Improving Estimation Efficiency In Structural Equation Models By An Easy Empirical Likelihood Approach

ShanWang and Hanxiang Peng

University of San Francisco
Department of Mathematics and Statistics
San Francisco, CA 94117, USA
swang151@usfca.edu

Indiana University Purdue University Indianapolis
Department of Mathematical Sciences
Indianapolis, IN 46202-3267, USA
hanxpeng@iu.edu

Abstract: In this article, we construct empirical likelihood (EL)-weighted estimators of linear functionals of a probability measure in the presence of side information. Motivated by nuisance parameters in semiparametric models with possibly infinite dimension, we consider the use of estimated constraint functions and allow the number of constraints to grow with the sample size. We study the asymptotic properties and efficiency gains. The results are used to construct improved estimators of parameters in structural equations models. The EL-weighted estimators of parameters are shown to have reduced variances in a SEM in the presence of side information of stochastic independence of the random error and random covariate. Some simulation results on efficiency gain are reported.

AMS 2000 subject classifications: Primary 62D05; secondary 62J05, 62F12, 62F40.
Keywords and phrases: Estimated constraints; Infinitely many constraints; Maximum empirical likelihood estimators; Side information; Structrual equation models.

1. Introduction

Structure equation models (SEM) is a popular multivariate technique for analyzing data in behavioral, medical and social sciences. It is an analysis of moment structures in which the variance-covariance matrix $\Sigma = \mathrm{Var}(\mathbf{Z})$ of a random vector $\mathbf{Z} \in \mathcal{R}^p$ is specified by a parametric matrix function, $\Sigma = \Sigma(\boldsymbol{\vartheta})$, $\boldsymbol{\vartheta} \in \Theta$ for some subset Θ of $\mathcal{R}^q$. Given independent and identically distributed (i.i.d.) observations $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ of $\mathbf{Z}$, one focuses on estimating the parameter vector $\boldsymbol{\vartheta}$. A moment-type estimator of $\boldsymbol{\vartheta}$ is based on the criterion of *minimum discrepancy function (MDF)*, see e.g. Shapiro (2007). For a $p \times p$ matrix $\mathbf{M}$, denote by vecs($\mathbf{M}$) the $p(p+1)/2$-dimensional vector formed by stacking its columns of the upper triangular matrix. Let Ξ be a subset of $\mathcal{R}^{p(p+1)/2}$ consisting of vecs($\mathbf{M}$) over all $p \times p$ semi-positive definite matrices $\mathbf{M}$. A function F on $\Xi \times \Xi$ is called a *discrepancy function* if it satisfies: (i) $F(\mathbf{t}, \boldsymbol{\xi}) \geq 0$ for all $\mathbf{t}, \boldsymbol{\xi} \in \Xi$, (ii)

$F(\mathbf{t}, \boldsymbol{\xi}) = 0$ if and only if $\mathbf{t} = \boldsymbol{\xi}$, (iii) $F(\mathbf{t}, \boldsymbol{\xi})$ is twice continuously differentiable, and (iv) For any fixed $\mathbf{s} \in \Xi$, $F(\mathbf{t}, \boldsymbol{\xi}) \to \infty$ as $\mathbf{t} \to \mathbf{s}$ and $\boldsymbol{\xi} \to \infty$. We shall abuse notation to write $F(\mathbf{A}, \mathbf{B}) = F(\text{vecs}(\mathbf{A}), \text{vecs}(\mathbf{B}))$ for matrices $\mathbf{A}, \mathbf{B}$.

