment of several commonly used statistics, see Arcones [1]. Let $\kappa : \mathcal{Z}^m \mapsto \mathcal{R}$ be argument-symmetric. Associated with κ there induces the cumulative distribution function (c.d.f.) $H(t) = P(\kappa(Z_1, \dots, Z_m) \le t)$. The MVUE $H_{nm}(t)$ of $H(t)$ is the U-statistic $H_{nm}(t) = U_{nm}(h(;t))$ with kernel $h(z_1, \dots, z_m; t) = \mathbf{1}[\kappa(z_1, \dots, z_m) \le t]$. The U-quantiles include the Hodges–Lehmann median, Gini's mean difference, Theil's slope estimator, and Kendall's tau, corresponding to the U-quantiles with $p_0 = 1/2$ and the kernels $\kappa(z_1, z_2) = 2^{-1}(z_1 + z_2)$, $|z_1 - z_2|$, $(y_1 - y_2)/(x_1 - x_2)$, and $(x_1 - x_2)(y_1 - y_2)$ respectively. Consider testing the null that the pth U-quantile q is equal to a specified value q_0 for some p_0, i.e. $H_0 : q = q_0$. Assume there is available the side information that the coefficient of variation is constant: $\sigma/\mu = c_0$ with $\mu = E(Z)$ and $\sigma^2 = \mathrm{Var}(Z)$. Using the U-STATISTICS REPRESENTATION APPROACH in Section 3, this can be described by

$$E\big(U_{n2}(\kappa_1)\big) = 0 \qquad \text{with } \kappa_1(z_1, z_2) = (1/2)\big(z_1^2 + z_2^2\big) - \big(1 + c_0^2\big)z_1 z_2.$$

One has the JEL $\mathscr{R}_n(\kappa_1)$ for $U_{n2}(\kappa_1)$. Let $\mathbf{h}(z_1, \dots, z_m; t) = \mathrm{Vec}(\mathbf{1}[\kappa(z_1, \dots, z_m) \le t], \kappa_1(z_1, z_2))$ and $\mathbf{h}_1(z_1; t) = E(\mathbf{h}(z_1, Z_2, \dots, Z_m; t))$. By Corollaries 1–2 the JEL test of size α for H_0 is $\mathbf{1}[-2\log\mathcal{R}_n(q_0) > \chi^2_{1-\alpha}(1)]$ provided that $\mathrm{Var}(\mathbf{h}_1(Z_1; q_0))$ is nonsingular.

Example 5. Testing the simplicial depth. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. r.v.'s on $\mathcal{R}^m$. Liu [10] defined the *simplicial depth function* (SDF) $D(\mathbf{x})$ of a point $\mathbf{x} \in \mathcal{R}^m$ w.r.t. a distribution as $D(\mathbf{x}) = P(\mathbf{x} \in \Delta(\mathbf{X}_1, \ldots, \mathbf{X}_{m+1}))$, where $\Delta(\mathbf{X}_1, \ldots, \mathbf{X}_{m+1})$ is the random simplex with vertices $\mathbf{X}_1, \ldots, \mathbf{X}_{m+1}$. $D(\mathbf{x})$ can be estimated by the sample SDF $D_n(\mathbf{x})$, the U-statistic $D_n(\mathbf{x}) = U_{n,m+1}(h(;\mathbf{x}))$ with kernel $h(\mathbf{x}_1, \ldots, \mathbf{x}_{m+1}; \mathbf{x}) = \mathbf{1}[\mathbf{x} \in \Delta(\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{m+1}})]$. When side information about the underlying distribution is available, tests based on $D_n(\mathbf{x})$ do not utilize it. We can apply the JEL for vector U-statistics to use such information to improve power. For a fixed number of constraints, we can construct the JEL with side information as in Example 2. Often we have partial information about the joint distribution, for example, one marginal is known. Such information is equivalent to an infinite number of constraints. Specifically, suppose the c.d.f. of the first component X_1 of $\mathbf{X} = (X_1, \ldots, X_m)^\top$ is known and equal to F_{10}. Then

$$\int a_k \, dF_{10} = 0, \qquad k = 1, 2, \ldots, \tag{4.9}$$

where a_k is an orthonormal basis of $L_{2,0}(F_{10})$. Assume F_{10} is continuous. This allows us to take $a_k = \phi_k(F_{10}), k = 1, 2, \ldots,$ where ϕ_k is the trigonometric basis of $L_{2,0}(\mathscr{U})$ with $\mathscr{U}$ the uniform measure on [0, 1] given by

$$\phi_k(t) = \sqrt{2}\cos(k\pi t), \qquad t \in [0, 1], k = 1, 2, \ldots. \tag{4.10}$$

