JACKKNIFE EMPIRICAL LIKELIHOOD GOODNESS-OF-FIT TESTS FOR VECTOR U-STATISTICS

By Fei Tan, Qun Lin, Wei Zheng and Hanxiang Peng

Indiana University Purdue University Indianapolis

Motivated by applications to goodness of fit U-statistic testing, the jackknife empirical likelihood (Jing, *et al.* [14]) is justified with an alternative approach and the Wilks theorems for vector U-statistics are proved. This generalizes Owen's empirical likelihood for vectors to vector U statistics and includes the JEL for U-statistics with side information as a special case. The results are extended to allow for the constraints to use estimated criteria functions and for the number of constraints to grow with the sample size. The developed theory is applied to derive the JEL tests or confidence sets for some useful vector U-statistics associated with chisquare statistic, Cronbach's coefficient alpha, Pearson's correlation, concordance correlation coefficient, Cohen's kappa, Goodman & Kruskal's Gamma, Kendall's τ_b, and interclass correlation; for a linear mixed effects model; for a balanced random effects model; for models with overdisperson and zero-inflated Poisson; for U-quantiles including Hodges-Lehmann median, Gini's mean difference and Theil's test; and for the simplicial depth function. A small simulation is conducted to evaluate the tests.

1. Introduction. To construct tests and confidence sets in a nonparametric setting, Owen [23–25] introduced the empirical likelihood approach. As a likelihood approach of nonparametric nature, it combines the effectiveness of likelihood and the reliability of nonparametrics. It does not require specification of a distribution for data and is particularly convenient to incorporate side information. It gives confidence sets with data-driven shapes and does not require estimation of variances. The empirical likelihood theory has been successfully extended to various branches of statistics with tremendous accomplishments. In this article, we shall develop the theory for vector U statistics with side information and apply it to a number of important examples from different topics in statistics.

Vector U-statistics are useful and each of many frequently used test statistics can be written as a function of vector U-statistics, see e.g. Kowalski and Tu [16], Lee [21] and Serfling [29]. See also the examples in Subsection 3.2. Relatively recently, Jing *et al.* [14] developed an empirical likelihood theory

AMS 2000 subject classifications: Primary 62G10; secondary 62G15,62G20

Keywords and phrases: empirical likelihood; Kendall's tau; simplicial depth; Theil's test; U-statistics; variance component; Wilks' theorem

for univariate U-statistics by exploiting the technique of jackknife pseudo values. The usual empirical likelihood for a U-statistic involves in the nonlinearality of the probabilities $\pi_j, j = 1, \ldots, n$ in the constraint equations that defines the maximization problem for the empirical likelihood. This leads to there being not available the usual explicit solutions for π_j's, which discourages the development of the theory. The jackknife technique circumvents this discouragement. In the meantime, it correctly estimates the variance, so that the Wilks type of theorems still hold and the advantages of the empirical likelihood approach are retained. As in the case of empirical likelihood for time series in Nordman and Lahiri [22], the independence or asymptotic independence which justifies the definition of empirical likelihood as a product of probabilities π_j's is not directly available for a U-statistic of which the summands are not independent but correlated. Jing *et al.* recognized the asymptotic independence of the jackknife pseudo values of a univariate U-statistic and defined the jackknife empirical likelihood (JEL) for the U-statistic.

In justifying the asymptotic independence, Jing *et al.* cited a theorem from Shi [31], who proved the asymptotic independence by an application of the zero-one law for a sequence of exchangeable random variables. Shi's result is not easily available as it was published in Chinese, we here present an alternative, somewhat straightforward, justification of the asymptotic independence based on the Hoeffding decomposition for a U-statistic. Based on this justification and the Hoeffding decomposition for a vector U-statistic, we define the JEL for the vector U-statistic, which takes the JEL for a vector U-statistic with the usual side information as a special case. See Section 2 for the details.

We proved the Wilks theorem for vector U-statistics with side information, which is allowed to be given by either a finite number or a growing number of known or estimated constraints with an increasing sample size. The latter is needed to handle naturally occurring nuisance parameters in semiparametric models. We applied the developed results to obtain the JEL-based tests and confidence sets with side information for a number of important examples from different areas of statistics. In Examples 1 and 2, we discussed the JEL tests for some useful vector U-statistics associated with the chisquare statistic, Cronbach's coefficient alpha, Pearson's correlation, concordance correlation coefficient, Cohen's kappa, Goodman & Kruskal's Gamma, Kendall's τ_b, and interclass correlation. In Example 3, we obtained the JEL tests and confidence sets for linear mixed effects models and variance component models. In particular, we have provided two JEL tests about the random effects. In Example 4, the JEL tests about Binomial and Poisson under/overdisperson are constructed. In Example 5, the JEL tests for U-quantiles with side information are described, including Hodges-Lehmann median, Kendall' tau, Gini's mean difference and Theil's test. In Example 6, the JEL test about the simplicial depth with a growing number of constraints are discussed. In the context of Example 3, we offer an example of a growing number of estimated constraints in Section 5.

As aforementioned, the empirical likelihood theory has been extended to various branches of statistics. These include Bartlett correction (DiCiccio, *et al.* [7]), generalized linear models (Kolaczyk [15]), heteroscedastic partially linear models (Lu [19]), partially linear models (Shi and Lau [30]; Wang and Jing [32]), parametric and semiparametric models in multiresponse regression (Chen and Van Keilegom [5]), right censored data (Li and Wang [17]), U-statistics with side information (Yuan, *et al.* [33]), and stratified samples with nonresponse (Fang, *et al.* [9]). Qin and Lawless [28] linked empirical likelihood with finitely many estimating equations and investigated maximum empirical likelihood estimators. Chen, *et al.* [4] obtained asymptotic normality for the number of constraints growing to infinity. Hjort, *et al.* [12] and Peng and Schick [26, 27] generalized the empirical likelihood approach to allow for the number of constraints to grow with the sample size and for the constraints to use estimated criteria functions. Algorithms, calibration and higher-order precision of the approach can be found in Hall and La Scala [10], Emerson and Owen [8] and Liu and Chen [18] among others.

The rest of the paper is structured as follows: In Section 2, the JEL is introduced with a justification based on the Hoeffding decomposition for U-statistics. In Section 3, the Wilks theorems for vector U-statistics and for U-statistics with growing number of constraints are proved. A number of important examples are given. Section 4 reports a small simulation. The asymptotic behaviors of the JEL with growing number of estimated constraints are studied in Section 5. The result is used to derive a joint confidence set in a random effects model. In Section 6, we first introduce the notation used throughout. Two general asymptotic results on the empirical likelihood are then described. We close this section with a general asymptotic result on the JEL for U-statistics with side information. Some of the proofs and a useful lemma are given in Section 7.

2. Jackknife empirical likelihood for vector U statistics. In this section, we first recall some facts about U-statistics. The JEL for vector U-statistics is then given with a justification.

Let $(\Omega, \mathscr{A})$ be a measurable space and P be a probability measure on this space. Let Z be a random element taking values in some measurable space $(\mathcal{Z}, \mathscr{S})$ with distribution Q under P. Let $Z_1, \ldots, Z_n$ be independent and identical copies of Z. Let h be a measurable function from $\mathcal{Z}^m$ to $\mathcal{R}$ which is argument-symmetric in its m arguments, that is, $h(z_1, \ldots, z_m) = h(z_{\pi_1}, \ldots, z_{\pi_m})$ for every $z_1, \ldots, z_m \in \mathcal{Z}$, where $\pi_1, \ldots, \pi_m$ is an arbitrary permutation of integers $1, \ldots, m$. A U-statistic with kernel h of order m is defined as

$$U_n := U_{nm}(h) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \ldots < i_m \leq n} h(Z_{i_1}, \ldots, Z_{i_m}), \quad n \geq 2.$$

Throughout we assume h is Q^m-square integrable, that is, $h \in L_2(Q^m)$, where $L_2(Q^m) = \{f : \int f^2 \, dQ^m < \infty\}$. We shall abbreviate $\theta = E(h) := E(h(Z_1, \ldots, Z_m)) = \int h \, dQ^m$, $P_n f = n^{-1} \sum_{j=1}^n f(Z_j)$ and $Pf = E(f(Z))$. Then U_n is an unbiased estimator of θ. Let $h_m = h$ and $h_c(z_1, \ldots, z_c) = E(h(z_1, \ldots, z_c, Z_{c+1}, \ldots, Z_m))$ for $c = 1, \ldots, m-1$. Then h_c is a version of the conditional expectation, that is,

$$h_c(z_1, \ldots, z_c) = E\big(h(Z_1, \ldots, Z_m) | Z_1 = z_1, \ldots, Z_c = z_c\big).$$

Let δ_z be the point mass at $z \in \mathcal{Z}$. We now define

$$h_c^*(z_1, \ldots, z_c) = (\delta_{z_1} - P) \ldots (\delta_{z_c} - P) P^{m-c} h, \quad c = 0, 1, \ldots, m.$$

Let $\tilde{f} = f - Pf$ denote the centered version of an integrable function f. Obviously $h_1^* = \tilde{h}_1$. With this notation the useful Hoeffding decomposition can be stated as

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

Let $U_{n-1}^{(-j)}$ denote the U-statistic of order m based on the $n-1$ observations $Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_n$. The jackknife pseudo values of the U-statistic $U_n(h)$ with kernel h are defined as

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

For ease of notation, we sometimes drop h and write $V_{nj} = V_{nj}(h)$ when there is no ambiguity. From (2.1) it follows

$$V_{nj} = \theta + m\tilde{h}_1(Z_j) + R_{nj}, \quad j = 1, \ldots, n, \tag{2.2}$$

where R_{nj} is the remainder given by

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

Using the Hoeffding decomposition (2.1) and the orthogonality property of $U_{nc}(h_c^*)$'s, we can easily prove the following.

Lemma 2.1. *The jackknife pseudo values* V_{nj} *of* $U_n(h)$ *satisfy*

$$E\big((V_{nj}-\theta-m\tilde{h}_1(Z_j))^2\big)=O(n^{-1}),\quad j=1,\dots,n. \tag{2.4}$$

For a complete proof please see (6.7) and thereafter. Thus from (2.4) it immediately follows

$$V_{nj}=\theta+m\tilde{h}_1(Z_j)+O_p(n^{-1/2}),\quad j=1,\dots,n. \tag{2.5}$$

This shows that each jackknife pseudo value V_{nj} depends asymptotically on Z_j so that $V_{nj}, j=1,\dots,n$ are approximately *independent* for large values of n. One nice property of the jackknife pseudo values V_{nj}'s is they satisfy

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

Thus a U-statistic can be expressed as the average of approximately independent random variables (the jackknife pseudo values). Furthermore, if π_j is a probability mass placed at Z_j, then approximately the same mass π_j is placed at the jackknife pseudo value V_{nj} for $j=1,\dots,n$. Therefore, the likelihood of the pseudo values V_{nj}'s is approximately the product of these π_j's. Suppose there is available side information about Q given by

$$\int \mathbf{g}\,dQ=0, \tag{2.7}$$

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

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

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

$$\mathscr{P}_n=\Big\{\boldsymbol{\pi}=(\pi_1,\dots,\pi_n)^\top\in[0,1]^n:\sum_{i=1}^{n}\pi_i=1\Big\}.$$

Here r is the number of equalities which give the side information, and these equalities are referred to as *constraints*.

REMARK 2.1. If we replace the jackknife pseudo values $\tilde{V}_{nj}(h)$ by $m\tilde{h}_1(Z_j)$ in (2.8), then the resulted supremum is $\mathscr{R}_n(mh_1, \mathbf{g})$, which is the usual empirical likelihood. We now consider estimating $m\tilde{h}_1(Z_j)$ by the jackknife pseudo values $\tilde{V}_{nj}(h)$ and work with the estimated constraints. The resulted supremum is then (2.8). This is in the spirit of empirical likelihood with estimated constraints of Hjort, *et al.* [12] and Peng and Schick [26, 27].

It must be mentioned that the preceding definition of JEL for U-statistics with side information covers the case that the side information is given by several U-statistics in view of the Hoeffding decompositions for U-statistics. This is indeed the case of vector U-statistics. Specifically, let $h^{(k)}$ be a kernel (i.e. argument-symmetric and square-integrable) from $\mathcal{Z}^{m_k}$ to $\mathcal{R}$ for $k = 1, \ldots, r$. Let $E\left(U_{nm_k}(h^{(k)})\right) = \theta_k$ and $\tilde{V}_{nj}(h^{(k)}) = V_{nj}(h^{(k)}) - \theta_k$ be the centered jackknife pseudo values of the U-statistic $U_{nm_k}(h^{(k)})$ of order m_k with kernel $h^{(k)}$. Let $\mathbf{h} = (h^{(1)}, \ldots, h^{(m_k)})^\top$, $\mathbf{U}_n(\mathbf{h}) = (U_{nm_1}(h^{(1)}), \ldots, U_{nm_r}(h^{(r)}))^\top$ and $\tilde{\mathbf{V}}_{nj}(\mathbf{h}) = (\tilde{V}_{nj}(h^{(1)}), \ldots, \tilde{V}_{nj}(h^{(m_k)}))^\top$. Based on the discussion, the JEL for the vector U-statistic $\mathbf{U}_n(\mathbf{h})$ is justified to be defined by

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

3. The Wilks theorems and examples. In this section, we first present the Wilks theorems for vector U-statistics and for U-statistics with side information described by a finite number of constraints, followed by the JEL for a U-statistic with a growing number of constraints. Several important examples are given in the end.

3.1. *The Wilks theorems for vector U-statistics.* Our first result generalizes Owen's vector empirical likelihood and Jing, *et al.*'s [14] JEL for univariate U-statistics to vector U-statistics. It holds under the same condition as required for the asymptotic normality of vector U-statistics. The proof is delayed to the last section.

THEOREM 3.1. *Suppose the variance-covariance matrix Var*$(\mathbf{mh}_1(Z))$ *exists and is 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 the chisquare distribution with* r *degrees of freedom, i.e.,*

$$-2\log \mathscr{R}_n(\mathbf{h}) \Rightarrow \chi^2(r).$$

A special case of Theorem 3.1 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 following definition of the empirical likelihood ratio,

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

In this case, we naturally concentrate on the subspace of probability measures Q constrained by (2.7) and look at the empirical likelihood ratio:

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

as it is larger than $\mathscr{R}_n(\mathbf{h},\mathbf{g})$ i.e.

$$\mathcal{R}_n(\mathbf{h},\mathbf{g}) \geq \mathscr{R}_n(\mathbf{h},\mathbf{g}).$$

Clearly the ratio $\mathcal{R}_n(\mathbf{h},\mathbf{g})$ can be expressed as

$$\mathcal{R}_n(\mathbf{h},\mathbf{g}) = \frac{\mathscr{R}_n(\mathbf{h},\mathbf{g})}{\mathscr{R}_n(\mathbf{g})}.$$

It also has an asymptotic Chisquare distribution as stated below.