Often Σ is estimated by the unstructured sample variance-covariance matrix,

$$\mathbb{S}_n = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^\top, \tag{1.1}$$

where $\bar{\mathbf{Z}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Z}_i$ is the sample version of $\boldsymbol{\mu} = E(\mathbf{Z})$. Let $\mathbf{s}_n = \text{vecs}(\mathbb{S}_n)$ and $\boldsymbol{\sigma} = \text{vecs}(\Sigma)$. The MDF estimator of $\boldsymbol{\vartheta}$ is any value $\boldsymbol{\vartheta}_n$ in Θ that satisfies

$$F(\mathbf{s}_n, \sigma(\boldsymbol{\vartheta}_n)) = \inf_{\boldsymbol{\vartheta} \in \Theta} F(\mathbf{s}_n, \boldsymbol{\sigma}(\boldsymbol{\vartheta})). \tag{1.2}$$

One commonly used discrepancy function is the maximum-likelihood (ML) discrepancy function of two matrix-valued variables given by

$$F_{ML}(\mathbf{S}, \Sigma) = \log |\Sigma| - \log |\mathbf{S}| + \text{trace}(\mathbf{S}\Sigma^{-1}) - p, \tag{1.3}$$

where $\mathbf{S}, \Sigma$ are positive definite matrices. Another is the generalized least squares (GLS) discrepancy function given by

$$F_{GLS}(\mathbf{S}, \Sigma) = \text{trace}((\mathbf{S} - \Sigma)\mathbf{W}^{-1}(\mathbf{S} - \Sigma)\mathbf{W}^{-1}), \tag{1.4}$$

where $\mathbf{W}$ is a symmetric matrix, for instance, $\mathbf{W} = \mathbb{S}_n$ and $\mathbf{W} = \mathbf{I}_p$. Suppose that there is available some information about SEM that can be expressed by a vector equation of expectation,

$$E(\mathbf{g}(\mathbf{Z})) = 0, \tag{1.5}$$

where $\mathbf{g}$ is some measurable function taking values in $\mathcal{R}^m$. While the moment estimator $\mathbb{S}_n$ completely ignores the information, the empirical likelihood (EL)-weighted estimator $\tilde{\mathbb{S}}_n$ utilizes the information to provide an improved estimator,

$$\tilde{\mathbb{S}}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^\top}{1 + \mathbf{g}(\mathbf{Z}_i)^\top \tilde{\boldsymbol{\zeta}}}, \tag{1.6}$$

where $\tilde{\boldsymbol{\zeta}}$ is a solution to the equation

$$\sum_{i=1}^{n} \frac{\mathbf{g}(\mathbf{Z}_i)}{1 + \mathbf{g}(\mathbf{Z}_i)^\top \tilde{\boldsymbol{\zeta}}} = 0. \tag{1.7}$$

Accordingly, an improved MDF estimator $\tilde{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ is any value in Θ that satisfies

$$F(\tilde{\mathbf{s}}_n, \sigma(\tilde{\boldsymbol{\vartheta}}_n)) = \inf_{\boldsymbol{\vartheta} \in \Theta} F(\tilde{\mathbf{s}}_n, \sigma(\boldsymbol{\vartheta})), \tag{1.8}$$

where $\tilde{\mathbf{s}}_n = \text{vecs}(\tilde{\mathbb{S}}_n)$.

In many semiparamatric models, $\mathbf{g}(\mathbf{z})$ involves in nuisance parmeters which must be estimated, leading to a plug-in estimator $\hat{\mathbf{g}}(\mathbf{z})$. Using it, we work with

$$\hat{\mathbb{S}}_n = \frac{1}{n}\sum_{i=1}^{n} \frac{(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^\top}{1 + \hat{\mathbf{g}}^\top(\mathbf{Z}_i)\hat{\boldsymbol{\zeta}}}, \tag{1.9}$$

where $\hat{\boldsymbol{\zeta}}$ is the solution to Eqt (1.7) by replacing $\mathbf{g}(\mathbf{Z}_i) = \hat{\mathbf{g}}(\mathbf{Z}_i)$. As a result, an imporved MDF estimator $\hat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}$ is any value in Θ that satisfies

$$F(\hat{\mathbf{s}}_n, \sigma(\hat{\boldsymbol{\vartheta}}_n)) = \inf_{\boldsymbol{\vartheta}\in\Theta} F(\hat{\mathbf{s}}_n, \sigma(\boldsymbol{\vartheta})), \tag{1.10}$$

where $\hat{\mathbf{s}}_n = \text{vecs}(\hat{\mathbb{S}}_n)$.