This suggests us to using the first r_n equations in (4.9) as constraints and construct the JEL for the U-statistic $D_n(\mathbf{x})$ with side information as follows:

$$\begin{aligned}\mathscr{R}_n(D, F_{10}) = \sup\Bigg\{ &\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j(V_{nj} - D) = 0, \\ &\sum_{j=1}^n \pi_j \phi_k\big(F_{10}(X_{1j})\big) = 0, k = 1, \ldots, r_n \Bigg\}, \qquad D \geq 0.\end{aligned}$$

Assume $m \geq 2$ and at least one of the components $X_2, \ldots, X_m$ is nondegenerate, i.e. $P(X_d = c) < 1$ for some $d \geq 2$ and any constant c. Then by Theorem 2 for fixed $D_0 \geq 0$ as r_n and n tend to infinity such that $r_n^3/n \to 0$,

$$\big(-2\log \mathscr{R}_n(D_0, F_{10}) - (r_n + 1)\big)/\sqrt{2(r_n + 1)} \Rightarrow \mathscr{N}(0, 1). \tag{4.11}$$

The details can be found in the Supplement at Peng and Tan [16].

5. Asymptotic behaviors of the JEL with a growing number of estimated constraints

In this section, we shall study the case that the kernel h is known but the constraint $\mathbf{g}_n$ must be estimated by some measurable function $\hat{\mathbf{g}}_n$. We allow the number of constraints to grow with the sample size. Recall $\mathbf{C}_n, \mathbb{W}_n$ in (4.6) and set $\hat{\mathbb{W}}_n = n^{-1}\sum_{j=1}^n \hat{\mathbf{g}}_n(\mathbf{Z}_j)^{\otimes 2}$. Generalizing Theorem 7.4 of Peng and Schick [14], we have the following result with the proof given in the Supplement at Peng and Tan [16].

Theorem 3. *Suppose $r_n\tilde{h}_1$ satisfies (L). Suppose $\hat{\mathbf{g}}_n$ is an estimator of $\mathbf{g}_n$ such that*

$$r_n \max_{1\le j\le n}\|\hat{\mathbf{g}}_n(\mathbf{Z}_j)\| = o_p\big(n^{1/2}\big), \tag{5.1}$$

$$\left\|\frac{1}{n}\sum_{j=1}^n \tilde{V}_{nj}\hat{\mathbf{g}}_n(\mathbf{Z}_j) - \mathbf{C}_n\right\| = o_p\big(r_n^{-1/2}\big), \qquad |\hat{\mathbb{W}}_n - \mathbb{W}_n|_o = o_p\big(r_n^{-1/2}\big) \tag{5.2}$$

for which $\mathscr{W}_n := \mathscr{W}(m^2\operatorname{Var}(\tilde{h}_1(\mathbf{Z})), \mathbf{C}_n, \mathbb{W}_n)$ satisfies (R), and that

$$\frac{1}{n}\sum_{j=1}^n \hat{\mathbf{g}}_n(\mathbf{Z}_j) = \frac{1}{n}\sum_{j=1}^n \mathbf{u}_n(\mathbf{Z}_j) + o_p\big(n^{-1/2}\big) \tag{5.3}$$

for some measurable function $\mathbf{u}_n : \mathcal{Z} \mapsto \mathcal{R}^{r_n}$ such that $\int \mathbf{u}_n\, dQ = 0$ and $\|\mathbf{u}_n\|$ satisfies (L). Assume further the dispersion matrix $\mathbb{U}_n$ of $\mathscr{W}_n^{-1/2}\mathbf{v}_n(\mathbf{Z})$,