COROLLARY 3.1. *Let $\mathbf{h}$ be a vector kernel and $\mathbf{m}$ be a vector of positive integers both in $\mathcal{R}^s$. Assume $\mathbf{g}$ is a measurable function from $\mathcal{Z}$ to $\mathcal{R}^r$ such that (2.7) holds. Suppose the variance-covariance matrix* $\operatorname{Cov}(\operatorname{Vec}(\mathbf{m}\mathbf{h}_1, \mathbf{g})(Z))$ *exists and is nonsingular. Then*

$$-2\log \mathscr{R}_n(\mathbf{h},\mathbf{g}) \Rightarrow \chi^2(r+s). \tag{3.1}$$

Hence,

$$-2\log \mathcal{R}_n(\mathbf{h},\mathbf{g}) = -2\log \mathscr{R}_n(\mathbf{h},\mathbf{g}) + 2\log \mathscr{R}_n(\mathbf{g}) \Rightarrow \chi^2(s). \tag{3.2}$$

PROOF. As a special case of Theorem 3.1, we obtain (3.1), from which and Cochran's theorem we derive (3.2), following a standard proof in textbooks for the logarithm of a parametric likelihood ratio to be asymptotically chisquare distributed. □

We now study the JEL when the number $r = r_n$ of constraints grows to infinity with the increasing sample size n. Recall $\mathbf{g}$ is a vector function of dimension r_n. To stress the dependence on n we write $\mathbf{g}_n = \mathbf{g}$. With (2.7) as side information, the JEL for a U statistic is as follows:

$$\mathscr{R}_n(h, \mathbf{g}_n) = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \mathrm{Vec}\big(\tilde{V}_{nj}(h), \mathbf{g}_n(Z_j)\big) = 0\Big\}.$$

As in Subsection 3.1, one defines $\mathcal{R}_n(h, \mathbf{g}_n)$ and has the relationship $\mathcal{R}_n(h, \mathbf{g}_n) = \mathscr{R}_n(h, \mathbf{g}_n)/\mathscr{R}_n(\mathbf{g}_n)$. We are interested in establishing

$$\frac{-2\log \mathscr{R}_n(h, \mathbf{g}_n) - (r_n+1)}{\sqrt{2(r_n+1)}} \Rightarrow \mathscr{N}(0,1) \tag{3.3}$$

as r_n tends to infinity slowly with n under suitable conditions.

Remark 3.1. One can interpret (3.3) as $-2\log \mathscr{R}_n(h, \mathbf{g}_n)$ approximately chisquare-distributed with r_n+1 degrees of freedom, whence $-2\log \mathcal{R}_n(h, \mathbf{g}_n)$ approximately chisquare-distributed with 1 degree of freedom.

To this goal, for a $r_n \times r_n$ symmetric matrix $\mathbb{M}_n$ denote by $\lambda_{\min}(\mathbb{M}_n)$ and $\lambda_{\max}(\mathbb{M}_n)$ the smallest and largest eigen values of $\mathbb{M}_n$, i.e.,

$$\lambda_{\min}(\mathbb{M}_n) = \inf_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbb{M}_n \mathbf{u} \quad \text{and} \quad \lambda_{\max}(\mathbb{M}_n) = \sup_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbb{M}_n \mathbf{u},$$

where $\|\cdot\|$ denotes the euclidean norm. Following Peng and Schick [27], 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 measurable vector functions $\{\mathbf{v}_n\}$ on $\mathcal{Z}$ is *Lindeberg* if for every $\epsilon > 0$,

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

Useful properties for Lindeberg sequences can be found in Peng and Schick [26, 27]. Here we quote three properties for our late use.

(L0) If $\{\mathbf{u}_n\}$ and $\{\mathbf{v}_n\}$ are Lindeberg, so are $\{\max(\|\mathbf{u}_n\|, \|\mathbf{v}_n\|)\}$ and $\{\mathbf{u}_n + \mathbf{v}_n\}$.
(L1) If $\{\mathbf{v}_n\}$ is Lindeberg, then $\max_{1\le j\le n} \|\mathbf{v}_n(Z_j)\| = o_p(n^{1/2})$.
(L2) If $\int \|\mathbf{v}_n\|^r\, dQ = o(n^{r/2-1})$ for some $r > 2$, then $\{\mathbf{v}_n\}$ is Lindeberg.

For matrices $\mathbb{A}$, $\mathbb{C}$ and $\mathbb{M}$ of compatible dimensions, we define the matrix function $\mathscr{W}$ by

$$\mathscr{W}(\mathbb{A}, \mathbb{C}, \mathbb{M}) = \begin{pmatrix} \mathbb{A} & \mathbb{C}^\top \\ \mathbb{C} & \mathbb{M} \end{pmatrix}. \tag{3.4}$$

Set

$$\mathbf{C}_n = \int mh_1 \mathbf{g}_n \, dQ, \quad \mathbb{W}_n = \int \mathbf{g}_n^{\otimes 2} \, dQ, \tag{3.5}$$

where $\mathbf{w}^{\otimes 2} = \mathbf{w}\mathbf{w}^\top$. As a special case of Theorem 5.1 below, the distribution of $-2\log \mathscr{R}_n(h, \mathbf{g}_n)$ is approximately a chisquare with $r_n + 1$ degrees of freedom as stated next. This generalizes Theorem 3.1 from a finite number of constraints to a growing number. The proof is delayed to the last section.

THEOREM 3.2. *Suppose $\mathbf{g}_n$ is a measurable function from $\mathcal{Z}$ to $\mathcal{R}^{r_n}$ such that $\int \mathbf{g}_n \, dQ = 0$. Suppose further the sequences $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ are Lindeberg such that $\mathscr{W}(Var(mh_1(Z_1)), \mathbf{C}_n, \mathbb{W}_n)$ satisfies (R). Then (3.3) holds as both r_n and n tend to infinity such that $r_n = o(n^{1/2})$.*

3.2. *Examples.* Here we apply Theorems 3.1 – 3.2 to derive the JEL-based tests and confidence sets for a number of frequently used test statistics.

EXAMPLE 1. GOODMAN & KRUSKAL'S GAMMA. Vector U-statistics are common as each of many frequently used test statistics can yield a vector U-statistic, see e.g. Kowalski and Tu [16]. These U-statistics extend the scope of the frequently used test statistics and provide additional useful tests. The JEL for vector U-statistics with side information and their asymptotic distributions can be obtained by using Theorem 3.1 or Corollary 3.1. Let us illustrate this using Goodman & Kruskal's Gamma. To this end, let $\mathbf{Z}_j = (X_j, Y_j), j = 1, \dots, n$ be i.i.d. copies of a random vector $\mathbf{Z} = (X, Y)$. The Goodman & Kruskal's Gamma is defined as $\gamma = (P(C) - P(D))/(P(C) + P(D))$, where $C = \{(X_1 - X_2)(Y_1 - Y_2) > 0\}$ and $D = \{(X_1 - X_2)(Y_1 - Y_2) < 0\}$. Associated with it a vector U-statistic $\mathbf{U}_n(\mathbf{h})$ of order 2 can be constructed with the kernel equal to

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

Similar vector U-statistics can be constructed, for example, for Kendall's tau-b, Pearson's correlation, Cohen's kappa, interclass correlation, concordance correlation coefficient, and Cronbach's coefficient alpha. Useful vector U-statistics can also be constructed from models for group comparison such as ANOVA for modeling variances, random-factor ANOVA, ROC analysis, and models for K-sample mean and variance. Another instance is given in Example 2.

The JEL with side information given by (2.7) for the Gamma is

$$\mathscr{R}_n(\boldsymbol{\theta}, \mathbf{g}) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j \mathrm{Vec}\big(\mathbf{V}_{nj}(\mathbf{h}) - \boldsymbol{\theta}, \mathbf{g}(\mathbf{Z}_j)\big) = 0 \Big\},$$

where $\mathbf{V}_{nj}(\mathbf{h})$ are the jackknife pseudo values of $\mathbf{U}_n(\mathbf{h})$ and $\boldsymbol{\theta} = (P(C), P(D))^\top$. As in Subsection 3.1, one defines $\mathcal{R}_n(\boldsymbol{\theta}, \mathbf{g})$ and has $\mathcal{R}_n(\boldsymbol{\theta}, \mathbf{g}) = \mathscr{R}_n(\boldsymbol{\theta}, \mathbf{g})/\mathscr{R}_n(\mathbf{g})$. An example of $\mathbf{g}$ is $\mathbf{g} = \mathrm{Vec}(g_1, g_2)$ with $g_1(x) = \mathbf{1}[x \le m_{10}] - 1/2$ and $g_2(y) = \mathbf{1}[y \le m_{20}] - 1/2$, where m_{10} and m_{20} are the known medians of X and Y respectively. Let $\mathbf{h}_1(\mathbf{z}) = E(\mathbf{h}(\mathbf{z}, \mathbf{Z}_2))$. Then by Corollary 3.1 one has

$$P\big(-2\log \mathcal{R}_n(\boldsymbol{\theta}_0, \mathbf{g}) = -2\log \mathscr{R}_n(\boldsymbol{\theta}_0, \mathbf{g}) + 2\log \mathscr{R}_n(\mathbf{g}) > \chi^2_{1-\alpha}(2)\big) \to \alpha,$$

provided that the dispersion matrix $\mathscr{W}(\mathrm{Var}(2\mathbf{h}_1(\mathbf{Z})), \mathbf{C}, \mathrm{Var}(\mathbf{g}(\mathbf{Z})))$ is nonsingular, where $\mathbf{C} = E(\mathbf{h}_1(\mathbf{Z}) \otimes \mathbf{g}(\mathbf{Z}))$. This shows that $\mathbf{1}[-2\log \mathcal{R}(\boldsymbol{\theta}_0, \mathbf{g}) > \chi^2_{1-\alpha}(2)]$ is an asymptotic test of size α for testing the hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. It is noteworthy that 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.

EXAMPLE 2. TESTING INDEPENDENCE BETWEEN TWO CATEGORICAL OUTCOMES. Let (U, V) be a bivariate categorical r.v. whose marginals have K, L levels indexed by r_k, s_l respectively. Let $(U_i, V_i), i = 1, \ldots, n$ be i.i.d. copies of (U, V). Based on the sample we are interested in testing the null hypothesis that U and V are independent. Chisquare or Fisher's exact tests are commonly used. Here we shall use the JEL for vector U-statistics to give an asymptotic test based on the vector U-statistic $\mathbf{U}_n(\mathbf{h})$, see also p. 260 of Kowalski and Tu [16]. To this end, set $\mathbf{Z} = (\mathbf{X}^\top, \mathbf{Y}^\top)^\top$ where

$$\mathbf{X} = (\mathbf{1}[U = r_1], \ldots, \mathbf{1}[U = r_K])^\top, \quad \mathbf{Y} = (\mathbf{1}[V = s_1], \ldots, \mathbf{1}[V = s_L])^\top.$$

Let $\mathbf{Z}_i = \mathrm{Vec}(\mathbf{X}_i, \mathbf{Y}_i), i = 1, \ldots, n$ be the corresponding i.i.d. copies of $\mathbf{Z}$. Independence is equivalent to the statement

$$\boldsymbol{\delta} = E\left(2^{-1}(\mathbf{X}_1 - \mathbf{X}_2) \otimes (\mathbf{Y}_1 - \mathbf{Y}_2)\right) = 0. \tag{3.6}$$

This suggests us to look at the JEL with side information given by (2.7) as follows:

$$\mathscr{R}_n(\mathbf{h}, \mathbf{g}) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \mathrm{Vec}(\mathbf{V}_{nj}, \mathbf{g}(\mathbf{Z}_j)) = 0 \Big\},$$

where $\mathbf{V}_{nj}$ are the jackknife pseudo values of the vector U-statistic $\mathbf{U}_n(\mathbf{h})$ of order 2 with the $(K-1)(L-1)$-dimensional kernel $\mathbf{h}$ given by

$$\mathbf{h}(\mathbf{z}_1, \mathbf{z}_2) = 2^{-1}(\mathbf{x}_1 - \mathbf{x}_2)[-K] \otimes (\mathbf{y}_1 - \mathbf{y}_2)[-L],$$

where $\mathbf{x}[-k]$ denotes the $(K-1)$-dimensional vector obtained from $\mathbf{x}$ with the deletion of the k-th component. Also cf. (3.6). Similar to Subsection 3.1, one defines $\mathcal{R}_n(\mathbf{h},\mathbf{g})$ and claims $\mathcal{R}_n(\mathbf{h},\mathbf{g}) = \mathscr{R}_n(\mathbf{h},\mathbf{g})/\mathscr{R}_n(\mathbf{g})$. In our simulation study below, the side information is that the marginal distributions are known:

$$p_{k\cdot} = p_{k\cdot}^{(0)}, \quad p_{l\cdot} = p_{\cdot l}^{(0)}, \quad k = 1,\dots,K, l = 1,\dots,L, \tag{3.7}$$

where $p_{k\cdot}^{(0)}, p_{\cdot l}^{(0)}$ are known probabilities. In this case, $\mathbf{g} = \mathrm{Vec}(\mathbf{g}_1, \mathbf{g}_2)$ with

$$\begin{aligned} \mathbf{g}_1(\mathbf{z}) &= \mathrm{Vec}(\mathbf{1}[U = r_{1,0}] - p_{1,\cdot}^{(0)}, \dots, \mathbf{1}[U = r_{K-1,0}] - p_{K-1,\cdot}^{(0)}), \\ \mathbf{g}_2(\mathbf{z}) &= \mathrm{Vec}(\mathbf{1}[V = s_{1,0}] - p_{\cdot,1}^{(0)}, \dots, \mathbf{1}[V = s_{L-1,0}] - p_{\cdot,L-1}^{(0)}), \end{aligned} \tag{3.8}$$

where $r_{k,0}, s_{l,0}$ are known numbers. Note that each of the two marginal probabilities sums up to one so there are only $(K-1)(L-1)$ independent cell probabilities. Let $\mathbf{h}_1(\mathbf{z}) = E(\mathbf{h}(\mathbf{z}, \mathbf{Z}_2)), \mathbf{z} \in \mathcal{R}^{KL}$. Then by Corollary 3.1 under the null hypothesis of independence one has

$$-2\log \mathcal{R}_n(\mathbf{h},\mathbf{g}) = -2\log \mathscr{R}_n(\mathbf{h},\mathbf{g}) + 2\log \mathscr{R}_n(\mathbf{g}) \Rightarrow \chi^2((K-1)(L-1)),$$

provided that the dispersion matrix $\Sigma = \mathrm{Var}(2\mathbf{h}_1(\mathbf{Z}_1))$ is nonsingular. We demonstrate in the last section that Σ is nonsingular if and only if all the marginal probabilities are nonzero, i.e.

$$p_{k\cdot} \neq 0,\ k = 1,\dots,K \quad \text{and} \quad p_{\cdot l} \neq 0,\ l = 1,\dots,L, \tag{3.9}$$

where $\mathbf{p} = (p_{lk})_{KL} = E(\mathbf{X}\mathbf{Y}^\top)$.