The improved MDF estimator $\tilde{\boldsymbol{\vartheta}}_n$ is more efficient than the usual MDF estimator $\boldsymbol{\vartheta}_n$. The efficiency criteria used are that of a least dispersed regular estimator or that of a locally asymptotic minimax estimator, and are based on the convolution theorems and on the lower bounds of the local asymptotic risk in LAN and LAMN families, see the monograph by Bickel, *et al.* (1993).

The side information contained in (1.5) is carried by the EL-weights $(n(1 + \mathbf{g}(\mathbf{Z}_i)^\top\tilde{\boldsymbol{\zeta}}))^{-1}$ based on the principle of the maximum empirical likelihood. There is an extensive amount of literature on the empirical likelihood. It was introduced by Owen [23, 24] to construct confidence intervals in a nonparametric setting. Soon it was used to construct point estimators. Qin and Lawless [30] studied maximum empirical likelihood estimators (MELE). Bravo [3] studied a class of M-estimators based on generalized empirical likelihood with side information and showed that the resulting class of estimators is efficient in the sense that it achieves the same asymptotic lower bound as that of the efficient GMM estimator with the same side information. Parente and Smith (2011 [25]) investigated generalized empirical likelihood estimators for irregular constraints.

Estmators of the preceding EL-weighted form were investigated in Zhang [38, 39] in M-estimation and quantile processes in the presence of auxiliary information. Hellerstein and Imbens [12] exploited such estimators for the least squares estimators in a linear regression model. Yuan *et al.* [37] explored such estimators in U-statistics. Tang and Leng [36] utilized the form to construct improved estimators of parameters in quantile regression. Asymptotic properties of the EL-weighted estimators were obtained for a finite number of known constraints.

Motivated by nuisance parameters in semiparametric models and the infinite dimension of such models, Peng and Schick [29] considered the use of estimated constraint functions and studied a growing number of constraints in MELE. Wang and Peng (2022) used the EL-weighted approach to construct efficient estimators of linear functionals of a probability measure in the presence of side information for two cases, viz, known marginal distributions and equal but unknowned marginals, each of which is equivalent to infinitely many constraints.

MELE enjoy high efficiency and is particularly convenient to incorporate side information. Just like any other optimization problems, however, it is not trivial to numerically find MELE especially for a large number of constraints. Peng and Schick [29] employed one-step estimators to construct MELE. The EL-weighing approach reduces the number of constraints and are thus computationally easier than general MELE.

The rest of the article is organized as follows. In Section 2, we shall construct the EL-weighted estimator of the linear functional of a probability measure in the presence of side information which is expressed by an finite or infinite number of known or estimated constraints, and present the asymptotic properties. In Section 3, we give examples of side information and study the asymptotic properties of the improved estimators in SEM. The form of SEM can be extended to great extent in a variety of ways. We shall focus on the extensions that have been described in Bollen (1989) as well as in the LISREL software manual (Joreskog and Sorbom, 1996). The components present in a general SEM are a path analysis, the conceptual synthesis of latent variable and measurement models, and general estimation procedures. In SEM, only information up to the second moments is used, while other forms of information such higher order moments, independence or symmetry of the random errors are ignored, which can be used by the EL-weighting method to improve efficiency. In Section 4, we report simulation results. Technical details are collected in Section 5.

2. The main results and side information

In this section, we give the main results and discuss side information.