$$\mathbb{U}_n = \mathscr{W}_n^{-1/2}\int \mathbf{v}_n\mathbf{v}_n^\top\, dQ\,\mathscr{W}_n^{-1/2} \qquad \textit{with } \mathbf{v}_n = \operatorname{Vec}(m\tilde{h}_1, \mathbf{u}_n) \tag{5.4}$$

satisfies $|\mathbb{U}_n|_o = O(1)$ and $r_n/\operatorname{trace}(\mathbb{U}_n^2) = O(1)$. Then as $r_n \to \infty$ but $r_n/n^{1/2} \to 0$,

$$\big(-2\log\mathscr{R}_n(h, \hat{\mathbf{g}}_n) - \operatorname{trace}(\mathbb{U}_n)\big)\Big/\sqrt{2\operatorname{trace}\big(\mathbb{U}_n^2\big)} \Rightarrow \mathscr{N}(0,1).$$

Example 6. JOINT CONFIDENCE SETS. Consider constructing a confidence set for $\boldsymbol{\theta} = (\mu, \sigma_u^2)^\top$ in the model (4.7). Let us motivate a U-statistic as test for σ_u^2. Note first that

$$\begin{aligned} E\big((\epsilon_{1j} - \epsilon_{2j})(\epsilon_{1j'} - \epsilon_{2j'})\big) &= 0, \qquad 1 \le j < j' \le J,\\ E\big((u_1 - u_2)(\epsilon_{1j'} - \epsilon_{2j'})\big) &= E\big((u_1 - u_2)(\epsilon_{1j} - \epsilon_{2j})\big) = 0, \end{aligned}$$

so that $E((Y_{1j} - Y_{2j})(Y_{1j'} - Y_{2j'})) = 2\sigma_u^2, 1 \le j < j' \le J$. Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})^\top, i = 1, \dots, n$. Clearly they are i.i.d. Let

$$h(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{2}{J(J-1)}\sum_{1\le j<j'\le J} 2^{-1}(Y_{1j} - Y_{2j})(Y_{1j'} - Y_{2j'}).$$

Then h is a kernel with $E(h(\mathbf{Y}_1, \mathbf{Y}_2)) = \sigma_u^2$. Thus an unbiased estimator of σ_u^2 based on all the observations is the U-statistic $U_{n2}(h)$ of order 2.

Normality is usually assumed for both u and ϵ. Here we shall relax normality to symmetry of the sums $\varepsilon_i = u_i + \epsilon_{i\cdot}, i = 1, \dots, n$ about zero and employ the JEL to incorporate the symmetry assumption as side information to improve power. Notice that under this (4.7) is the usual symmetric location model. Let F be the c.d.f. of ε_1 and $L_{2,0}(F, \text{odd}) \subset L_{2,0}(F)$ consisting of the odd functions. Assume F is continuous. Symmetry of ε_1 implies

$$E\big(a_k(\varepsilon_1)\big) = E\big(a_k(Y_{1\cdot} - \mu_0)\big) = 0, \qquad k = 1, 2, \dots, \tag{5.5}$$

where a_k's is an orthonormal basis of $L_{2,0}(F, \text{odd})$. This also implies that $2F(t) - 1$ is an odd function. As $L_{2,0}(\mathcal{U}, \text{odd})$ has the orthonormal basis

$$\psi_k(s) = \sin(k\pi s), \qquad s \in [-1, 1], k = 1, 2, \dots, \tag{5.6}$$

where $\mathcal{U}$ is the uniform measure on $[-1, 1]$, the composites $\psi_k(2F(t) - 1)$ is a basis of $L_{2,0}(F, \text{odd})$. This justifies that we can take $a_k = \psi_k(2F - 1)$. But F is unknown, we estimate it using $\varepsilon_i = Y_{i\cdot} - \mu_0$ by the symmetrized empirical c.d.f. $\mathbb{F}_{\mu_0}(t)$, where

$$\mathbb{F}_\mu(t) = \frac{1}{n}\sum_{i=1}^n \frac{1}{2}\big(\mathbf{1}[Y_{i\cdot} - \mu \le t] + \mathbf{1}\big[-(Y_{i\cdot} - \mu) < t\big]\big), \qquad t \in \mathcal{R}.$$