Example 3. Linear mixed effects models. In a longitudinal study with n subjects and J assessment points, the response Y_{ij}, fixed effect $\mathbf{x}_{ij}$, random effect $\mathbf{u}_i$, and random error ϵ_{ij} satisfy

$$Y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{w}_{ij}^\top \mathbf{u}_i + \epsilon_{ij}, \quad j = 1,\dots,J,\ i = 1,\dots,n, \tag{3.10}$$

where $\mathbf{z}_i = \mathrm{Vec}(\mathbf{x}_i, \mathbf{w}_i)$ with $\mathbf{x}_i = (\mathbf{x}_{i1},\dots,\mathbf{x}_{iJ})^\top$ and $\mathbf{w}_i = (\mathbf{w}_{i1},\dots,\mathbf{w}_{iJ})^\top$ are n i.i.d. copies of $\mathbf{z} = \mathrm{Vec}(\mathbf{x},\mathbf{w})$ and have finite second moments, the $\boldsymbol{\epsilon}_i = (\epsilon_{i1},\dots,\epsilon_{iJ})^\top$ are i.i.d. with mean zero and finite variance-covariance matrix $\Sigma_\epsilon = \mathrm{Var}(\epsilon_i) = \mathrm{Diag}(\sigma_1^2,\dots,\sigma_J^2)$, the $\mathbf{u}_i$ are i.i.d. with finite variance-covariance matrix $\Sigma_u = \mathrm{Var}(\mathbf{u}_i)$, and $\mathbf{z}_i$, $\mathbf{u}_i$ and $\boldsymbol{\epsilon}_i$ are independent. To introduce the distribution-free U-statistics based- generalized estimating equations (UGEE), set $\mathbf{y}_i = (Y_{i1},\dots,Y_{iJ})^\top$, $\boldsymbol{\theta} = \mathrm{Vec}(\boldsymbol{\beta}, \Sigma_u, \Sigma_\epsilon)$, and $\mathbf{f} = \mathrm{Vec}(\mathbf{f}_1, \mathbf{f}_2)$, where

$$\begin{aligned} &\mathbf{f}_1 = (f_{11},\dots,f_{1J})^\top \text{ with } f_{1j}(\mathbf{y}_i, \mathbf{y}_{i'}) = Y_{ij} + Y_{i'j}, \\ &\mathbf{f}_2 = (f_{211}, f_{212},\dots,f_{2(J-1)J}, f_{2JJ})^\top,\ f_{2jj'}(\mathbf{y}_i,\mathbf{y}_{i'}) = (Y_{ij} - Y_{i'j})(Y_{ij'} - Y_{i'j'}). \end{aligned}$$

The conditional means $\boldsymbol{\mu} = \mathrm{Vec}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ are easily calculated as follows:

$$\begin{aligned}
&\boldsymbol{\mu}_1 = (\mu_{11}, \dots, \mu_{1J})^\top, \quad \boldsymbol{\mu}_2 = (\mu_{211}, \mu_{212}, \dots, \mu_{2(J-1)J}, \mu_{2JJ})^\top, \\
&\mu_{1j}(\boldsymbol{\theta}; \mathbf{z}_i, \mathbf{z}_{i'}) = E(f_{1j}(\mathbf{y}_i, \mathbf{y}_{i'})|\mathbf{z}_i, \mathbf{z}_{i'}) = (\mathbf{x}_{ij} + \mathbf{x}_{i'j})^\top \boldsymbol{\beta}, \\
&\mu_{2jj'}(\theta; \mathbf{z}_i, \mathbf{z}_{i'}) = \boldsymbol{\beta}^\top (\mathbf{x}_{ij} - \mathbf{x}_{i'j})(\mathbf{x}_{ij'} - \mathbf{x}_{i'j'})^\top \boldsymbol{\beta} \\
&\qquad\qquad + \mathbf{u}_{ij}^\top \Sigma_u \mathbf{u}_{i'j} + \mathbf{u}_{ij'}^\top \Sigma_u \mathbf{u}_{i'j'} + 2\sigma_j^2 \mathbf{1}_{jj'}.
\end{aligned}$$

The UGEE for estimating $\boldsymbol{\theta}$ is $\mathbf{U}_{n2}(\mathbf{h}(\cdot; \boldsymbol{\theta})) = 0$ with the kernel $\mathbf{h}$ given by

$$\mathbf{h}(\mathbf{t}_1, \mathbf{t}_2; \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\theta}} (\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})(\mathbf{f}(\mathbf{y}_1, \mathbf{y}_2) - \boldsymbol{\mu}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})),$$

where $\mathbf{t} = \mathrm{Vec}(\mathbf{y}, \mathbf{z})$. This suggests us to look at the JEL for the vector U-statistic as follows:

$$\mathscr{R}_n(\boldsymbol{\theta}) = \sup \Big\{ \prod_{i=1}^n n\pi_i : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{i=1}^n \pi_i \mathbf{V}_{ni}(\mathbf{h}, \boldsymbol{\theta}) = 0 \Big\},$$

where $\mathbf{V}_{ni}(\mathbf{h}, \boldsymbol{\theta})$ is the jackknife pseudo values of the vector U-statistics $\mathbf{U}_{n2}(\mathbf{h}(\cdot; \boldsymbol{\theta}))$. Let $\mathbf{h}_1(\mathbf{t}; \boldsymbol{\theta}) = E(\mathbf{h}(\mathbf{t}_1, \mathbf{t}_2; \boldsymbol{\theta})|\mathbf{t}_1 = \mathbf{t})$. By Theorem 3.1, if $\mathrm{Var}(\mathbf{h}_1(\mathbf{t}; \boldsymbol{\theta}_0))$ is nonsingular then an asymptotic test of size α for the null hypothesis of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is $\mathbf{1}[-2\log \mathscr{R}_n(\boldsymbol{\theta}_0) > \chi^2_{1-\alpha}(\dim(\boldsymbol{\theta}_0))]$.

Joint confidence sets for variance components. As a special case of the LMM (3.10), the balanced one-way random effects model is

$$Y_{ij} = \mu + u_i + \epsilon_{ij}, \quad i = 1, \dots, n, \ j = 1, \dots, J (J \geq 2), \tag{3.11}$$

where μ is the mean response, the ϵ_{ij}'s are i.i.d. with mean zero and variance $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_{ij})$, the u_i's are i.i.d. with mean zero and variance $\sigma_u^2 = \mathrm{Var}(u_j)$, and ϵ_{ij}'s and u_i's are independent and have finite fourth moments.

The commonly used confidence regions for the variances heavily depend on the assumption of normality of the model. Here we employ the JEL to give confidence sets for the variances. It is well known that a random variable X has a finite fourth moment then the minimum variance unbiased estimator (MVUE) of the variance $\mathrm{Var}(X)$ is the U-statistic of order two with the kernel equal to $2^{-1}(X_1 - X_2)^2$, where X_1, X_2 are i.i.d. copies of X, see e.g. Heffernan [11]. We shall exploit the MVUE's of variances in our forthcoming study of confidence sets.

Following Arvesen [2], put

$$\mathbf{X}_i = \begin{pmatrix} Y_{i\cdot} \\ (J-1)^{-1} \sum_{j=1}^J (Y_{ij} - Y_{i\cdot})^2 \end{pmatrix}, \quad i = 1, \dots, n, \tag{3.12}$$

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, with $\kappa(\mathbf{X}_i) = (J-1)^{-1}\sum_{j=1}^{J}(Y_{ij} - Y_{i\cdot})^2$,

$$h^{(1)}(\mathbf{X}_i, \mathbf{X}_{i'}) = 2^{-1}(\kappa(\mathbf{X}_i) + \kappa(\mathbf{X}_{i'})), \quad h^{(2)}(\mathbf{X}_i, \mathbf{X}_{i'}) \ = 2^{-1}(Y_{i\cdot} - Y_{i'\cdot})^2.$$

Then one readily calculates

$$E(h^{(1)}(\mathbf{X}_1, \mathbf{X}_2)) = \sigma_\epsilon^2, \quad E(h^{(2)}(\mathbf{X}_1, \mathbf{X}_2)) = \sigma^2 := \sigma_u^2 + J^{-1}\sigma_\epsilon^2. \tag{3.13}$$

Therefore the vector U-statistic

$$\mathbf{U}_n(\mathbf{h}) = \mathrm{Vec}(U_n(h^{(1)}), U_n(h^{(2)}))$$

is an unbiased estimators of $\boldsymbol{\theta} = (\sigma_\epsilon^2, \sigma^2)^\top$. The JEL for the vector U-statistic $\mathbf{U}_n(\mathbf{h})$ is then given by

$$\mathscr{R}_n(\boldsymbol{\theta}) = \sup\Big\{\prod_{i=1}^{n} n\pi_i : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{i=1}^{n} \pi_i(\mathbf{V}_{ni}(\mathbf{h}) - \boldsymbol{\theta}) = 0\Big\}, \ \boldsymbol{\theta} \in \mathcal{R}^+ \times \mathcal{R}^+,$$

where $\mathbf{V}_{ni}(\mathbf{h}) = \mathrm{Vec}\big(V_{ni}(h^{(1)}), V_{ni}(h^{(2)})\big)$ is the vector whose components are the jackknife pseudo values of the U-statistics $U_n(h^{(1)})$ and $U_n(h^{(2)})$ respectively. By Theorem 3.1, if $\mathrm{Var}(\mathbf{h}_1(\mathbf{X}))$ is nonsingular then an asymptotic joint confidence set for $\boldsymbol{\theta}$ at the $1-\alpha$ level of significance is given by

$$\left\{\boldsymbol{\theta} \in \mathcal{R}^+ \times \mathcal{R}^+ : \ -2\log \mathscr{R}_n(\boldsymbol{\theta}) \le \chi^2_{1-\alpha}(2)\right\}.$$

It is noteworthy that a confidence set for $\vartheta = (\sigma_\epsilon^2, \sigma_u^2)^\top$ can be obtained by the transformation $\vartheta_1 = \theta_1, \vartheta_2 = \theta_2 - \theta_1/J$. Also, a confidence set for σ_u^2 can be obtained by $J \to \infty$. The null hypothesis $H_0 : (\sigma_\epsilon^2, \sigma_u^2) = (\sigma_{\epsilon 0}^2, \sigma_{u0}^2)$ must be rejected if $(\sigma_{\epsilon 0}^2, \sigma_{u0}^2 + J^{-1}\sigma_{\epsilon 0}^2)$ does not belong to the above confidence set. In particular, the null hypothesis $H_0 : \sigma_\epsilon^2 = \sigma_{\epsilon 0}^2, \sigma_u^2 = 0$ must be rejected if $(\sigma_{\epsilon 0}^2, J^{-1}\sigma_{\epsilon 0}^2)$ does not belong to the above confidence set.

It is easy to calculate

$$\begin{aligned} h_1^{(1)}(\mathbf{x}_1) &= E(h^{(1)}(\mathbf{x}_1, \mathbf{X}_2)) = 2^{-1}(\kappa(\mathbf{x}_1) + \sigma_\epsilon^2), \\ h_1^{(2)}(\mathbf{x}_1) &= E(h^{(2)}(\mathbf{x}_1, \mathbf{X}_2)) = 2^{-1}((y_{1\cdot} - \mu)^2 + \sigma^2). \end{aligned} \tag{3.14}$$

Testing the random effects. Let $\psi = h^{(2)} - J^{-1}h^{(1)}$. Clearly ψ is argument-symmetric and from (3.13) it follows that

$$\psi(\mathbf{X}_1, \mathbf{X}_2) = 2^{-1}\big((Y_{1\cdot} - Y_{2\cdot})^2 - J^{-1}(\kappa(\mathbf{X}_1) + \kappa(\mathbf{X}_2))\big), \quad E(\psi(\mathbf{X}_1, \mathbf{X}_2)) = \sigma_u^2.$$

This suggests to look at the JEL as follows:

$$\mathscr{R}_n(\sigma_u^2) = \sup\Big\{\prod_{i=1}^n n\pi_i : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{i=1}^n \pi_i(\mathbf{V}_{ni}(\psi) - \sigma_u^2) = 0\Big\},\ \sigma_u^2 \in [0,\infty),$$

where $\mathbf{V}_{ni}(\psi)$ are the jackknife pseudo values of the U-statistic $U_n(\psi)$. Let $\psi_1(\mathbf{x}_1) = E(\psi(\mathbf{x}_1, \mathbf{X}_2))$. It follows from Theorem 3.1 that if $\sigma^2(\psi_1) = \mathrm{Var}(\psi_1(\mathbf{X}))$ is nonsingular then an asymptotic test of size α for the null hypotheis $H_0 : \sigma_u^2 = 0$ is $\mathbf{1}[-2\log \mathscr{R}_n(0) > \chi^2_{1-\alpha}(1)]$. By (3.14) one has $\psi_1(\mathbf{X}_1) = 2^{-1}\big((Y_{1\cdot} - \mu)^2 - J^{-1}\kappa(\mathbf{X}_1)\big) - 2^{-1}\sigma_u^2$. Thus with some algebra one shows $\sigma^2(\psi_1) > 4^{-1}E(u_1^4) \geq 0$ when $J \geq 3$.

EXAMPLE 4. MODELS FOR OVERDISPERSON. In a GEE model, the mean μ_i of the response Y_i is modeled as a function a covariate vector $\mathbf{X}_i$ as follows:

$$g(\mu_i) = \mathbf{X}_i^\top \boldsymbol{\beta}, \quad i = 1, \ldots, n, \tag{3.15}$$

where g is a link function and $\boldsymbol{\beta}$ is a parameter. Here $\mathbf{T}_i = \mathrm{Vec}(Y_i, \mathbf{X}_i), i = 1, \ldots, n$ is a random sample of $\mathbf{T} = \mathrm{Vec}(Y, \mathbf{X})$. The parameter $\boldsymbol{\beta}$ can be estimated as a solution to the GEE. Suppose the conditional variance of each Y_i given $\mathbf{X}_i$ is a function of the mean μ_i and some parameter $\boldsymbol{\alpha}$, i.e., $\mathrm{Var}(Y_i|\mathbf{X}_i) = V(\mu_i, \boldsymbol{\alpha})$. By simultaneously modeling the mean and variance, more efficiency can be gained and other issues such as overdisperson can be handled. To introduce the UGEE, set $\mathbf{f} = \mathrm{Vec}(f_1, f_2)$ where

$$f_1(y_1, y_2) = y_1 + y_2, \quad f_2(y_1, y_2) = (y_1 - y_2)^2, \quad y_1, y_2 \in \mathcal{R}.$$

Then it is easy to calculate $\mathbf{h}(\mu_1, \mu_2, \boldsymbol{\alpha}) = E(\mathbf{f}(Y_1, Y_2)|\mathbf{X}_1, \mathbf{X}_2)$ with $\mathbf{h} = \mathrm{Vec}(h_1, h_2)$ where

$$h_1(\mu_1, \mu_2, \boldsymbol{\alpha}) = \mu_1 + \mu_2, \quad h_2(\mu_1, \mu_2, \boldsymbol{\alpha}) = V(\mu_1, \boldsymbol{\alpha}) + V(\mu_2, \boldsymbol{\alpha}) + (\mu_1 - \mu_2)^2.$$

The UGEE for estimating $\boldsymbol{\theta} = \mathrm{Vec}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is then given by

$$\mathbf{U}_n(\boldsymbol{\kappa}) = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \mathbf{G}(\mathbf{X}_i, \mathbf{X}_j, \boldsymbol{\alpha})\big(\mathbf{f}(Y_i, Y_j) - \mathbf{h}(\mu_i, \mu_j, \boldsymbol{\alpha})\big) = 0, \tag{3.16}$$

where $\mathbf{G}(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\alpha})$ is some argument-symmetric matrix function of $\mathbf{x}_1, \mathbf{x}_2$. The choice of $\mathbf{G}$ is not unique and the consistency of the UGEE estimates is independent of the selection of $\mathbf{G}$. In most cases, we can choose $\mathbf{G} = \mathbf{D}\mathbf{V}^{-1}$, where $\mathbf{D} = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$ and $\mathbf{V}$ is some compatible matrix which may depend on some other parameter $\boldsymbol{\nu}$. Extensive discussions can be found in the literature about distribution-free models. This suggests us to look at the JEL as follows:

$$\mathscr{R}_n(\boldsymbol{\theta}) = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j \mathbf{V}_{nj}(\boldsymbol{\theta}) = 0 \Big\}, \tag{3.17}$$

where $\mathbf{V}_{nj}(\boldsymbol{\theta})$ are the jackknife pseudo values of the vector U-statistic $\mathbf{U}_n(\boldsymbol{\kappa})$. Let $\boldsymbol{\kappa}_1(\mathbf{t}) = E(\boldsymbol{\kappa}(\mathbf{T}_1, \mathbf{T}_2; \boldsymbol{\theta})|\mathbf{T}_1 = \mathbf{t})$. By Theorem 3.1, if $\mathrm{Var}(\boldsymbol{\kappa}_1(\mathbf{T}))$ is non-singular then an asymptotic test of size α for the null hypothesis of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is $\mathbf{1}[-2\log \mathscr{R}_n(\boldsymbol{\theta}_0) > \chi^2_{1-\alpha}(\dim(\boldsymbol{\theta}_0))]$.