2.1. The main results

Suppose that $Z_1, \dots, Z_n$ are i.i.d. random variables with a common distribution Q taking values in a measurable space $\mathcal{Z}$. We are interested in efficient estimation of the linear functional $\boldsymbol{\theta} = \int \boldsymbol{\psi}\, dQ$ of Q for some square-integrable function $\boldsymbol{\psi}$ from $\mathcal{Z}$ to $\mathcal{R}^r$ when side information is available through

(C) $\mathbf{u}$ is a measurable function from $\mathcal{Z}$ to $\mathcal{R}^m$ such that $\int \mathbf{u}\, dQ = 0$ and the variance-covariance matrix $\mathbf{W} = \int \mathbf{u}\mathbf{u}^\top dQ$ is nonsingular.

To utilize the information contained (C), consider the the empirical likelihood,

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \mathbf{u}(Z_j) = 0 \Big\},$$

where $\mathscr{P}_n = \{\pi \in [0, 1]^n : \sum_{j=1}^n \pi_j = 1\}$ is the unit probability simplex. Following Owen, one uses Lagrange multipliers to get the maximizers,

$$\tilde{\pi}_j = \frac{1}{n} \frac{1}{1 + \mathbf{u}(Z_j)^\top \tilde{\boldsymbol{\zeta}}}, \quad j = 1, \dots, n, \tag{2.1}$$

where $\tilde{\boldsymbol{\zeta}}$ is the solution to the equation

$$\frac{1}{n}\sum_{j=1}^{n}\frac{\mathbf{u}(Z_j)}{1+\mathbf{u}(Z_j)^\top\tilde{\boldsymbol{\zeta}}}=0. \tag{2.2}$$

These $\tilde{\pi}_j$'s incorporte the side information, and a natural estimator of $\boldsymbol{\theta} = \int \boldsymbol{\psi}\, dQ$ is the EL-weighted estimator,

$$\tilde{\boldsymbol{\theta}}=\sum_{j=1}^{n}\tilde{\pi}_j\boldsymbol{\psi}(Z_j)=\frac{1}{n}\sum_{j=1}^{n}\frac{\boldsymbol{\psi}(Z_j)}{1+\mathbf{u}(Z_j)^\top\tilde{\boldsymbol{\zeta}}}. \tag{2.3}$$

For $\boldsymbol{\psi}_{\mathbf{t}}(\mathbf{z}) = \mathbf{1}[\mathbf{z} \le \mathbf{t}]$ for fixed $\mathbf{t} \in \mathcal{R}^p$, one obtains the distribution function $\theta = P(\mathbf{Z} \le \mathbf{t})$. For $\psi(\mathbf{z}) = z_1 \cdots z_p$, $\theta = E(Z_1 \cdots Z_p)$ is the mixed moment.

Write $\|\mathbf{a}\|$ for the euclidean norm of $\mathbf{a}$ and $\mathbf{a}\otimes\mathbf{b}$ for the Kronecker product of $\mathbf{a}$ and $\mathbf{b}$. For $\mathbf{x} = (x_1, \dots, x_p), \mathbf{y} = (y_1, \dots, y_p)$, write $\mathbf{x} \le \mathbf{y}$ for $x_1 \le y_1, \dots, x_p \le x_p$. Let $L_2^m(Q) = \{\mathbf{f} = (f_1, \dots, f_m)^\top : \int \|\mathbf{f}\|^2\, dQ^m < \infty\}$, and let $L_{2,0}^m(Q) = \{\mathbf{f} \in L_2^m(Q) : \int \mathbf{f}\, dQ^m = 0\}$. For $\mathbf{f} \in L_2^m(Q)$, write $[\mathbf{f}]$ for the closed linear span of the components $f_1, \dots, f_m$ in $L_2(Q)$. Let Z be an i.i.d. copy of Z_1. Let $\boldsymbol{\phi}_0$ be the projection of $\boldsymbol{\psi}$ onto the closed linear span $[\mathbf{u}]$ of $\mathbf{u}$, so that $\boldsymbol{\phi}_0 = \Pi(\boldsymbol{\psi}|[\mathbf{u}]) = E(\boldsymbol{\psi}(Z) \otimes \mathbf{u}^\top(Z))\mathbf{W}^{-1}\mathbf{u}$. Let $\Sigma_0 = \mathrm{Var}(\boldsymbol{\psi}(Z)) - \mathrm{Var}(\boldsymbol{\phi}_0(Z))$. We now give the first result with the proof delayed.