Again one verifies $2\mathbb{F}_{\mu_0}(t) - 1$ is odd. This suggests to utilizing the first r_n equations in (5.5) as side information to construct the JEL for U-statistic $U_{n2}(h)$ as follows:

$$\mathscr{R}_n\big(\mu, \sigma_u^2\big) = \sup\Bigg\{\prod_{i=1}^n n\pi_i : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{i=1}^n \pi_i\big(V_{ni}(h) - \sigma_u^2\big) = 0,$$
$$\sum_{i=1}^n \pi_i \psi_k\big(2\mathbb{F}_\mu(Y_{i\cdot} - \mu) - 1\big) = 0, k = 1, \dots, r_n\Bigg\}.$$

We shall allow r_n to grow to infinity with the sample size n such that r_n^4/n tends to zero. If further $\sigma_\epsilon > 0$, then by Theorem 3 one has

$$\big(-2\log \mathscr{R}_n\big(\mu_0, \sigma_{u0}^2\big) - (r_n + 1)\big)/\sqrt{2(r_n + 1)} \Rightarrow \mathscr{N}(0, 1). \tag{5.7}$$

This shows under the null $-2\log \mathscr{R}_n(\mu_0, \sigma_{u0}^2)$ is approximately $\chi^2(r_n + 1)$. The details can be found in the Supplement at Peng and Tan [16]. A simulation study based on this example was conducted, see Tables 2 and 3.

6. Simulations

In this section, we report some simulation results for the JEL tests for U-statistics with r constraints (UJELr, $r = 0, 1, \dots, 5$) based on Examples 1 and 6. Two UJEL tests of size 0.05 for testing the null hypothesis $H_0 : d =: E(X) - E(Y) = 0$ are as follows:

$$T_1 = \mathbf{1}\big[-2\log \mathcal{R}_n(\mathbf{h}) > \chi^2_{0.95}(1)\big], \qquad T_2 = \mathbf{1}\big[-2\log \mathscr{R}_n(\mathbf{h}) > \chi^2_{0.95}(r + 1)\big].$$

Table 2. Simulated power of the UJEL tests for $H_0 : \sigma_u^2 = 0$ in a random effect model $Y = \mu + u + \epsilon$ with side information of symmetry of $u + \epsilon$ with the size adjusted to 0.05 and data generated from (a)

J	n	σ_u^2	ρ	T	F	UJEL0	UJEL1	UJEL2	UJEL3	UJEL4	UJEL5
5	15	0.6	0.004	T_1	0.0795	0.1135	0.1240	0.1080	0.0830	0.0730	0.0725
				T_2	0.0795	0.1135	0.1255	0.0985	0.0695	0.0680	0.0690
		0.8	0.006	T_1	0.0730	0.1150	0.1210	0.1130	0.0995	0.0820	0.0640
				T_2	0.0730	0.1150	0.1170	0.0975	0.0815	0.0820	0.0780
		1.0	0.007	T_1	0.0840	0.1525	0.1695	0.1565	0.1165	0.0860	0.0675
				T_2	0.0840	0.1525	0.1655	0.1220	0.0845	0.0870	0.0710
		1.2	0.008	T_2	0.0915	0.1430	0.1640	0.1510	0.1025	0.0825	0.0835
				T_2	0.0915	0.1430	0.1540	0.1250	0.0780	0.0765	0.0810
		1.4	0.010	T_1	0.0930	0.1540	0.1720	0.1735	0.1210	0.0770	0.0745
				T_2	0.0930	0.1540	0.1710	0.1525	0.0930	0.0790	0.0850
	30	0.6	0.004	T_1	0.0570	0.0810	0.0825	0.0840	0.0860	0.0795	0.0735
				T_2	0.0570	0.0810	0.0775	0.0820	0.0700	0.0670	0.0655
		0.8	0.006	T_1	0.0615	0.0955	0.0980	0.1020	0.1015	0.1015	0.1000
				T_2	0.0615	0.0955	0.1090	0.1110	0.1075	0.1030	0.0855
		1.0	0.007	T_1	0.0560	0.0815	0.0885	0.1000	0.0960	0.0835	0.0780
				T_2	0.0560	0.0815	0.0935	0.0885	0.0855	0.0740	0.0630
		1.2	0.008	T_1	0.0705	0.1020	0.1115	0.1140	0.1165	0.1080	0.0985
				T_2	0.0705	0.1020	0.1045	0.1065	0.1020	0.0870	0.0810
		1.4	0.010	T_1	0.0545	0.1035	0.1150	0.1165	0.1190	0.1215	0.1165
				T_2	0.0545	0.1035	0.1050	0.1090	0.1050	0.1050	0.0855
	45	0.6	0.004	T_1	0.0535	0.0655	0.0635	0.0685	0.0625	0.0630	0.0640
				T_2	0.0535	0.0655	0.0690	0.0690	0.0600	0.0610	0.0660
		0.8	0.006	T_1	0.0585	0.0790	0.0835	0.0845	0.0875	0.0850	0.0825
				T_2	0.0585	0.0790	0.0830	0.0780	0.0795	0.0790	0.0840
		1.0	0.007	T_1	0.0565	0.0760	0.0820	0.0830	0.0830	0.0815	0.0865
				T_2	0.0565	0.0760	0.0815	0.0795	0.0890	0.0785	0.0755
		1.2	0.008	T_1	0.0570	0.0990	0.1020	0.1025	0.0930	0.0940	0.0885
				T_2	0.0570	0.0990	0.0940	0.0900	0.0880	0.0910	0.0800
		1.4	0.010	T_1	0.0760	0.0850	0.0855	0.0895	0.0960	0.0870	0.0810
				T_2	0.0760	0.0850	0.0830	0.0790	0.0840	0.0775	0.0755