Due to data clustering, overdisperson often occurs. This includes Binomial under/overdisperson and Poisson under/overdisperson. The usual logistic or Poisson log-linear models are not appropriate. In the Poisson log-linear, the link g is the log function. If we take the conditional variance function to be $V(\mu_i, \lambda) = \lambda^2 \mu_i$ for some parameter λ. Then $h_2(\mu_1, \mu_2, \lambda) = \lambda^2(\mu_1 + \mu_2) + (\mu_1 - \mu_2)^2$. We can test under- or overdisperson by considering the hypotheses: Underdisperson: $H_0 : \lambda^2 = 1$ vs $H_1 : \lambda^2 < 1$ and Overdisperson: $H_0 : \lambda^2 = 1$ vs $H_1 : \lambda^2 > 1$. Since the negative binomial model is a substitute for the overdispersed Poisson data, we may take $V(\mu_i) = \mu_i(1 + \alpha\mu_i)$, then $h_2(\mu_1, \mu_2, \alpha) = \mu_1 + \mu_2 + \alpha(\mu_1^2 + \mu_2^2) + (\mu_1 - \mu_2)^2$. Clearly overdisperson can be tested by looking at the hypothesis $H_0 : \alpha = 0$ versus $H_1 : \alpha > 0$.

For Binomial under/overdisperson, the usual choice for the link g is the logistic, and the variance function can be chosen as $V(\mu_i, \lambda) = \lambda^2 \mu_i(1 - \mu_i)$ for some parameter λ. Then we can test under/overdisperson as in the above Poisson model.

In the zero-inflated Poisson model, the distribution can be model as a two-component mixture as follows:

$$f(y_i|\mathbf{X}_i) = \alpha(\mathbf{U}_i) f_0(y_i) + (1 - \alpha(\mathbf{U}_i)) f_P(y_i|\mathbf{V}_i), \quad y_i = 0, 1, \ldots, \tag{3.18}$$

where $\mathbf{U}_i, \mathbf{V}_i$ are (possibly overlapped) subsets of $\mathbf{X}_i$, f_0 denotes a degenerate distribution at 0 to account for the structural zeros, f_P is the Poisson distribution to model the remaining observations and α a mixing proportion of the model. The proportion $\alpha(\mathbf{U}_i)$ and the mean $\mu(\mathbf{V}_i)$ of Y_i can be modeled as follows:

$$\log \alpha_i/(1 - \alpha_i) = \mathbf{U}_i^\top \boldsymbol{\beta}_1, \quad \log(\mu_i) = \mathbf{V}_i^\top \boldsymbol{\beta}_2,$$

where $\alpha_i = \alpha(\mathbf{U}_i)$, $\mu_i = \mu(\mathbf{V}_i)$ and $\boldsymbol{\beta} = \mathrm{Vec}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. In this case, $h_1(\mu_i, \mu_j, \boldsymbol{\beta}) = (1 - \alpha_i)\mu_i + (1 - \alpha_j)\mu_j$ and $h_2(y_i, y_j, \boldsymbol{\beta})$ is equal to

$$\mu_i(1 - \alpha_i)(1 + \alpha_i\mu_i) + \mu_j(1 - \alpha_j)(1 + \alpha_j\mu_j) + \big(\mu_i(1 - \alpha_i) - \mu_j(1 - \alpha_j)\big)^2.$$

EXAMPLE 5. U-QUANTILES. The theory of U-quantile provides a unified treatment of several commonly used statistics, see Arcones [1]. Let $\kappa : \mathcal{Z}^m \mapsto \mathcal{R}$ be a measurable argument-symmetric function. Associated with κ there induces a distribution function $H(t) = P\big(\kappa(Z_1, \ldots, Z_m) \le t\big), t \in \mathcal{R}$. The MVUE of $H(t)$ is the U-statistic of order m given by

$$H_n(t) := H_{nm}(t) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} \mathbf{1}[\kappa(Z_{i_1}, \ldots, Z_{i_m}) \le t], \quad t \in \mathcal{R},$$

Following Arcones [1], the κ shall be referred to as the kernel (of the U-quantile). The U-quantiles include the Hodges-Lehmann median estimator, Gini's mean difference, Theil's estimator of the slope in a simple linear model and Kendall's tau. They correspond 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.

As $H(t)$ is a distribution function, its p-th quantile q is well defined by $q = \inf\{t : H(t) \ge p\}$ for $p \in [0, 1]$. We are interested in testing the null hypothesis that the p-th quantile q is equal to some specified value $q_0 \in \mathcal{R}$ for a known value p_0, i.e., $H_0 : q = q_0$. Suppose there is available additional information about the underlying distribution given by (2.7). For example, (i) Q has zero median and (ii) Q has zero mean. The former corresponds to taking $g(z) = \text{sign}(z)$, while the latter to $g(z) = z$. The JEL for the U-statistic $H_n(q_0)$ with side information can be constructed as in Example 1.

EXAMPLE 6. THE SIMPLICIAL DEPTH FUNCTION. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. with a distribution Q on $\mathcal{R}^m$. Liu [21] introduced the *simplicial depth function* $D(\mathbf{x})$ of a point $\mathbf{x} \in \mathcal{R}^m$ with respect to distribution Q as follows:

$$D(\mathbf{x}) = P\big(\mathbf{x} \in \Delta(\mathbf{X}_1, \ldots, \mathbf{X}_{m+1})\big), \quad \mathbf{x} \in \mathcal{R}^m,$$

where $\Delta(\mathbf{X}_1, \ldots, \mathbf{X}_{m+1})$ denotes the random simplex with vertices $\mathbf{X}_1, \ldots, \mathbf{X}_{m+1}$, i.e., the closed simplex with vertices $\mathbf{X}_1, \ldots, \mathbf{X}_{m+1}$. Note that $D(\mathbf{x})$ is the population simplicial depth of a point $\mathbf{x}$ and can be estimated by the sample simplicial depth $D_n(\mathbf{x})$ of point $\mathbf{x}$ given by the U-statistic

$$D_n(\mathbf{x}) = \binom{n}{m+1}^{-1} \sum_{1 \le i_1 < \ldots < i_{m+1} \le n} \mathbf{1}[\mathbf{x} \in \Delta(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_{m+1}})], \quad \mathbf{x} \in \mathcal{R}^m.$$

The depth function can be used to define the multivariate median and to give an ordering of data points in space from center outward. When additional information is available about the underlying distribution Q, tests or confidence sets based the usual sample depth do not utilize the information. We can use the developed JEL theory for vector U-statistics to incorporate side information. For a fixed number of constraints, we can construct the JEL with side information similar to Example 1. We now consider the case of *growing number of constraints*. Often we have partial information that the joint distribution, for example, two marginal distributions are independent or the marginal distribution is known. These are equivalent to an infinite number of constraints. Let us now use the latter as an example to illustrate our approach. Let the distribution of the first component X_1 of $\mathbf{X} = (X_1, \dots, X_m)^\top$ be known and equal to F_{10}. This implies

$$\int a_k \, dF_{10} = 0, \quad k = 1, 2, \dots, \tag{3.19}$$

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

$$\phi_k(t) = \sqrt{2}\cos(k\pi t), \quad t \in [0, 1], k = 1, 2, \dots. \tag{3.20}$$

This suggests us to look at the JEL with side information as follows:

$$\begin{aligned}\mathscr{R}_n(D, F_{10}) = \sup \Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{j=1}^n \pi_j (V_{nj} - D) = 0, \\ \sum_{j=1}^n \pi_j \phi_k(F_{10}(X_{1j})) = 0, \quad k = 1, \dots, r_n \Big\}, \ D \in \mathcal{R}^+,\end{aligned}$$

where X_{1j} is the first component of the jth observation $\mathbf{X}_j$. Here we have used the first r_n equations in (3.19). As before one defines $\mathcal{R}_n(D, F_{10})$ and obtains $\mathcal{R}_n(D, F_{10}) = \mathscr{R}_n(D, F_{10})/\mathscr{R}_n(F_{10})$. Let us assume $m \geq 2$ and at least one of the components in $X_2, \dots, X_m$ is nondegenerate, i.e. $P(X_d = c) < 1$ for some $d \geq 2$ and arbitrary constant c. Then by Theorem 3.2 one has

$$\frac{-2 \log \mathscr{R}_n(D_0, F_{10}) - (r_n + 1)}{\sqrt{2(r_n + 1)}} \Rightarrow \mathscr{N}(0, 1), \tag{3.21}$$

as both r_n and n tend to infinity such that r_n^3/n tends to zero, where $D_0 = D(\mathbf{x}_0)$. This shows $-2 \log \mathscr{R}_n(D_0, F_{10})$ is approximately chisquare-distributed with $r_n + 1$ degrees of freedom, whence $-2 \log \mathcal{R}_n(D_0, F_{10}) = -2 \log \mathscr{R}_n(D_0, F_{10}) + 2 \log \mathscr{R}_n(F_{10})$ is approximately chisquare-distributed with 1 degrees of freedom. The details are delayed to the last section.

TABLE 1
Simulated powers of the UJEL tests with side information of known marginal probabilities $p_{1\cdot}^{(0)} = .3$, $p_{\cdot 1}^{(0)} = .6$, sample size n and correlation coefficient ρ.

n	ρ	Chisq/Fisher	UJEL0	UJEL1	UJEL2
30	Ind	.045	.050	.050	.050
	-.7	.645	.480	.350	.405
	-.5	.325	.175	.120	.130
	-.3	.155	.055	.045	.050
	-.1	.050	.020	.020	.020
	.1	.045	.030	.035	.035
	.3	.060	.070	.070	.070
	.5	.245	.175	.155	.170
	.7	.555	.505	.420	.455
50	Ind	.045	.050	.050	.050
	-.7	.915	.890	.890	.895
	-.5	.510	.500	.495	.520
	-.3	.205	.185	.185	.195
	-.1	.050	.045	.040	.050
	.1	.055	.050	.045	.065
	.3	.130	.145	.145	.145
	.5	.425	.440	.440	.450
	.7	.780	.805	.805	.845
80	Ind	.050	.050	.050	.050
	-.7	.980	.980	.975	.985
	-.5	.700	.700	.700	.715
	-.3	.320	.320	.320	.335
	-.1	.055	.055	.050	.070
	.1	.030	.040	.035	.030
	.3	.235	.275	.270	.235
	.5	.610	.625	.620	.640
	.7	.945	.965	.960	.950

4. Simulation study. In this section, we report some simulation results on the U-statistic-based JEL (UJEL) tests about the independence of two categorical variables U, V given in Example 2. One of our intentions is the power improvement of the UJEL tests with the incorporation of side information, which is the knowledge about the marginal distributions of U, V given in (3.7) and (3.8). In practice, the national census data could be a source of such information and provide nearly exact information of moments of the marginal distributions of economic variables, see Imbens and Lancaster [13] and the references therein. Hence these tests can be used to detect the correlation between two categorical variables such as gender and salary, where the known marginal probabilities of gender and salary could be obtained from the census data. We looked at the usual Chisquare test

Table 2
Simulated powers of the UJEL tests with side information of known marginal probabilities $p_{1\cdot}^{(0)} = .3$, $(p_{\cdot 1}^{(0)}, p_{\cdot 2}^{(0)}) = (.25, .5)$, ***sample size*** n ***and correlation coefficient*** ρ.

n	ρ	Chisq/Fisher	UJEL0	UJEL1	UJEL2	UJEL3
30	Ind	.045	.050	.050	.045	.050
	-.7	.680	.660	.120	.510	.370
	-.5	.345	.305	.070	.285	.200
	-.3	.145	.135	.055	.165	.100
	-.1	.070	.105	.040	.095	.080
	.1	.095	.090	.030	.025	.065
	.3	.170	.170	.035	.010	.085
	.5	.400	.355	.075	.005	.180
	.7	.685	.730	.105	.040	.390
50	Ind	.055	.050	.050	.050	.050
	-.7	.920	.870	.865	.885	.905
	-.5	.505	.465	.460	.465	.490
	-.3	.235	.180	.175	.195	.200
	-.1	.055	.065	.070	.070	.080
	.1	.100	.090	.085	.110	.105
	.3	.265	.220	.225	.230	.235
	.5	.600	.540	.535	.560	.560
	.7	.895	.860	.855	.870	.875
80	Ind	.050	.050	.050	.050	.050
	-.7	.985	.980	.980	.980	.990
	-.5	.760	.750	.745	.770	.795
	-.3	.285	.295	.290	.315	.345
	-.1	.065	.055	.055	.070	.085
	.1	.050	.045	.040	.050	.055
	.3	.335	.320	.315	.335	.365
	.5	.770	.770	.765	.780	.795
	.7	.975	.980	.980	.980	.980
100	Ind	.050	.050	.050	.050	.050
	-.7	1	.995	.995	.995	1
	-.5	.885	.875	.875	.900	.895
	-.3	.460	.440	.435	.445	.440
	-.1	.105	.090	.090	.090	.095
	.1	.130	.120	.120	.135	.130
	.3	.415	.410	.410	.420	.430
	.5	.890	.885	.885	.880	.895
	.7	1	1	1	1	1

(or Fisher's exact test if appropriate) which neglect the side information, and several UJEL tests: UJELr : $r = 0, 1, \dots, 4$ which use r known marginal probabilities of U, V (r constraints), see Example 2 for more details.

The data were generated from the discretization of simulated independent

TABLE 3

Simulated powers of the UJEL tests with side information of known marginal probabilities $p_{1\cdot}^{(0)} = .3$, $(p_{\cdot 1}^{(0)}, p_{\cdot 2}^{(0)}, p_{\cdot 3}^{(0)}) = (.25, .2, .3)$, sample size n and correlation coefficient ρ.

n	ρ	Chisq/Fisher	JEL0	JEL1	JEL2	JEL3	JEL4
30	Ind	.065	.050	.050	.050	.050	.050
	-.7	.665	.625	.135	.510	.455	.260
	-.5	.310	.295	.070	.290	.325	.220
	-.3	.130	.105	.050	.165	.220	.135
	-.1	.080	.100	.055	.095	.145	.120
	.1	.095	.095	.075	.025	.085	.100
	.3	.215	.155	.080	.010	.080	.140
	.5	.415	.380	.100	.005	.095	.200
	.7	.705	.665	.145	.040	.110	.265
50	Ind	.040	.050	.050	.050	.050	.050
	-.7	.900	.875	.240	.500	.660	.600
	-.5	.505	.500	.140	.255	.390	.350
	-.3	.210	.200	.105	.130	.215	.170
	-.1	.060	.100	.100	.100	.130	.120
	.1	.065	.090	.055	.055	.065	.070
	.3	.265	.220	.075	.075	.050	.120
	.5	.570	.470	.145	.140	.060	.240
	.7	.925	.870	.245	.245	.110	.500
80	Ind	.050	.050	.050	.050	.050	.050
	-.7	.985	.980	.980	.980	.980	.980
	-.5	.785	.775	.775	.765	.785	.795
	-.3	.320	.355	.340	.340	.360	.365
	-.1	.060	.070	.070	.070	.075	.075
	.1	.050	.055	.045	.050	.050	.065
	.3	.365	.365	.365	.375	.370	.380
	.5	.785	.775	.775	.805	.815	.810
	.7	.985	.995	.995	.990	.990	.990
100	Ind	.050	.050	.050	.050	.050	.050
	-.7	.995	.995	.995	.995	1	1
	-.5	.875	.885	.880	.860	.880	.885
	-.3	.420	.435	.430	.430	.450	.445
	-.1	.115	.120	.120	.110	.105	.115
	.1	.125	.120	.115	.120	.115	.100
	.3	.420	.405	.395	.390	.400	.400
	.5	.890	.870	.870	.850	.865	.860
	.7	1	1	1	1	1	1

(the null hypothesis) and correlated bivariate t distributions with correlation coefficient $\rho = (2i+1)/10, i = -4, -3, \ldots, 3$ and degrees of freedom 3. For a fair comparison, each power of the tests was size-adjusted, that is, the critical values obtained from the simulated null distributions were used to determine whether a test value obtained under the alternative hypothesis was significant. In this way, all the type I errors of the tests were calibrated to be 0.05 so that the power comparison among them could be meaningful.