Theorem 1. *Assume (C) with m fixed. Then $\tilde{\boldsymbol{\theta}}$ given in (2.3) satisfies the stochastic expansion,*

$$\tilde{\boldsymbol{\theta}} = \bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\phi}}_0 + o_p(n^{-1/2}), \tag{2.4}$$

Thus if $\Sigma_0 = \mathrm{Var}(\boldsymbol{\psi}(Z)) - \mathrm{Var}(\boldsymbol{\phi}_0(Z))$ is nonsingular then $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal with mean zero and asymptotic covariance matrix Σ_0, that is,

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Longrightarrow \mathscr{N}(0, \Sigma_0).$$

Theorem 1 exhibits that the EL-weighted estimator $\tilde{\boldsymbol{\theta}}$ has a smaller asymptotic variance than that of the sample mean $\bar{\boldsymbol{\psi}}$, and the amount of reduction is $\mathrm{Var}(\boldsymbol{\phi}_0(Z))$. It is, in fact, the MELE of $\boldsymbol{\theta}$.

Remark 1. Haberman (1984) studied minimum Kullback-Leibler divergence -type estimators for the linear functionals of a probability measure, and more general problems involving a fixed number of side information. The EL-weighted estimator $\tilde{\boldsymbol{\theta}}$ in Theorem 1 is asymptotically equivalent to Haberman's estimator, see his page 976. This shows that Haberman's estimator is semiparametrically efficient.

In semiparametric models, the constraint function $\mathbf{u}$ contains nuisance parameters and must be estimated. Let $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_m)^\top$ be an estimate of $\mathbf{u}$. With it we now work with the EL-weights

$$\hat{\pi}_j = \frac{1}{n}\frac{1}{1+\hat{\mathbf{u}}(Z_j)^\top\hat{\boldsymbol{\zeta}}}, \quad j = 1, \dots, n, \tag{2.5}$$

where $\hat{\boldsymbol{\zeta}}$ is the solution to the equation (2.2) with $\mathbf{u} = \hat{\mathbf{u}}$. In the same fashion, a natural estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \sum_{j=1}^{n} \hat{\pi}_j \boldsymbol{\psi}(Z_j) = \frac{1}{n}\sum_{j=1}^{n} \frac{\boldsymbol{\psi}(Z_j)}{1 + \hat{\mathbf{u}}(Z_j)^\top \hat{\boldsymbol{\zeta}}}. \tag{2.6}$$

Set $\hat{\mathbf{W}} = n^{-1}\sum_{j=1}^{n} \hat{\mathbf{u}}\hat{\mathbf{u}}^\top(Z_j)$. Let $|\mathbf{W}|_o$ denote the spectral norm (largest eigenvalue) of a matrix $\mathbf{W}$. We have

Theorem 2. *Assume (C) with m fixed. Let $\hat{\mathbf{u}}$ be an estimator of $\mathbf{u}$ such that*

$$\max_{1\le j\le n} \|\hat{\mathbf{u}}(Z_j)\| = o_p(n^{1/2}), \tag{2.7}$$

$$|\hat{\mathbf{W}} - \mathbf{W}|_o = o_p(1), \tag{2.8}$$

$$\frac{1}{n}\sum_{j=1}^{n} \big(\boldsymbol{\psi}(Z_j) \otimes \hat{\mathbf{u}}(Z_j) - E\big(\boldsymbol{\psi}(Z_j) \otimes \hat{\mathbf{u}}(Z_j)\big)\big) = o_p(1), \tag{2.9}$$

and that there exists some measurable function $\mathbf{v}$ that satisfies (C) such that

$$\frac{1}{n}\sum_{j=1}^{n} E\left(\|\hat{\mathbf{u}}(Z_j) - \mathbf{v}(Z_j)\|^2\right) = o(1), \tag{2.10}$$

$$\frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{u}}(Z_j) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{v}(Z_j) + o_p(n^{-1/2}). \tag{2.11}$$