Reported on Table 1 is the simulated power of the tests for sample sizes $n = 150, 200$, $d = 0, 0.10, 0.15$ and 2000 repetitions. For comparison, we also simulated the power of the asymptotic normality test (ANT). Here X was generated from the standard exponential, the standard log normal, the standard Poisson, and the binomial with parameters $(0.7, 10)$, while Y was generated the same way as X but with the value of d added to each generated value with the exception of the log normal in which the parameters were adjusted so that $E(X) - E(Y) = d$. Our findings are as follows: (i) Relatively larger sample sizes were needed for the significance levels of the UJEL tests to reach the nominal level than the usual EL tests. This might be due to the use of the jackknife technique. Thus for small sample sizes we suggest to perform size adjusted JEL tests.

Table 3. Simulated power of the UJEL tests for $H_0 : \sigma_u^2 = 0$ in $Y = \mu + u + \epsilon$ with side information of symmetry of $u + \epsilon$ with the size adjusted to 0.05. Data generated with $u \sim \mathscr{N}(0, \sigma_u^2)$

J	n	σ_u^2	ρ	T	F	UJEL0	UJEL1	UJEL2	UJEL3	UJEL4	UJEL5
					$\epsilon \sim 0.005\mathscr{U}(-100, -3) + 0.99\mathscr{N}(0, 1) + 0.005\mathscr{U}(3, 100)$						
3	30	0.6	0.021	T_1	0.4515	0.5410	0.5295	0.5080	0.4950	0.4915	0.4485
				T_2	0.4515	0.5410	0.4840	0.4650	0.4080	0.3550	0.2595
		0.8	0.028	T_1	0.4950	0.6215	0.6155	0.6085	0.6005	0.5755	0.5570
				T_2	0.4950	0.6215	0.5840	0.5570	0.5215	0.4720	0.3880
		1.0	0.035	T_1	0.5070	0.6440	0.6370	0.6210	0.6210	0.6195	0.6100
				T_2	0.5070	0.6440	0.6205	0.6000	0.5745	0.5280	0.4465
		1.2	0.041	T_2	0.4770	0.6390	0.6295	0.6285	0.6210	0.6200	0.6125
				T_2	0.4770	0.6390	0.6275	0.6120	0.5995	0.5680	0.5125
		1.4	0.048	T_1	0.5155	0.6860	0.6795	0.6735	0.6680	0.6630	0.6575
				T_2	0.5155	0.6860	0.6705	0.6575	0.6490	0.6190	0.5355
	60	0.6	0.021	T_1	0.3900	0.4845	0.4820	0.4790	0.4800	0.4825	0.4830
				T_2	0.3900	0.4845	0.4750	0.4805	0.4570	0.4520	0.4510
		0.8	0.028	T_1	0.3855	0.5085	0.5065	0.5045	0.5015	0.5010	0.5065
				T_2	0.3855	0.5085	0.4985	0.4920	0.4885	0.4900	0.4800
		1.0	0.035	T_1	0.4085	0.5110	0.5110	0.5070	0.5075	0.5080	0.5070
				T_2	0.4085	0.5110	0.5035	0.4955	0.4950	0.4955	0.4965
		1.2	0.041	T_1	0.4660	0.5175	0.5170	0.5190	0.5190	0.5220	0.5215