Reported in Tables 1, 2 and 3 are the simulated powers of the tests for different sample size n and repetitions 200. We observed for small sample sizes the classic Chi-square/Fisher's exact test were more powerful than the UJEL tests in most cases. This seemed to indicate that data of too small sample sizes in this example were not able to accommodate many constraints. But we also noticed that even when sample sizes were small or moderate the UJEL tests had higher power than the classical test in some cases of weak correlation. This was prominent in Tables 2 and 3. Specifically, the powers of the UJEL tests in Table 2 for $n = 30$ and $\rho = -0.1$ were higher than that of the classical test, and the UJEL2 and UJEL3 tests for $n = 50$ and $\rho = -0.1, 0.1$ were more powerful. Such phenomenon was even more noticeable for UJEL3 in Table 3 for $n = 30$ and $\rho = -0.5, -0.3, -0.1$ and the UJEL4 test for $\rho = 0.1$, where the gain in the power of UJEL3 test could be as large as almost 10% compared to that of the classical test. When the sample size was slightly larger, i.e. $n = 50$, the UJEL3 and UJEL4 tests were more powerful than the classical test for $\rho = -0.3, -0.1$ and $\rho = 0.1$, respectively. These seemed to indicate that for small or moderate samples the UJEL tests and additional information could be used to improve power for detecting weak relationships.

For $n = 50, 80$ in Table 1 and $n = 80$ in Tables 2 and 3, the UJEL tests with the incorporation of side information were more powerful than the Chisquare/Fisher's exact test. These indicated that data of large sample sizes were able explain more constraints, the UJEL tests and side information could be used to improve power. However, when sample size kept increasing ($n = 80$ in Table 1, $n = 100$ in Tables 2 and 3) the advantage of the UJEL tests and additional information may still exist but became less dominant. We think that when sample size was sufficiently large the side information incorporated into the testing procedure through marginal probabilities could be well recovered from the sample itself by the classical test. Hence in this case the incorporation of the known marginal probabilities did not bring much extra gain compared to the classical test.

An final note is that the powers of all the tests corresponding to small values of the coefficient $|\rho|$ (weak correlation) are very low as they are "close" to the null hypothesis of independence for a bivariate t distribution.

Discussion. As noted in the literature (e.g. Chen, *et al.* [3]; Emerson and Owen [8]), the empirical likelihood tests based on the asymptotic distribution for small sample sizes may yield larger type I error than the nominal level due to the convex hull constraint. To deal with such a problem and calibrate the type I error to the nominal level, Chen, *et al.* [3] proposed the addition of an artificial data point to the data set. In our simulations when type I errors were not calibrated using the finite sample critical values, the observed type I errors were also higher than the nominal value when the sample sizes were small in relation to the dimension. Therefore we applied the method of Chen, *et al.* in the simulations and it well shrank the type I errors to or much closer to the nominal level than the usual EL method. Hence in applications we suggest to use their method.

5. Asymptotic behaviors of the JEL with a growing number of estimated constraints. In this section, we shall consider the case that the kernel h is known but the constraint function $\mathbf{g}_n$ must estimated by some measurable function $\hat{\mathbf{g}}_n$. We allow the number of constraints to grow with the sample size and study the asymptotic behaviors of the JEL

$$\mathscr{R}_n(h, \hat{\mathbf{g}}_n) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \mathrm{Vec}\big(\tilde{V}_{nj}(h), \hat{\mathbf{g}}_n(Z_j)\big) = 0\Big\}.$$

To derive its asymptotic distribution, recall $\mathbf{C}_n, \mathbb{W}_n$ in (3.5) and set $\hat{\mathbb{W}}_n = n^{-1}\sum_{j=1}^n \hat{\mathbf{g}}_n(Z_j)^{\otimes 2}$. We have the following result with the proof delayed to the last section.

THEOREM 5.1. *Suppose $r_n\tilde{h}_1$ is Lindeberg. Suppose $\hat{\mathbf{g}}_n$ is an estimator of $\mathbf{g}_n$ such that*

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

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

for which $\mathscr{W}_n := \mathscr{W}(m^2\, Var(h_1(Z)), \mathbf{C}_n, \mathbb{W}_n)$ satisfies (R), and that

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

for some measurable function $\mathbf{u}_n$ from $\mathcal{Z}$ into $\mathcal{R}^{r_n}$ satisfying that $\int \mathbf{u}_n\, dQ = 0$ and $\|\mathbf{u}_n\|$ is Lindeberg. Assume further the dispersion matrix of $\mathscr{W}_n^{-1/2}\mathbf{v}_n(Z)$ with $\mathbf{v}_n = (m\tilde{h}_1, \mathbf{u}_n^\top)^\top$,

$$\mathbb{U}_n = \mathscr{W}_n^{-1/2} \int \mathbf{v}_n\mathbf{v}_n^\top\, dQ \mathscr{W}_n^{-1/2}, \tag{5.4}$$

satisfies $|\mathbb{U}_n|_o = O(1)$ *and* $r_n/\mathrm{trace}(\mathbb{U}_n^2) = O(1)$. *Then, as* r_n *tends to infinity with* n *such that* $r_n = o(n^{1/2})$,

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

Even though the asymptotic distribution of the JEL with estimated constraints in Theorem 5.1 is not a chisquare in general, it is still possible that it is approximately a chisquare and hence asymptotically distribution free. Below is such an example.

Example 3 (continued). Joint confidence sets for the mean and variance component. In a balanced one-way random effects model, we are interested in constructing joint confidence set for $\boldsymbol{\theta} = (\mu, \sigma_u^2)^\top$. Let us now motivate a U-statistic as a test statistic, see also Nobre, *et al.* [20]. We shall exploit MVUE's and "averaging out" j gives

$$U_{ii'} = \binom{J}{2}^{-1} \sum_{1\le j<j'\le J} 2^{-1}(Y_{ij} - Y_{i'j'})^2, \quad i, i' = 1, \dots, n.$$

For the between-treatment, $i \neq i'$, so $E((u_i - u_{i'})(\epsilon_{ij} - \epsilon_{i'j'})) = 0$, hence

$$E(U_{ii'}) = E\big(2^{-1}(u_i - u_{i'})^2 + 2^{-1}(\epsilon_{ij} - \epsilon_{i'j'})^2\big) = \sigma_u^2 + \sigma_\epsilon^2 := \sigma^2, \tag{5.5}$$

whereas for the within-treatment, $i = i'$, thus

$$E(U_{ii}) = E(2^{-1}(\epsilon_{i1} - \epsilon_{i2})^2) = \sigma_\epsilon^2.$$

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})^\top$ denote the observation vector in the i-th treatment. Obviously $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are i.i.d. Note that $U_{ii'} - 2^{-1}(U_{ii} + U_{i'i'})$ is a function of $\mathbf{Y}_i$ and $\mathbf{Y}_{i'}$ only, say $\kappa(\mathbf{Y}_i, \mathbf{Y}_{i'})$. Clearly $E(\kappa(\mathbf{Y}_i, \mathbf{Y}_{i'})) = \sigma_u^2$ for every pair (i, i') of subject indices with $i \neq i'$, so every such $\kappa(\mathbf{Y}_i, \mathbf{Y}_{i'})$ is an unbiased estimator of σ_u^2. Since $\kappa(\mathbf{y}_1, \mathbf{y}_2)$ is not argument-symmetric, we symmetrize it to get the argument-symmetric kernel $h(\mathbf{y}_1, \mathbf{y}_2) = 2^{-1}(\kappa(\mathbf{y}_1, \mathbf{y}_2) + \kappa(\mathbf{y}_2, \mathbf{y}_1)), \mathbf{y}_1, \mathbf{y}_2 \in \mathbf{R}^J$. Thus an unbiased estimator of σ_u^2 based on all the observations is the U-statistic with the kernel h given by

$$U_n(h) = \binom{n}{2}^{-1} \sum_{1\le i<i'\le n} h(\mathbf{Y}_i, \mathbf{Y}_{i'}).$$

It can be verified that the above kernel h and the kernel ψ in Example 3 satisfy the relationship

$$h(\mathbf{Y}_1, \mathbf{Y}_2) = \psi(\mathbf{X}_1, \mathbf{X}_2) + \binom{J}{2}^{-1} \sum_{j=1}^{J} 2^{-1}(Y_{1j} - Y_{1\cdot})(Y_{2j} - Y_{2\cdot}). \tag{5.6}$$

This yields that $h_1(\mathbf{y}) = E(h(\mathbf{y}, \mathbf{Y}_2))$ and $\psi_1(\mathbf{x}) = E(\psi(\mathbf{x}, \mathbf{X}_2))$ satisfy

$$h_1(\mathbf{y}) = \psi_1(\mathbf{x}), \quad \mathbf{y} = (y_1, \dots, y_J)^\top, \tag{5.7}$$

where $\mathbf{x}$ has components $x_1 = y.$ and $x_2 = (J-1)^{-1}\sum_{j=1}^{J}(y_j - y.)^2$.

Suppose there is available additional information about the model, for instance, ε as an i.i.d. copy of $\varepsilon_i = u_i + \epsilon_{i\cdot}$ is *symmetric* about zero. In this formulation, the model (3.11) becomes

$$Y_{i\cdot} = \mu + \varepsilon_i, \quad i = 1, \dots, n.$$

This is the well known symmetric location model. With symmetry as side information we now construct an empirical-likelihood-based confidence set for $\boldsymbol{\theta}$. To this end, let F denote the distribution function of ε, and $L_{2,0}(F, \mathrm{odd})$ be the subspace of $L_{2,0}(F)$ consisting of the odd functions. Assume F is continuous. Symmetry of ε about zero implies

$$E(a_k(\varepsilon)) = E(a_k(Y_{1\cdot} - \mu)) = 0, \quad k = 1, 2, \dots, \tag{5.8}$$

where a_k's is an orthonormal basis of $L_2(F, \mathrm{odd})$ and μ denotes the true value of parameter. Since ε and $-\varepsilon$ have an identical distribution, it follows

$$-\big(2F(-t)-1\big) = 1-2P(\varepsilon \le -t) = 1-2P(\varepsilon \ge t) = 1-2(1-F(t)) = 2F(t)-1.$$

This shows that $2F(t)-1$ is an odd function. Note that $\psi_k(t) = \sin(k\pi t), t \in [-1, 1], k = 1, 2, \dots$ is an orthonormal basis of $L_{2,0}(\mathcal{U}, \mathrm{odd})$ (the square-integrable odd functions with respect to the uniform measure $\mathcal{U}$ on $[-1, 1]$). Hence the composites $\psi_k(2F(t) - 1)$ is a basis of $L_{2,0}(F, \mathrm{odd})$ since the composite of two odd functions is odd. This justifies that we can take $a_k = \psi_k(2F(t) - 1)$. But F is unknown, we estimate it using the residuals $\varepsilon_i = Y_{i\cdot} - \mu_0, i = 1, \dots, n$ by the symmetrized empirical distribution function,

$$\mathbb{F}_{\mu_0}(t) = \frac{1}{2n}\sum_{i=1}^{n}\big(\mathbf{1}[Y_{i\cdot} - \mu_0 \le t] + \mathbf{1}[-(Y_{i\cdot} - \mu_0) < t]\big), \quad t \in \mathcal{R}.$$

Again we must justify $2\mathbb{F}_{\mu_0}(t) - 1$ is odd. This is easy to prove. Indeed,

$$\begin{aligned}-\big(2\mathbb{F}_{\mu_0}(-t) - 1\big) &= 1 - \frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{1}[\varepsilon_i \le -t] + \mathbf{1}[-\varepsilon_i < -t]\big)\\ &= 1 - \frac{1}{n}\sum_{i=1}^{n}\big(2 - \mathbf{1}[-\varepsilon_i < t] - \mathbf{1}[\varepsilon_i \le t]\big) = 2\mathbb{F}_{\mu_0}(t) - 1.\end{aligned}$$

This motivates us to utilize the first r_n equalities in (5.8) as constraints to construct the JEL with side information as follows:

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

where $V_{nj}(h)$'s are the jackknife pseudo values of the U-statistic $U_n(h)$. As before one can also define $\mathcal{R}_n(\mu,\sigma_u^2)$. We shall allow r_n to grow to infinity with the sample n such that r_n^4/n tends to zero. Suppose $r_n h_1$ is Lindeberg. Then by Theorem 5.1 one has

$$\frac{-2\log\mathscr{R}_n(\mu_0,\sigma_{u0}^2)-(r_n+1)}{\sqrt{2(r_n+1)}}\Rightarrow\mathscr{N}(0,1), \tag{5.9}$$

where $(\mu_0,\sigma_{u0}^2)\in\mathcal{R}\times\mathcal{R}^+$ denote the true values of parameter. The proof of (5.9) can be found in the last section. This shows that under the null hypothesis $-2\log\mathscr{R}_n(\mu_0,\sigma_{u0}^2))$ is approximately chisquare-distributed with r_n+1 degrees of freedom, whence $-2\log\mathcal{R}_n(\mu_0,\sigma_{u0}^2))$ is approximately chisquare-distributed with 1 degree of freedom. It is not difficult to calculate

$$\tilde{h}_1(\mathbf{Y}_1)=\binom{J}{2}^{-1}\sum_{1\le j<j'\le n}2^{-1}\big((Y_{1j}-\mu)(Y_{1j'}-\mu)-\sigma_u^2\big). \tag{5.10}$$

6. General results. In this section, we first introduce the notation we use throughout. We then state some results from Peng and Schick [27] which are tailored for our use. Based on these results, we prove Lemma 2.1 and a useful general theorem in the end.

6.1. *Notation.* Throughout we write boldface lower case letters for vectors. Write $\mathrm{Vec}(\mathbf{a},\mathbf{b})=(\mathbf{a}^\top,\mathbf{b}^\top)^\top$ for the stacking operation of vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{A}^{\otimes 2}=\mathbf{A}\mathbf{A}^\top$ and $\mathbf{A}\otimes\mathbf{A}$ for the Kronecker product of a vector or matrix $\mathbf{A}$. Let $\|\mathbb{A}\|$ denote the euclidean norm of a matrix $\mathbb{A}$ and $|\mathbb{A}|_o$ be the operator (or spectral) norm which are defined by

$$\|\mathbb{A}\|^2=\mathrm{trace}(\mathbb{A}^\top\mathbb{A})=\sum_{i,j}\mathbb{A}_{ij}^2,\quad |\mathbb{A}|_o=\sup_{\|u\|=1}|\mathbb{A}\mathbf{u}|=\sup_{\|\mathbf{u}\|=1}(\mathbf{u}^\top\mathbb{A}^\top\mathbb{A}\mathbf{u})^{1/2}.$$

Clearly the inequality $|\mathbb{A}|_o\le\|\mathbb{A}\|$ holds. Thus we have

$$|\mathbb{A}\mathbf{x}|\le|\mathbb{A}|_o\|\mathbf{x}\|\le\|\mathbb{A}\|\|\mathbf{x}\|$$

for compatible vectors $\mathbf{x}$. We should also point out that

$$|\mathbb{A}|_o = \sup_{\|\mathbf{u}\|=1} \sup_{\|\mathbf{v}\|=1} \mathbf{u}^\top \mathbb{A} \mathbf{v}$$

which simplifies to

$$|\mathbb{A}|_o = \sup_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbb{A} \mathbf{u}$$

if $\mathbb{A}$ is a nonnegative definite symmetric matrix. Using this and the Cauchy-Schwartz inequality it is easy to see that

$$\Big| \int \mathbf{f}^{\otimes 2} \, d\mu \Big|_o \le \int \|\mathbf{f}\|^2 \, d\mu, \tag{6.1}$$

whenever μ is a measure and $\mathbf{f}$ is a measurable function into $\mathcal{R}^s$ such that $\int \|\mathbf{f}\|^2 \, d\mu$ is finite.