Then $\hat{\boldsymbol{\theta}}$ given in (2.6) satisfies the stochastic expansion,

$$\hat{\boldsymbol{\theta}} = \bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\phi}} + o_p(n^{-1/2}), \tag{2.12}$$

where $\boldsymbol{\phi} = \Pi(\boldsymbol{\psi}|[\mathbf{v}])$. Thus if $\Sigma = \mathrm{Var}(\boldsymbol{\psi}(Z)) - \mathrm{Var}(\boldsymbol{\phi}(Z))$ is nonsingular then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Longrightarrow \mathscr{N}(0, \Sigma).$$

We now allow the number of constraints to depend on n, $m = m_n$, and grow to infinity with increasing n. To stress the dependence, let us write

$$\mathbf{u}_n = (u_1, \dots, u_{m_n})^\top, \quad \hat{\mathbf{u}}_n = (\hat{u}_1, \dots, \hat{u}_{m_n})^\top,$$

and $\tilde{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}$ for the corresponding estimators of $\boldsymbol{\theta}$, that is,

$$\tilde{\boldsymbol{\theta}}_n = \frac{1}{n}\sum_{j=1}^{n} \frac{\boldsymbol{\psi}(Z_j)}{1 + \mathbf{u}_n(Z_j)^\top \tilde{\boldsymbol{\zeta}}_n} \quad \text{and} \quad \hat{\boldsymbol{\theta}}_n = \frac{1}{n}\sum_{j=1}^{n} \frac{\boldsymbol{\psi}(Z_j)}{1 + \hat{\mathbf{u}}_n(Z_j)^\top \hat{\boldsymbol{\zeta}}_n}, \tag{2.13}$$

where $\tilde{\boldsymbol{\zeta}}_n$ and $\hat{\boldsymbol{\zeta}}_n$ solves Eqt (2.2) with $\mathbf{u} = \mathbf{u}_n$ and $\mathbf{u} = \hat{\mathbf{u}}_n$, respectively. Denote by $[\mathbf{u}_\infty]$ the closed linear span of $\mathbf{u}_\infty = (u_1, u_2, \dots)$. Set

$$\mathbf{W}_n = \mathrm{Var}(\mathbf{u}_n(Z)), \quad \bar{\mathbf{W}}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{u}_n\mathbf{u}_n^\top(Z_j), \quad \hat{\mathbf{W}}_n = \frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^\top(Z_j).$$

Peng and Schick [26] introduced that a sequence $\mathbf{W}_n$ of $m_n \times m_n$ dispersion matrices is *regular* if

$$0 < \inf_n \inf_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbf{W}_n \mathbf{u} \le \sup_n \sup_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbf{W}_n \mathbf{u} < \infty.$$

Note that if $\mathbf{W} = \mathbf{W}_n$ is independent of n then the regularity of $\mathbf{W}$ simplifies to its nonsingularity. We have

Theorem 3. *Suppose that* $\mathbf{u}_n = (u_1, \dots, u_{m_n})^\top$ *satisfies (C) for each* $m = m_n$ *such that*

$$\max_{1\le j\le n} \|\mathbf{u}_n(Z_j)\| = o_p(m_n^{-3/2} n^{1/2}), \tag{2.14}$$

that the sequence of $m_n \times m_n$ *dispersion matrices* $\mathbf{W}_n$ *is regular and satisfies*

$$|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o = o_p(m_n^{-1}), \tag{2.15}$$

$$\frac{1}{n}\sum_{j=1}^n \big(\boldsymbol{\psi}(Z_j) \otimes \mathbf{u}_n(Z_j) - E\big(\boldsymbol{\psi}(Z_j) \otimes \mathbf{u}_n(Z_j)\big)\big) = o_p(m_n^{-1/2}). \tag{2.16}$$