				T_2	0.4660	0.5175	0.5150	0.5090	0.5050	0.5075	0.5080
		1.4	0.048	T_1	0.4610	0.5250	0.5260	0.5260	0.5235	0.5250	0.5285
				T_2	0.4610	0.5250	0.5200	0.5175	0.5155	0.5080	0.5110
					$\epsilon \sim 0.005\mathscr{N}(-51.5, 172.8) + 0.99\mathscr{N}(0, 1) + 0.005\mathscr{N}(51.5, 172.8)$						
5	15	0.6	0.021	T_1	0.4950	0.4600	0.4770	0.4025	0.2530	0.1535	0.0965
				T_2	0.4950	0.4600	0.4390	0.3145	0.1185	0.0955	0.0905
		0.8	0.028	T_1	0.5275	0.5225	0.5205	0.4795	0.2780	0.1095	0.1140
				T_2	0.5275	0.5225	0.5020	0.3775	0.1475	0.1085	0.1005
		1.0	0.035	T_1	0.5630	0.6165	0.6075	0.5810	0.4690	0.2965	0.1460
				T_2	0.5630	0.6165	0.5810	0.4955	0.1985	0.1405	0.1430
		1.2	0.041	T_1	0.5785	0.6580	0.6550	0.6130	0.4460	0.2670	0.1595
				T_2	0.5785	0.6580	0.6500	0.5510	0.1890	0.1550	0.1515
		1.4	0.048	T_1	0.5625	0.6765	0.6565	0.6070	0.5065	0.3020	0.1610
				T_2	0.5625	0.6765	0.6340	0.5305	0.2005	0.1680	0.1545
	30	0.6	0.021	T_1	0.3655	0.5030	0.4900	0.4935	0.4835	0.4780	0.4520
				T_2	0.3655	0.5030	0.4710	0.4620	0.4240	0.3860	0.2905
		0.8	0.028	T_1	0.4715	0.5600	0.5540	0.5530	0.5565	0.5480	0.5370
				T_2	0.4715	0.5600	0.5440	0.5330	0.5125	0.4925	0.4270
		1.0	0.035	T_1	0.4470	0.5545	0.5490	0.5500	0.5475	0.5465	0.5435
				T_2	0.4470	0.5545	0.5420	0.5370	0.5335	0.5240	0.4750
		1.2	0.041	T_1	0.4850	0.5720	0.5680	0.5645	0.5665	0.5625	0.5600
				T_2	0.4850	0.5720	0.5560	0.5470	0.5430	0.5300	0.4910
		1.4	0.048	T_1	0.5235	0.5735	0.5750	0.5745	0.5760	0.5770	0.5795
				T_2	0.5235	0.5735	0.5730	0.5720	0.5620	0.5415	0.5070

(ii) The power of most UJELr tests was substantially higher than that of ANT, and the power of T_1 was increasing with the number r of constraints. (iii) T_1 was more powerful than T_2 and UJEL2 was the most powerful.

Based on Example 6, Tables 2 and 3 reports the simulated power size-adjusted to 0.05 for testing the null hypothesis of no random effect $H_0 : \sigma_u^2 = \sigma_{u0}^2 = 0$ in the presence of side information of symmetry of $u + \epsilon$. Two JEL tests of size 0.05 are as follows:

$$T_1 = \mathbf{1}\big[-2\log \mathcal{R}_n\big(\mu_0, \sigma_{u0}^2\big) > \chi_{0.95}^2(1)\big],$$
$$T_2 = \mathbf{1}\big[-2\log \mathscr{R}_n\big(\mu_0, \sigma_{u0}^2\big) > \chi_{0.95}^2(r_n + 1)\big], \qquad r_n = 0, 1, \ldots, 5.$$