6.2. *General results.* Let $\mathcal{T}_{n1}, \dots, \mathcal{T}_{nn}$ be r_n-dimensional random vectors. With them we associate the empirical likelihood

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{j=1}^n \pi_j \mathcal{T}_{nj} = 0 \Big\}.$$

To study the asymptotic behavior of $\mathscr{R}_n$ we introduce

$$\mathcal{T}_n^* = \max_{1 \le j \le n} \|\mathcal{T}_{nj}\|, \quad \bar{\mathcal{T}}_n = \frac{1}{n} \sum_{j=1}^n \mathcal{T}_{nj}, \quad \mathbb{S}_n = \frac{1}{n} \sum_{j=1}^n \mathcal{T}_{nj}^{\otimes 2},$$

and

$$\mathcal{T}_n^{(\nu)} = \sup_{\|u\|=1} \frac{1}{n} \sum_{j=1}^n (u^\top \mathcal{T}_{nj})^\nu, \quad \nu = 3, 4.$$

Let $\lambda_n = \lambda_{\min}(\mathbb{S}_n)$ and $\Lambda_n = \lambda_{\max}(\mathbb{S}_n)$. We first quote Theorem 6.1 of Peng and Schick [27] below for our use.

LEMMA 6.1. *Let $r_n = r$ for all n. Suppose*

$$\mathcal{T}_n^* = o_p(n^{1/2}), \quad n^{1/2} \bar{\mathcal{T}}_n \Rightarrow \mathcal{T}, \quad \mathbb{S}_n = \mathscr{S} + o_p(1) \tag{6.2}$$

for some random vector $\mathcal{T}$ and $r \times r$ positive definite matrix $\mathscr{S}$. Then

$$-2 \log \mathscr{R}_n \Rightarrow \mathcal{T}^\top \mathscr{S}^{-1} \mathcal{T}.$$

Peng and Schick [27] also introduced the following conditions.

(A1) $\mathcal{T}_n^* = o_p(r_n^{-1/2} n^{1/2})$.
(A2) $\|\bar{\mathcal{T}}_n\| = O_p(r_n^{1/2} n^{-1/2})$.
(A3) There is a sequence of regular $r_n \times r_n$ dispersion matrices $\{\mathscr{S}_n\}$ such that

$$|\mathbb{S}_n - \mathscr{S}_n|_o = o_p(r_n^{-1/2}).$$

(A4) $\mathcal{T}_n^{(3)} = o_p(r_n^{-1} n^{1/2})$ and $\mathcal{T}_n^{(4)} = o_p(r_n^{-3/2} n)$.

They also gave the following inequalities

$$\mathcal{T}_n^{(3)} \le \lambda_{\max}(\mathbb{S}_n)\mathcal{T}_n^*, \quad \mathcal{T}_n^{(4)} \le \lambda_{\max}(\mathbb{S}_n)(\mathcal{T}_n^*)^2. \tag{6.3}$$

These imply that a sufficient condition for (A1) and (A4) is

$$\mathcal{T}_n^* = o_p(r_n^{-1} n^{1/2}). \tag{6.4}$$

They studied the case when r_n increases with the increasing sample size n. The following is quoted from Theorem 6.2 of Peng and Schick [27].

LEMMA 6.2. *Let (A1)–(A4) hold. Suppose that r_n increases with n to infinity and that there are $r_n \times r_n$ dispersion matrices $\mathbb{V}_n$ such that $r_n/\mathrm{trace}(\mathbb{V}_n^2) = O(1)$ and*

$$\frac{n\bar{\mathcal{T}}_n^\top \mathbb{W}_n^{-1}\bar{\mathcal{T}}_n - \mathrm{trace}(\mathbb{V}_n)}{\sqrt{2\mathrm{trace}(\mathbb{V}_n^2)}} \Rightarrow \mathscr{N}(0,1). \tag{6.5}$$

Then

$$\frac{-2\log \mathscr{R}_n - \mathrm{trace}(\mathbb{V}_n)}{\sqrt{2\mathrm{trace}(\mathbb{V}_n^2)}} \Rightarrow \mathscr{N}(0,1). \tag{6.6}$$

6.3. *General results for U-statistics with side information.* We now apply Lemma 6.1 to derive the asymptotic behaviors of the JEL for U-statistics with side information. Recall that the kernel h is square-integrable. Here we further assume throughout that h is nondegenerate, that is, $\mathrm{Var}(h_1(Z)) > 0$. Let $\mathbf{T}_{n1}, \dots, \mathbf{T}_{nn}$ be r_n-dimensional random vectors. With them we associate the JEL for the U-statistic with side information as follows:

$$\mathscr{R}_n(h) = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \mathrm{Vec}(\tilde{V}_{nj}(h), \mathbf{T}_{nj}) = 0\Big\},$$

where $\tilde{V}_{nj}(h)$'s are the centered jackknife pseudo values of the U-statistic $U_n(h)$. Recall the remainder R_{nj} in (2.3) and use some algebra to express

$$R_{nj} = \sum_{c=2}^m \binom{m}{c}\Big(cU_{(n-1)(c-1)}(h^*_{(c-1)j}) - (c-1)U^{(-j)}_{(n-1)c}(h^*_c)\Big), \tag{6.7}$$

where $h^*_{(c-1)j}(z_1,\dots,z_{c-1}) := h^*_c(Z_j, z_1,\dots,z_{c-1})$. Using now the inequality $(\sum_{j=1}^m a_j)^2 \le m\sum_{j=1}^m a_j^2$ for real numbers a_j, we derive

$$E(R_{n1}^2) \le 2m\sum_{c=2}^m \binom{m}{c}^2 \Big(c^2\mathrm{Var}(U_{(n-1)(c-1)}(h^*_{(c-1)1})) + (c-1)^2\mathrm{Var}(U^{(-1)}_{(n-1)c}(h^*_c))\Big).$$

It is well known that the variances of the above U-statistics satisfy (see e.g. p. 189, Serfling [29])

$$\mathrm{Var}\big(U_{(n-1)(c-1)}(h^*_{(c-1)1})\big) = O(n^{-c+1}), \quad \mathrm{Var}\big(U^{(-1)}_{(n-1)c}(h^*_c)\big) = O(n^{-c}).$$

Thus by (2.2) we conclude (2.4) and prove Lemma 2.1. Moreover,

$$\sum_{j=1}^n \big(\tilde V_{nj} - m\tilde h_1(Z_j)\big)^2 = O_p(1), \tag{6.8}$$

as the expected value of the above sum is $O(1)$. Let us prepare with the following result.

Lemma 6.3. *Suppose that $\mathbf{T}_{n1},\dots,\mathbf{T}_{nn}$ are $\mathcal{R}^{r_n}$-valued random vectors such that (A1) – (A4) are met. Suppose that the jackknife pseudo values $V_{n1},\dots,V_{nn}$ satisfy (A1) and (A4). Assume that there exists some r_n-dimensional vector $\mathbf{C}_n$ such that*

$$\|\frac{1}{n}\sum_{j=1}^n \tilde V_{nj}\mathbf{T}_{nj} - \mathbf{C}_n\| = o_p(r_n^{-1/2}), \tag{6.9}$$

and that the sequence of matrices $\mathscr{W}_n = \mathscr{W}\big(Var(m\tilde h_1(Z)), \mathbf{C}_n, \mathbb{W}_n\big)$ satisfies (R). Then $\mathcal{T}_{nj} = \mathrm{Vec}(\tilde V_{nj}, \mathbf{T}_{nj})$ satisfy (A1)–(A4) with $\mathscr{S}_n = \mathscr{W}_n$ as r_n tends to infinity such that $r_n = o(n)$.

Proof. As both $\mathbf{T}_{nj}$ and V_{nj} satisfy (A1) and (A4), so do $\mathcal{T}_{nj}$, while they satisfy (A2) as $\mathbf{T}_{nj}$ satisfy (A2) and in view of (2.6) and $n\mathrm{Var}(U_n) = O(1)$. By Cauchy inequality for real numbers a_j and b_j,

$$\big|\frac{1}{n}\sum_{j=1}^n (a_j^2 - b_j^2)\big|^2 \le \frac{1}{n}\sum_{j=1}^n (a_j-b_j)^2 \frac{1}{n}\sum_{j=1}^n 2(a_j^2+b_j^2).$$

Applying this with $a_j = \tilde{V}_{nj}$ and $b_j = m\tilde{h}_1(Z_j)$ and in view of (6.8), we get

$$\Big|\frac{1}{n}\sum_{j=1}^{n}\big(\tilde{V}_{nj}^2 - m^2 E(\tilde{h}_1(Z)^2)\big)\Big| = O_p(n^{-1/2}). \tag{6.10}$$

This, (6.9) and the fact that $\mathbf{T}_{nj}$ satisfy (A3) yield

$$\Big|\frac{1}{n}\sum_{j=1}^{n}\mathcal{T}_{nj}^{\otimes 2} - \mathscr{W}_n\Big|_o = o_p(r_n^{-1/2}). \tag{6.11}$$

This shows $\mathcal{T}_n$ satisfies (A3) and completes the proof. □

Consider now the case of a fixed number $r_n = r$ of constraints. We have the following.

THEOREM 6.1. *Let $r_n = r$ for all n. Suppose*

$$\mathbf{T}_n^* = o_p(n^{1/2}), \quad \frac{1}{n}\sum_{j=1}^{n} m\tilde{h}_1(Z_j)\mathbf{T}_{nj} \xrightarrow{P} \mathbf{C}, \quad \frac{1}{n}\sum_{j=1}^{n}\mathbf{T}_{nj}^{\otimes 2} \xrightarrow{P} \mathbb{W} \tag{6.12}$$

for some r-dimensional vector $\mathbf{C}$ and $r\times r$ matrix $\mathbb{W}$ such that $\mathscr{W} := \mathscr{W}(\mathit{Var}(mh_1(Z)), \mathbf{C}, \mathbb{W})$ is nonsingular. Assume

$$n^{-1/2}\sum_{j=1}^{n}\Big(m\tilde{h}_1(Z_j), \mathbf{T}_{nj}^\top\Big)^\top \Rightarrow \mathcal{T}, \tag{6.13}$$

for some $(r+1)$-dimensional random vector $\mathcal{T}$. Then

$$-2\log \mathscr{R}_n(h) \Rightarrow \mathcal{T}^\top \mathscr{W}^{-1}\mathcal{T}.$$

PROOF. We shall apply Lemma 6.1 to prove the result by verifying its three conditions in (6.2) with $\mathcal{T}_{nj} = (\tilde{V}_{nj}, \mathbf{T}_{nj}^\top)^\top$. First by Markov's inequality and in view of (2.4), we derive for any $\epsilon > 0$,

$$\begin{aligned} P\Big(\max_{1\le j\le n}|\tilde{V}_{nj}| > n^{1/2}\epsilon\Big) &\le \sum_{j=1}^{n} P\left(|\tilde{V}_{nj}| > n^{1/2}\epsilon\right) \\ &\le \epsilon^{-2} E\left(|\tilde{V}_{n1}|^2 \mathbf{1}[|\tilde{V}_{n1}| > n^{1/2}\epsilon]\right) \to 0, \quad n\to\infty. \end{aligned}$$

Thus

$$\max_{1\le j\le n}|\tilde{V}_{nj}| = o_p(n^{1/2}). \tag{6.14}$$

This and the first equality in (6.12) imply the first condition in (6.2). It is well known (e.g. page 188, Serfling (1980)) that

$$U_n(h) - \theta = \frac{m}{n}\sum_{j=1}^{n} \tilde{h}_1(Z_j) + O(n^{-1}). \tag{6.15}$$

This, (2.6) and (6.13) yield the second condition in (6.2). We now verify that $\mathbf{T}_{nj}$ satisfy (A1) – (A4). Note first that $\mathscr{W}$ is nonsingular hence the submatrix $\mathbb{W}$ is also nonsingular. Applying the inequalities in (6.3) to $\mathbf{T}_{nj}$ and noticing that $r = r_n$ is fixed and (6.4) is a sufficient condition for (A1) and (A4), we derive by the first equality in (6.12) that $\mathbf{T}_{nj}$ satisfy (A1) and (A4). It follows from (6.13) that $\mathbf{T}_{nj}$ satisfy (A2), while $\mathbf{T}_{nj}$ satisfying (A3) follows from the third equality in (6.12). We show next that (6.9) is also met with $\mathbf{C}_n$ equal to the $\mathbf{C}$ given in (6.12). This, in fact, follows from the limit of the second component in (6.13), (6.8), and the following inequalities

$$\begin{aligned}\|\frac{1}{n}\sum_{j=1}^{n} \tilde{V}_{nj}\mathbf{T}_{nj} - m\tilde{h}_1(Z_j)\mathbf{T}_{nj}\|^2 &\le \frac{1}{n}\sum_{j=1}^{n}(\tilde{V}_{nj} - m\tilde{h}_1(Z_j))^2 \frac{1}{n}\sum_{j=1}^{n}\|\mathbf{T}_{nj}\|^2 \\ = O_p(n^{-1})\mathrm{trace}\Big(\frac{1}{n}\sum_{j=1}^{n}\mathbf{T}_{nj}^{\otimes 2}\Big) &= O_p(n^{-1})(\mathrm{trace}(\mathbb{W}) + o_p(1)) = O_p(n^{-1}).\end{aligned}$$

Thus the conditions of Lemma 6.3 are met and hence $\mathcal{T}_{nj}$ satisfy (A1)–(A4), in particular, satisfy the third condition in (6.2). We now apply Lemma 6.1 to complete the proof. □

7. Proofs. In this section, we provide proofs of several theorems and the details for the examples introduced in the previous sections.

For $a \in \mathcal{R}$, $\mathbf{c}_r \in \mathcal{R}^r$ and $r \times r$ identity matrix $\mathbb{I}_r$, let $\mathbb{M}_{r+1}$ be the $(r+1) \times (r+1)$ matrix defined by $\mathbb{M}_{r+1} = \mathscr{W}(a^2, \mathbf{c}_r, \mathbb{I}_r)$, where $\mathscr{W}$ is the matrix operation defined in (3.4). Denote the determinant of $\mathbb{M}$ by $|\mathbb{M}|$. Using Laplace's formula to express the determinant of a matrix in terms of its minors and mathematical induction we can easily prove the following.

LEMMA 7.1. *For $\lambda \in \mathcal{R}$ and integer $r \ge 1$, the characteristic polynomial of $\mathbb{M}_{r+1}$ is given by*

$$|\mathbb{M}_{r+1} - \lambda\mathbb{I}_{r+1}| = (1-\lambda)^{r-1}\big(\lambda^2 - (1+a^2)\lambda + a^2 - \|\mathbf{c}_r\|^2\big). \tag{7.1}$$

Thus the sequence of matrices $\mathbb{M}_{r+1}$ satisfies (R) if $c^2 =: \lim_{r\to\infty}\|\mathbf{c}_r\|^2 = \sum_{i=1}^{\infty} c_i^2 < \infty$ such that $b^2 = a^2 - c^2 > 0$.