Then $\tilde{\boldsymbol{\theta}}_n$ *satisfies, as* m_n *grows to infinity with* n*, the stochastic expansion,*

$$\tilde{\boldsymbol{\theta}}_n = \bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\varphi}}_0 + o_p(n^{-1/2}), \tag{2.17}$$

where $\boldsymbol{\varphi}_0 = \Pi(\boldsymbol{\psi}|[\mathbf{u}_\infty])$*. Thus if* $\Sigma_0 = \mathrm{Var}(\boldsymbol{\psi}(Z)) - \mathrm{Var}(\boldsymbol{\varphi}_0(Z))$ *is nonsingular,*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \Longrightarrow \mathscr{N}(0, \Sigma_0).$$

Theorem 4. *Suppose that* $\mathbf{u}_n = (u_1, \dots, u_{m_n})^\top$ *satisfies (C) for each* $m = m_n$*. Let* $\hat{\mathbf{u}}_n$ *be an estimator of* $\mathbf{u}_n$ *such that*

$$\max_{1\le j\le n} \|\hat{\mathbf{u}}_n(Z_j)\| = o_p(m_n^{-3/2} n^{1/2}), \tag{2.18}$$

$$|\hat{\mathbf{W}}_n - \mathbf{W}_n|_o = o_p(m_n^{-1}) \tag{2.19}$$

for which the $m_n \times m_n$ *dispersion matrices* $\mathbf{W}_n$ *is regular,*

$$\frac{1}{n}\sum_{j=1}^n \big(\boldsymbol{\psi}(Z_j) \otimes \hat{\mathbf{u}}_n(Z_j) - E\big(\boldsymbol{\psi}(Z_j) \otimes \hat{\mathbf{u}}_n(Z_j)\big)\big) = o_p(m_n^{-1/2}), \tag{2.20}$$

and that there exists some measurable function $\mathbf{v}_n$ *from* $\mathcal{Z}$ *into* $\mathcal{R}^{m_n}$ *such that (C) is met for every* $m = m_n$*, the dispersion matrix* $\mathbf{U}_n = \mathbf{W}_n^{-1/2} \int \mathbf{v}_n \mathbf{v}_n^\top \, dQ \mathbf{W}_n^{-\top/2}$ *satisfies* $|\mathbf{U}_n|_o = O(1)$*, and*

$$\frac{1}{n}\sum_{j=1}^n E\left(\|\hat{\mathbf{u}}_n(Z_j) - \mathbf{v}_n(Z_j)\|^2\right) = o(m_n^{-1}), \quad \textit{and} \tag{2.21}$$

$$\frac{1}{n}\sum_{j=1}^n \hat{\mathbf{u}}_n(Z_j) = \frac{1}{n}\sum_{j=1}^n \mathbf{v}_n(Z_j) + o_p(m_n^{-1/2} n^{-1/2}). \tag{2.22}$$

Then $\hat{\boldsymbol{\theta}}$ *satisfies, as* m_n *tends to infinity, the stochastic expansion,*

$$\hat{\boldsymbol{\theta}}_n = \bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\varphi}} + o_p(n^{-1/2}), \tag{2.23}$$

where $\boldsymbol{\varphi} = \Pi(\boldsymbol{\psi}|[\mathbf{v}_\infty])$. *Thus if* $\Sigma = \mathrm{Var}(\boldsymbol{\psi}(Z)) - \mathrm{Var}(\boldsymbol{\varphi}(Z))$ *is nonsingular,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \Longrightarrow \mathscr{N}(0, \Sigma).$$

2.2. Side information

In Section 3, we shall discuss side information in SEM. Wang and Peng (2023) constructed the EL-weighted estimators of linear functionals of a bivariate probability measure with *known marginal distributions* and with *equal but unknow marginal distributions*, and proved that the estimators are semiparametrically efficient for both cases. Such side information including *stochastic independene, distributional symmetry, etc.* is equivalent to infinitely many constraints. *Known marginal means or medians* are examples of side information expressed via finitely many constraints.