For comparison, the power of the F-test with df $n - 1$ and $(J - 1)n$ was also simulated. The u and ϵ were respectively, generated from the normal $\mathscr{N}(0, \sigma_u^2)$ for several values of σ_u^2 and the contaminated normal (a) $0.025\mathscr{N}(-51.5, 172.8) + 0.95\mathscr{N}(0, 1) + 0.025\mathscr{N}(51.5, 172.8)$, (b) $0.025\mathscr{U}(-100, -3) + 0.95\mathscr{N}(0, 1) + 0.025\mathscr{U}(3, 100)$, (c) $0.005\mathscr{N}(-51.5, 35.4) + 0.99\mathscr{N}(0, 1) + 0.005\mathscr{N}(51.5, 35.4)$. The results on Tables 2 and 3 were based on (a) and (b), (c) respectively. We also listed the values of ICC (interclass correlation coefficient) $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_\epsilon^2)$. One sees that the JEL tests outperformed the F test except for the cases $J = 5$, $n = 15$ and $\sigma_u^2 = 0.6, 0.8$ on Table 2. The power of the JEL tests was increasing with number r of constraints. For a given sample size, however, the increasing trend may stop as a larger sample is needed to accommodate more constraints. To choose an optimal number r of constraints one can use the bootstrap method. See page 21, Peng and Schick [13].

Supplementary Material

Supplement to "Jackknife empirical likelihood goodness-of-fit tests for U-statistics based general estimating equations" (DOI: 10.3150/16-BEJ884SUPP; .zip). In this Supplement, we introduce the notation and prove the theorems and provide the details to the examples.

References

[1] Arcones, M.A. (1996). The Bahadur–Kiefer representation for U-quantiles. *Ann. Statist.* **24** 1400–1422. MR1401857

[2] Arvesen, J.N. (1969). Jackknifing U-statistics. *Ann. Math. Stat.* **40** 2076–2100. MR0264805

[3] Chen, S.X., Peng, L. and Qin, Y.-L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96** 711–722. MR2538767

[4] Heffernan, P.M. (1997). Unbiased estimation of central moments by using U-statistics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **59** 861–863. MR1483220

[5] Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068

[6] Jing, B.-Y., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *J. Amer. Statist. Assoc.* **104** 1224–1232. MR2562010

[7] Kowalski, J. and Tu, X.M. (2008). *Modern Applied U-Statistics. Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2368050

[8] Lee, A.J. (1990). *U-Statistics*: *Theory and Practice*. *Statistics*: *Textbooks and Monographs* **110**. New York: Dekker. MR1075417

[9] Li, M., Peng, L. and Qi, Y. (2011). Reduce computation in profile empirical likelihood method. *Canad*. *J*. *Statist*. **39** 370–384. MR2839486

[10] Liu, R.Y. (1990). On a notion of data depth based on random simplices. *Ann*. *Statist*. **18** 405–414. MR1041400

[11] Nordman, D.J. and Lahiri, S.N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann*. *Statist*. **34** 3019–3050. MR2329476

[12] Owen, A. (2001). *Empirical Likelihood*. London: Chapman & Hall/CRC.

[13] Peng, H. and Schick, A. (2005). Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown marginals: The least-squares approach. *J*. *Multivariate Anal*. **95** 385–409. MR2170403

[14] Peng, H. and Schick, A. (2013). Empirical likelihood approach to goodness of fit testing. *Bernoulli* **19** 954–981. MR3079302

[15] Peng, H. and Schick, A. (2016). Asymptotic normality of quadratic forms with random vectors of increasing dimension. *Statist*. *Sinica*. To appear.

[16] Peng, H. and Tan, F. (2016). Supplement to "Jackknife empirical likelihood goodness-of-fit tests for U-statistics based general estimating equations." DOI:10.3150/16-BEJ884SUPP.

[17] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley. MR0595165

[18] Shi, X. (1984). Approximate independence of jackknife pseudo values and the bootstrap method. *J*. *Wuhan Institute of Hydra-Electric Engineering* (*Chinese*) **2** 83–90.

[19] Yuan, A., He, W., Wang, B. and Qin, G. (2012). *U*-statistic with side information. *J*. *Multivariate Anal*. **111** 20–38. MR2944403

Received November 2015 and revised June 2016