PROOF. Note first that $\mathbb{M}_{r+1}$ has three distinct roots, $\lambda = 1$ with multiplicity $r-1$ and

$$\lambda_1 = (1 + a^2 + \sqrt{\Delta})/2, \quad \lambda_2 = (1 + a^2 - \sqrt{\Delta})/2,$$

where $\Delta = (a^2-1)^2 + 4\sum_{i=1}^{r} c_i^2$. Since $0 \le \Delta \le \delta =: (a^2-1)^2 + 4c^2$, it follows $0 < 1 + a^2 \le 2\lambda_1 \le 1 + a^2 + \sqrt{\delta} < \infty$ and

$$0 < 2b^2/(1 + a^2 + \sqrt{\delta}) \le \lambda_2 \le (1 + a^2)/2 < \infty.$$

This shows that $\mathbb{M}_{r+1}$ has $r+1$ eigen values which bounded away from both zero and infinity uniformly in $r = 1, 2, \dots,$ hence the sequence of matrices $\mathbb{M}_{r+1}$ satisfies (R). □

REMARK 7.1. Let $\mathcal{H}_1$ and $\mathcal{H}$ be two Hilbert spaces such that $\mathcal{H}_1$ is a true subspace of $\mathcal{H}$. Let $a_k : k = 1, 2, \dots$ be an orthonormal basis of $\mathcal{H}_1$. For $\varphi \in \mathcal{H}$, the projection φ_p of φ onto $\mathcal{H}_1$ is given by the Fourier series $\varphi_p = \sum_{k=1}^{\infty} c_k a_k$, where c_k are the Fourier coefficients. Suppose $\varphi \notin \mathcal{H}_1$. By the Hilbert space theory (see e.g. Theorem 4.13, Conway [6]), $\|\varphi_p\|^2 = \sum_{k=1}^{\infty} c_k^2 < \|\varphi\|^2$. Since a_k is orthonormal, the $r \times r$ matrix whose (i,j)-entry is the inner product of a_i and a_j is the $r \times r$ identity matrix $\mathbb{I}_r$. Consequently, it follows from Lemma 7.1 that the sequence of matrices $\mathscr{W}(\|\varphi\|^2, \mathbf{c}_r, \mathbb{I}_r), r = 1, 2, \dots$ satisfies (R).

Let $\boldsymbol{\phi}_n = (\phi_1, \dots, \phi_{r_n})^\top$ where ϕ_k is the trigonometric basis given in (3.20). Since these basis functions are bounded by $\sqrt{2}$, we see that $\boldsymbol{\phi}(t)$ and its derivative $\boldsymbol{\phi}'(t)$ satisfy

$$\|\boldsymbol{\phi}_n(t)\|^2 \le 2r_n, \quad \|\boldsymbol{\phi}'(t)\|^2 \le 2\pi^2 r_n^3, \quad t \in [0,1]. \tag{7.2}$$

PROOF OF (3.9). To show this, let us first calculate Σ. Write h^{kl} for the (k,l)-th entry of $\mathbf{h}$. One easily computes

$$h_1^{kl}(\mathbf{z}_1) = E(h^{kl}(\mathbf{Z}_1, \mathbf{Z}_2)|\mathbf{Z}_1 = \mathbf{z}_1) = 2^{-1}(x_{1k} - p_{k\cdot})(y_{1l} - p_{\cdot l}) + 2^{-1}\delta_{kl},$$

thus the centered version is $\tilde{h}_1^{kl}(\mathbf{z}_1) = 2^{-1}\big((x_{1k} - p_{k\cdot})(y_{1l} - p_{\cdot l}) - \delta_{kl}\big)$. Let $\boldsymbol{\alpha} = (p_{1\cdot}, \dots, p_{(K-1)\cdot})^\top$ and $\boldsymbol{\beta} = (p_{\cdot 1}, \dots, p_{\cdot (L-1)})^\top$ be the distributions of U and V respectively. Under the null the components of $\boldsymbol{\delta}$ satisfy $\delta_{kl} = 0$, hence the variances and covariances of the components of $\mathbf{X}$ are given by

$$\mathrm{Cov}(X_k, X_k) = \alpha_k - \alpha_k^2, \quad \mathrm{Cov}(X_k, X_{k'}) = -\alpha_k \alpha_{k'}, \quad k \ne k'.$$

Similar formulas for Y_l also hold. Let $\mathbf{A} = \mathrm{Diag}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}^{\otimes 2}$ and $\mathbf{B} = \mathrm{Diag}(\boldsymbol{\beta}) - \boldsymbol{\beta}^{\otimes 2}$. One easily verifies

$$E\big(\tilde{h}_1^{kl}(\mathbf{Z}_1)\tilde{h}_1^{k'l'}(\mathbf{Z}_1)\big) = 4^{-1}\mathrm{Cov}(X_k, X_{k'})\mathrm{Cov}(Y_l, Y_{l'}).$$

These yield $\Sigma = \mathrm{Var}(2\mathbf{h}_1(\mathbf{Z}_1)) = 4E\big(\tilde{\mathbf{h}}_1(\mathbf{Z}_1)^{\otimes 2}\big) = \mathbf{A} \otimes \mathbf{B}$. Hence in view of $|\Sigma| = |\mathbf{A}|^{K-1}|\mathbf{B}|^{L-1}$ we see that Σ is nonsingular if and only if both $\mathbf{A}$ and $\mathbf{B}$ are nonsingular, which is equivalent to (3.9). To prove this, it suffices to show $|\mathbf{A}| = \prod_{k=1}^{K} \alpha_k$ where $\alpha_K = p_{K\cdot}$, whereas the same also holds for $\mathbf{B}$. This can be verified by Laplace's formula and mathematical induction. □

Proof of (3.21). We shall apply Theorem 3.2 to prove the result. Note first that the kernel is $h(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{1}[x_0 \in \Delta(\mathbf{X}_1, \ldots, \mathbf{X}_m)]$ so that $h_1(\mathbf{x}) = P(x_0 \in \Delta(\mathbf{x}, \mathbf{X}_2, \ldots, \mathbf{X}_m)), \mathbf{x} \in \mathbf{R}^m$ which is bounded by 1. Also $\mathbf{g}_n = \boldsymbol{\phi}_n \circ F_{10}$ hence $\|\mathbf{g}_n\| \le \sqrt{2r_n}$ by (7.2). Since $r_n^3 = o(n)$, it follows that $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ are Lindeberg. We are now left to show the regularity (R). Recall that F is the distribution of $\mathbf{X} = (X_1, \ldots, X_m)^\top$ and F_{10} is the distribution of X_1. Since there is at least one component in $X_2, \ldots, X_m$ that is nondegenerate and $m \ge 2$, it follows $\mathcal{H}_1 = L_{2,0}(F_{10})$ is a true subspace of $\mathcal{H} = L_{2,0}(F)$. Clearly $\tilde{h}_1$ lives in $\mathcal{H}$ but not in $\mathcal{H}_1$. It follows from Remark 7.1 that (R) holds. This completes the proof. □

Proof of (5.9). We shall prove this by applying Theorem 5.1 with

$$\mathbf{g}_n(y) = \boldsymbol{\psi}_n(2F(y - \mu_0) - 1) \quad \text{and} \quad \hat{\mathbf{g}}_n(y) = \boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(y - \mu_0) - 1)$$

with $\mathbb{W}_n = \mathbb{I}_{r_n}$ and $\mathbf{C}_n = E(2\tilde{h}_1(\mathbf{Y}_1)\mathbf{g}_n(J^{-1}\mathbf{1}^\top\mathbf{Y}_1))$, where $\boldsymbol{\psi}_n = (\psi_1, \ldots, \psi_{r_n})^\top$. By assumption, $r_n h_1$ is Lindeberg. Using $\|\mathbf{g}_n\| \le \sqrt{r_n}$, we derive (5.1) in view of $r_n^3 = o(n)$. We now show $\mathscr{W}_n = \mathscr{W}(4\mathrm{Var}(h_1(\mathbf{Y}_1)), \mathbf{C}_n, I_{r_n})$ satisfies (R). Let D be the common distribution function of ϵ_{1j} and G be the joint distribution of $Y_{1j} - \mu_0, j = 1, \ldots, J$. Then by the independence between the random effect u_1 and error ε_{1j} we derive that the distribution of ε_1 satisfies $F(t) = P(\varepsilon_1 \le t) = P(u_1 + \epsilon_{1\cdot} \le t) = E\big(D(t - u_1)\big), t \in \mathcal{R}$ and

$$G(\mathbf{t}) = P(u_1 + \epsilon_{1j} \le t_j,\ j = 1, \ldots, J) = E\Big(\prod_{j=1}^{J} D(t_j - u_1)\Big), \quad \mathbf{t} \in \mathcal{R}^J.$$

Since F is continuous and $J \ge 2$, it follows $F \ne G$. Thus the Hilbert space $\mathcal{H}_1 = L_{2,0}(F)$ is a true subspace of $\mathcal{H} = L_{2,0}(G)$. Furthermore, by (5.10) $\tilde{h}_1 \in \mathcal{H}$ but $\tilde{h}_1 \notin \mathcal{H}_1$. Thus from Remark 7.1 it follows that the matrices $\mathscr{W}_n$ satisfies (R). Let us now prove the first equality in (5.2). To this end, note first the inequalities

$$\|\boldsymbol{\psi}_n\|^2 \le r_n, \quad \|\boldsymbol{\psi}_n'\| \le \sqrt{2}\pi r_n^{3/2}, \quad |2(\mathbf{t}^\top\boldsymbol{\psi}_n')(\mathbf{t}^\top\boldsymbol{\psi}_n)| \le 4\pi r_n^2\|\mathbf{t}\|^2. \tag{7.3}$$

We break

$$
\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_{ni}\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i)) - \mathbf{C}_n\|^2 \\
&\le \|\frac{1}{n}\sum_{i=1}^{n}\big(\tilde{V}_{ni} - 2\tilde{h}_1(\mathbf{Y}_i)\big)^2 \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i))\|^2\| \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}4\tilde{h}_1(\mathbf{Y}_i)^2\frac{1}{n}\sum_{i=1}^{n}\big(\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i)) - \boldsymbol{\psi}_n(2F(\varepsilon_i))\big)^2 \\
&\quad + \|\frac{1}{n}\sum_{i=1}^{n}\big(2\tilde{h}_1(\mathbf{Y}_i)\boldsymbol{\psi}_n(2F(\varepsilon_i)) - E\big[2\tilde{h}_1(\mathbf{Y}_1)\boldsymbol{\psi}_n(2F(\varepsilon_1))\big]\big)\|^2 \\
&:= A_n + B_n + C_n.
\end{aligned}
$$

By (6.8) and in view of (7.3) and $r_n^2 = o(n)$, we derive $A_n \le O_p(n^{-1})r_n = o_p(r_n^{-1})$, while by the square-integrability of h_1 and in view $r_n^4 = o(n)$, we obtain

$$
B_n \le O_p(1)2\pi^2 r_n^3 4 \sup_{-\infty<t<\infty}|\mathbb{F}_{\mu_0}(t) - F(t)|^2 = O_p(r_n^3/n) = o_p(r_n^{-1}).
$$

Now it is not difficult to calculate

$$
\begin{aligned}
E(C_n) &= n^{-1}E\big[\mathrm{Var}\big(2\tilde{h}_1(\mathbf{Y}_1)\|\boldsymbol{\psi}_n(2F(\varepsilon_1))\|\big)\big] \\
&\le n^{-1}E\big(4\tilde{h}_1(\mathbf{Y}_1)^2\|\boldsymbol{\psi}_n(2F(\varepsilon_1))\|^2\big) \\
&\le n^{-1}E\big(4\tilde{h}_1(\mathbf{Y}_1)^2\big)r_n = O(r_n n^{-1}) = o(r_n^{-1}).
\end{aligned}
$$

This shows the first equality of (5.2). We prove below that (5.3) hold with $\mathbf{u}_n = \mathbf{g}_n$, that is,

$$
n^{-1/2}\sum_{i=1}^{n}\big(\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i) - 1) - \boldsymbol{\psi}_n(2F(\varepsilon_i) - 1)\big) = o_p(1). \tag{7.4}
$$

Clearly $\int \mathbf{u}_n\, dQ = 0$ and $\|\mathbf{u}_n\|$ is Lindeberg in view of $\|\mathbf{u}_n\| \le \sqrt{r_n}$. Thus $\mathbf{v}_n = (2h_1, \mathbf{u}_n^\top)^\top$, $\int \mathbf{v}_n^{\otimes 2}\, dQ = \mathscr{W}_n$, and $\mathbb{U}_n = \mathbb{I}_{r_n+1}$, which implies $|\mathbb{U}_n|_o = 1$ and $r_n/\mathrm{trace}(\mathbb{U}_n^2) = r_n/(r_n+1) = O(1)$. We are now left to prove (7.4) and the second equality in (5.2) which is implied by

$$
\sup_{\|\mathbf{t}\|=1}\Big|\frac{1}{n}\sum_{i=1}^{n}(\mathbf{t}^\top\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i) - 1))^2 - 1\Big| = o_p(r_n^{-1/2}). \tag{7.5}
$$

Since $\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}-1)$ and $\boldsymbol{\psi}_n(2F_{\mu_0}-1)$ are odd functions, the above two equations can be written as

$$\sup_{\|\mathbf{t}\|=1}\Big|\frac{1}{n}\sum_{i=1}^n(\mathbf{t}^\top\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(|\varepsilon_i|)-1))^2-1\Big|=o_p(r_n^{-1/2}), \tag{7.6}$$

$$n^{-1/2}\sum_{i=1}^n \operatorname{sign}(\varepsilon_i)\big(\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(|\varepsilon_i|)-1)-\boldsymbol{\psi}_n(2F(|\varepsilon_i|)-1)\big)=o_p(1). \tag{7.7}$$

Also, we have almost surely the identity

$$\begin{aligned}2\mathbb{F}_{\mu_0}(|\varepsilon_k|)-1&=\frac{1}{n}\sum_{i=1}^n\big(\mathbf{1}[\varepsilon_i\le|\varepsilon_k|]+\mathbf{1}[-\varepsilon_i<|\varepsilon_k|]-1\big)\\&=\frac{1}{n}\sum_{i=1}^n\big(\mathbf{1}[|\varepsilon_i|\le|\varepsilon_k|]-\frac{1}{n}\sum_{i=1}^n\mathbf{1}[\varepsilon_i=|\varepsilon_k|]\big)\\&=(R_k-\mathbf{1}[\varepsilon_k\ge 0])/n,\end{aligned}$$

where $R_1,\dots,R_n$ are the ranks of $|\varepsilon_1|,\dots,|\varepsilon_n|$. Using the bounds (7.3) and $r_n^3=o(n)$, it is sufficient for us to prove (7.6) and (7.7) with $2\mathbb{F}_{\mu_0}(|\varepsilon_i|)-1$ replaced by R_i/n. Let a be a Lipschitz function on $[0,1]$ with Lipschitz constant L. Then we approximate the sum by an integral as follows:

$$\frac{1}{n}\sum_{i=1}^n a(R_i/n)=\frac{1}{n}\sum_{i=1}^n a(i/n)=\int_0^1 a(x)\,dx+\sum_{i=1}^n\int_{(i-1)/n}^{i/n}(a(i/n)-a(x))\,dx$$

and therefore

$$\Big|\frac{1}{n}\sum_{i=1}^n a(R_i/n)-\int_0^1 a(x)\,dx\Big|\le L/n.$$

For $a=(\mathbf{u}^\top\boldsymbol{\psi}_n)^2$ and noting $\int(\mathbf{u}^\top\boldsymbol{\psi}_n(x))^2\,dx=1$ as $\psi_1,\psi_2,\dots$ are also orthonormal with respect to the uniform measure on $[0,1]$, we get

$$\Big|\frac{1}{n}\sum_{i=1}^n(\mathbf{u}^\top\boldsymbol{\psi}_n)^2(R_i/n)-1\Big|\le 4\pi r_n^2/n=o_p(r_n^{-1}).$$

This shows (7.6). Let

$$\mathbf{T}_n=n^{-1/2}\sum_{i=1}^n\operatorname{sign}(\varepsilon_i)\big(\boldsymbol{\psi}_n(R_i/n)-\boldsymbol{\psi}_n(2F(|\varepsilon_i|)-1)\big).$$

Since ε is symmetric, it follows that $\mathrm{sign}(\varepsilon)$ and $|\varepsilon|$ are independent, $\mathrm{sign}(\varepsilon)$ is uniformly distributed on $\{-1,1\}$ and $|\varepsilon|$ has distribution given by $G(t) = 2F(t)-1, t \in \mathcal{R}^+$. From this we immediately derive

$$\begin{aligned} E\big(\|\mathbf{T}_n\|^2 \big| |\varepsilon_1|,\dots,|\varepsilon_n|\big) &= \frac{1}{n}\sum_{i=1}^n \|\boldsymbol{\psi}_n(R_i/n) - \boldsymbol{\psi}_n(G(|\varepsilon_i|)\|^2 \\ &= 2\pi^2 r_n^3 \frac{1}{n}\sum_{i=1}^n |R_i/n - G(|\varepsilon_i|)|^2 = O_p(r_n^3/n). \end{aligned}$$

This shows $\mathbf{T}_n = o_p(1)$ and hence the desired (7.7). □

Proof of Theorem 3.1. We apply Theorem 6.1 with $\tilde{V}_{nj}(h) = \tilde{V}_{nj}(h^{(1)})$ and $\mathbf{T}_{nj} = (\tilde{V}_{nj}(h^{(2)}),\dots,\tilde{V}_{nj}(h^{(r)}))^\top$ so that $\mathrm{Vec}(\tilde{V}_{nj}(h), \mathbf{T}_{nj}) = \tilde{\mathbf{V}}_{nj}(\mathbf{h})$. Set $\mathbf{m} = (m_1,\dots,m_r)^\top$ and $\mathbf{h}_1 = (h_1^{(1)},\dots,h_1^{(r)})^\top$ and $\mathbf{w} = \mathbf{m}\tilde{\mathbf{h}}_1$. By (6.8) and the Cauchy inequality, we derive

$$\begin{aligned} &\Big|\frac{1}{n}\sum_{j=1}^n \Big(\tilde{V}_{nj}(h^{(k)})\tilde{V}_{nj}(h^{(l)}) - m_k h_1^{(k)}(Z_j) m_l h_1^{(l)}(Z_j)\Big)\Big|^2 \\ &\quad\le 2\frac{1}{n}\sum_{j=1}^n \Big(\tilde{V}_{nj}(h^{(k)}) - m_k h_1^{(k)}(Z_j)\Big)^2 \frac{1}{n}\sum_{j=1}^n \tilde{V}_{nj}(h^{(l)})^2 \\ &\quad + 2\frac{1}{n}\sum_{j=1}^n \big(m_k h_1^{(k)}(Z_j)\big)^2 \frac{1}{n}\sum_{j=1}^n \Big(\tilde{V}_{nj}(h^{(l)}) - m_l h_1^{(l)}(Z_j)\Big)^2 \\ &= O_p(n^{-1}) = o_p(1), \quad k,l = 1,\dots,r. \end{aligned}$$

Hence by the law of large numbers we get

$$\frac{1}{n}\sum_{j=1}^n \tilde{\mathbf{V}}_{nj}(\mathbf{h})^{\otimes 2} \xrightarrow{P} \mathrm{Var}(\mathbf{m}\mathbf{h}_1) = E\big(\mathbf{w}(Z)^{\otimes 2}\big),$$

which establishes the second and third equalities of (6.12). If $\mathrm{Var}(\mathbf{m}\mathbf{h}_1)$ is nonsingular, then an application of the central limit theorem gives

$$n^{-1/2}\sum_{j=1}^n \mathbf{w}(Z_j) \Rightarrow \mathscr{N}(0, \mathrm{Var}(\mathbf{m}\mathbf{h}_1)),$$

which yields (6.13). Moreover, an analogous argument to (6.14) yields the first equality in (6.12). We now apply Theorem 6.1 to complete the proof. □

Proof of Theorem 5.1. We verify that the conditions of Lemma 6.2 are satisfied with $\mathcal{T}_{nj} = \hat{\mathbf{w}}_{nj} = \mathrm{Vec}(\tilde{V}_{nj}(h), \hat{\mathbf{g}}_n(Z_j))$. Note first that as $r_n\tilde{h}_1$ is Lindeberg one has

$$r_n \max_{1\le j\le n} |\tilde{V}_{nj}| = o_p(n^{1/2}) \tag{7.8}$$

in view of $r_n^2 = o(n)$. Indeed, for $\epsilon > 0$,

$$P(r_n \max_{1\le j\le n} |\tilde{V}_{nj}| > n^{1/2}\epsilon) \le P(r_n \max_{1\le j\le n} |\tilde{V}_{nj} - m\tilde{h}_1(Z_j)| > n^{1/2}\epsilon/2) + P(r_n \max_{1\le j\le n} |m\tilde{h}_1(Z_j)| > n^{1/2}\epsilon/2).$$

By the Lindeberg property (L1), the last probability converges to zero, whereas the second probability is bounded by

$$\sum_{j=1}^n P(r_n|\tilde{V}_{nj} - m\tilde{h}_1(Z_j)| > n^{1/2}\epsilon/2) \le \frac{4}{\epsilon^2} r_n^2 E(|\tilde{V}_{n1} - m\tilde{h}_1(Z_1)|^2) = \frac{4}{\epsilon^2}\frac{r_n^2}{n},$$

which converges to zero as n tends to infinity, where the last equality follows from (2.5). This proves (7.8) hence (A1) in view of (5.1). With the aid of (6.10), we conclude (A3) from (5.2). The equality (6.15), (2.6), (5.3) in which $\mathbf{v}_n = (m\tilde{h}_1, \mathbf{u}_n^\top)^\top$ and $r_n = o(n)$ imply

$$\frac{1}{n}\sum_{j=1}^n \hat{\mathbf{w}}_{nj} = \frac{1}{n}\sum_{j=1}^n \mathbf{v}_n(Z_j) + o_p(n^{-1/2}). \tag{7.9}$$

Clearly $\int \mathbf{v}_n\, dQ = 0$ and $\mathbf{v}_n$ is Lindeberg by (L0) as $\mathbf{u}_n$ and h_1 are Lindeberg. Let $\xi_{nj} = \mathscr{W}_n^{-1/2}\mathbf{v}_n(Z_j)$ and set

$$\bar{\mathbf{v}}_n = \frac{1}{n}\sum_{j=1}^n \mathbf{v}_n(Z_j) \quad \text{and} \quad \bar{\mathbf{T}}_n = \frac{1}{n}\sum_{j=1}^n \hat{\mathbf{w}}_n(Z_j).$$

It follows from (C) that $|\mathscr{W}_n^{1/2}|_o + |\mathscr{W}_n^{-1/2}|_o = O(1)$. Using this and the Lindeberg property of $\|\mathbf{v}_n\|$ we derive

$$L_n(\epsilon) = E\left(\|\xi_{n1}\|^2 \mathbf{1}[\|\xi_{n1}\| > \epsilon\sqrt{n}]\right) \to 0, \quad \epsilon > 0.$$

Note that $\mathrm{trace}(\mathbb{U}_n) \le r_n|\mathbb{U}_n|_o = O(r_n)$. Then we have $\mathrm{trace}(\mathbb{U}_n)/\mathrm{trace}(\mathbb{U}_n^2) \le |\mathbb{U}_n|_o r_n/\mathrm{trace}(\mathbb{U}_n^2) = O(1)$ and conclude $\mathrm{trace}(\mathbb{U}_n^2) \to \infty$. Thus Theorem 2 in Peng and Schick [26] yields

$$\frac{n\bar{\mathbf{v}}_n^\top \mathscr{W}_n^{-1}\bar{\mathbf{v}}_n - \mathrm{trace}(\mathbb{U}_n)}{\sqrt{2\mathrm{trace}(\mathbb{U}_n^2)}} \Rightarrow \mathscr{N}(0,1).$$

Next we calculate

$$nE\left(\|\bar{\mathbf{v}}_n\|^2\right) = E\left(\|\mathbf{v}_n(Z)\|^2\right) \le |\mathscr{W}_n^{1/2}|_o^2 E\big(\|\mathscr{W}_n^{-1/2}\mathbf{v}_n(Z)\|^2\big) \le |\mathscr{W}_n^{1/2}|_o^2 \mathrm{trace}(\mathbb{U}_n).$$

This shows that $n\|\bar{\mathbf{v}}_n\|^2 = O_p(r_n)$. Thus we derive with the help of (7.9) and $r_n/\mathrm{trace}(\mathbb{U}_n^2) = O(1)$, that $n\|\bar{\mathbf{T}}_n\|^2 = O_p(r_n)$ and

$$\frac{n\bar{\mathbf{T}}_n\mathscr{W}_n^{-1}\bar{\mathbf{T}}_n - \mathrm{trace}(\mathbb{U}_n)}{\sqrt{2\mathrm{trace}(\mathbb{U}_n^2)}} \Rightarrow \mathscr{N}(0,1).$$

Thus conditions (A1)–(A4) hold with $\mathcal{T}_{nj} = \hat{\mathbf{w}}_{nj}$ in view of (6.4) and $\mathcal{T}_n^* = o_p(r_n^{-1}n^{1/2})$. The desired result now follows from Lemma 6.2. □

Proof of Theorem 3.2. We shall prove the result by applying Theorem 5.1 with $\hat{\mathbf{g}} = \mathbf{g}$. It follows immediately that (5.3) automatically holds with $\mathbf{u}_n = \mathbf{g}_n$, whence $\mathbf{v}_n = \mathrm{Vec}(m\tilde{h}_1, \mathbf{g}_n)$ yields $\int \mathbf{v}_n^{\otimes 2}\, dQ = \mathscr{W}_n$ and hence $\mathbb{U}_n = \mathscr{W}_n^{-1/2} \int \mathbf{v}_n\mathbf{v}_n^\top\, dQ \mathscr{W}_n^{-\top/2} = I_{r_n+1}$ satisfies the required conditions. As $r_n\|\mathbf{g}_n\|$ is Lindeberg, we see that (5.1) holds in view of the Lindeberg property (L1). Next we show the Lindeberg property (L1) also implies (5.2). In fact, as $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ are Lindeberg, it follows from (L0) that $\mathbf{v}_n$ is also Lindeberg. Fix $\epsilon > 0$. Let $\mathbf{t}_n = \mathbf{v}_n\mathbf{1}[\|(r_n+1)\mathbf{v}_n\| \le \epsilon\sqrt{n}]$ so that $\mathbf{s}_n = \mathbf{v}_n - \mathbf{t}_n = \mathbf{v}_n\mathbf{1}[\|(r_n+1)\mathbf{v}_n\| > \epsilon\sqrt{n}]$. Set

$$\bar{\mathscr{W}}_{n,1} = \frac{1}{n}\sum_{j=1}^n \mathbf{t}_n^{\otimes 2}(Z_j), \quad \bar{\mathscr{W}}_{n,2} = \frac{1}{n}\sum_{j=1}^n \mathbf{s}_n^{\otimes 2}(Z_j).$$

As $\mathscr{W}_n$ satisfies (R), one has $\lambda_{\max}(\mathscr{W}_n) \le B$ for some $B > 0$ and all n, so that

$$E(\|\mathbf{v}_n\|^2(Z)) = \mathrm{trace}(E(\mathbf{v}_n^{\otimes 2}(Z))) = \mathrm{trace}(\mathscr{W}_n) \le B(r_n+1).$$

Then we find

$$nE[\|\bar{\mathscr{W}}_{n,1} - E[\bar{\mathscr{W}}_{n,1}]\|^2] = \sum_{i=1}^{r_n+1}\sum_{k=1}^{r_n+1} \mathrm{Var}(t_{n,i}(Z)t_{n,k}(Z))$$
$$\le E[\|\mathbf{t}_n\|^4(Z)] \le \frac{\epsilon^2 n}{(r_n+1)^2}E[\|\mathbf{v}_n\|^2(Z)] \le \frac{\epsilon^2 nB(r_n+1)}{(r_n+1)^2},$$

and

$$P(\bar{\mathscr{W}}_{n,2} \ne 0) \le P\big(\max_{1\le j\le n} \|(r_n+1)\mathbf{v}_n(Z_j)\| > \epsilon\sqrt{n}\big) \to 0,$$

and using (6.1),

$$|E[\bar{\mathscr{W}}_{n,2}]|_o \le E[\|\mathbf{v}_n\|^2(Z)\mathbf{1}[\|(r_n+1)\mathbf{v}_n(Z)\| > \epsilon\sqrt{n}]] = o(r_n^{-2}).$$

From these inequalities it immediately yields the desired (5.2). □

References.

[1] Arcones M.A. (1996). The Bahadur-Kiefer Representation for U-Quantiles. *Ann. Statist.* **24**: 1400 – 1422.

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

[3] Chen J, Variyath A., Abraham B. (2008). Adjusted Empirical Likelihood and its Properties. *J. Comp. Graph. Stat.* **17(2)**, 426-443.

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

[5] Chen, S.X. and Van Keilegom, I. (2009). A goodness-of-fit test for parametric and semiparametric models in multiresponse regression. *Bernoulli* **15**: 955–976.

[6] Conway, J.B. (1985). *A Course in Functional Analysis.* Springer-Verlag, New York Inc.

[7] DiCiccio, T.J., Hall, P. and Romano, J.P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**: 1053–1061.

[8] Emerson, S. and Owen, A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* **3**: 1161–1192.

[9] Fang, F., Hong, Q. and Shao, J. (2009). A pseudo empirical likelihood approach for stratified samples with nonresponse. *Ann. Statist.* **37**: 371–393.

[10] Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Review*, **58**: 109–127.

[11] Heffernan, P. M. (1997). Unbiased Estimation of Central Moments by using U-statistics. *J. R. Statist. Soc. B* **59**: 861 – 863.

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

[13] Imbens G, Lancaster T. (1994). Combining Micro and Macro Data in Microeconometric Models. *Rev. Econ. Stud.* **61(4)**, 655-680.

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

[15] Kolaczyk, E.D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4**: 199–218.

[16] Kowalski, J. and Tu, X.M. (2008). *Modern Applied U-Statistics.* John Wiley & Sons, Inc.

[17] Li, G. and Wang, Q.-H. (2003). Empirical likelihood regression analysis for right censored data. *Statist. Sinica* **13**: 51–68.

[18] Liu, Y. and Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *Ann. Statist.* **38**: 1341–1362.

[19] Lu, X. (2009). Empirical likelihood for heteroscedastic partially linear models. *J. Multivariate Anal.* **100**: 387–396.

[20] Nobre, J. S., Singer, J. M. and Silvapulle, M. J. (2008). U-tests for variance components in one-way random effects models. *Beyond Parametric in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen* Vol. **1**: 197 – 210.

[21] Lee, A. J. (1990). *U-Statistics: Theory and Practice.* Marcel Dekker, Inc.

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

[23] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**: 237–249.

[24] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**: 90–120.

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

[26] Peng, H. and Schick, A. (2013a). Asymptotic normality of quadratic forms with random vectors of increasing dimension. The *Manuscript* can be found at the URL: `https://math.iupui.edu/~hpeng/preprints_hp.html`.
[27] Peng, H. and Schick, A. (2013c). An empirical likelihood approach of goodness of fit testing. *Bernoulli* **19**: 954–981
[28] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**: 30–325.
[29] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc.
[30] Shi, J. and Lau, T.S. (2000). Empirical likelihood for partially linear models. *J. Multiv. Anal.* **72**: 132–148.
[31] Shi, X. (1984). Approximate Independence of Jackknife Pseudo Values and the Bootstrap Method. *Journal of Wuhan Institute of Hydra-Electric Engineering* (Chinese) **2**, 83 – 90.
[32] Wang, Q.-H. and Jing, B.-Y. (2003). Empirical likelihood for partially linear models. *Ann. Inst. Statist. Meth.* **55**: 585–595.
[33] Yuan, A., He, W., Wang, B. and Qin, G. (2012). U-statistics with side information. *J. Multiv. Analy.* **111**: 20 – 38.

Indiana University Purdue University Indianapolis
Department of Mathematical Sciences
Indianapolis, IN 46202-3267, USA
E-mail: hpeng@math.iupui.edu, E-mail: qlin@math.iupui.edu,
E-mail: ftan@math.iupui.edu, E-mail: weizheng@math.iupui.edu