Another type of side information which is often available is as follows. Consider estimating the density $g(y)$ of a random variable Y based on a random sample $Y_1, \dots, Y_n$ on Y. One popular estimator is the kernel estimator,

$$\tilde{g}(y) = \frac{1}{n}\sum_{j=1}^{n} K_h(Y_i - y), \quad y \in \mathcal{R},$$

where $K(t)$ is a kernel and $K_h(t) = K(t/h)/h$ with $h > 0$ a bandwidth. Often there is available a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ on $\mathbf{X}$ which contain information about Y. To use the information, one can conduct a regression of Y on $\mathbf{X}$,

$$Y_i = h(\mathbf{X}_i; \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where h is a link function, $\boldsymbol{\beta}$ is a regression parameter vector, and ε_i's are random errors with zero mean. In this case, the estimated constraints are $\hat{u}_n(\mathbf{X}_i, Y_i) = Y_i - \hat{h}_n(\mathbf{X}_i; \hat{\boldsymbol{\beta}}_n)$, where $\hat{h}_n$ is an estimator of h (if it is uknown), and $\hat{\boldsymbol{\beta}}_n$ is an estimator of $\boldsymbol{\beta}$. The EL-weighted estimator of $g(y)$ is then given by

$$\hat{g}(y) = \frac{1}{n}\sum_{j=1}^{n} \frac{K_h(Y_j - y)}{1 + \hat{u}(\mathbf{X}_j, Y_j)\hat{\zeta}}, \quad y \in \mathcal{R},$$

where $\hat{\zeta}$ is similarly calculated. It can be shown that $\hat{g}(y)$ improves the efficiency of $\tilde{g}(y)$ for each y, but we shall not pursue this in this article.

More generally, one can construct the EL-weighted estimator for the conditional expected value of $\mathbf{T}$ given $Y = y$, i.e., $g(y) = E(\mathbf{T}|Y = y)$, when there is available a random sample $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ on $\mathbf{X}$ which contains information about $\mathbf{T}$. This estimator can be shown to improve the efficiency of the kernel estimator of $g(y)$ under suitable conditions.

3. That asyptotic properties of EL-weighted MDF estimators

In this section, we give two examples of side information and discuss the reduction in the asymptotic coviariance matrix of the improved MDF estimator $\hat{\boldsymbol{\vartheta}}_n$ given in (1.10) compared with the sample MDF estimator $\boldsymbol{\vartheta}_n$.

3.1. Examples and side information

We introduce the SEM and discuss side information.

Example 1. Consider the *combined model* of latent variable and measurement error,

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \Gamma\boldsymbol{\xi} + \boldsymbol{\zeta}, \quad \mathbf{Y} - \boldsymbol{\mu}_y = \Lambda_y\boldsymbol{\eta} + \boldsymbol{\epsilon}, \quad \mathbf{X} - \boldsymbol{\mu}_x = \Lambda_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3.1}$$

where $\mathbf{B}$, Γ, Λ_x, Λ_y, $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are compatible parameter matrices and vectors, $\mathbf{X}$ and $\mathbf{Y}$ are random vectors having finite fourth moments, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are latent endogenous and exogenous random vectors, respectively, and $\boldsymbol{\zeta}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are disturbances (random errors) that satisfy

$$\begin{aligned} &E(\boldsymbol{\zeta}) = 0,\ E(\boldsymbol{\epsilon}) = 0,\ E(\boldsymbol{\delta}) = 0, \quad \mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\eta}) = 0, \quad \mathrm{Cov}(\boldsymbol{\delta}, \boldsymbol{\xi}) = 0, \\ &\mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\zeta}) = 0, \quad \mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\zeta}) = 0, \quad \mathrm{Cov}(\boldsymbol{\delta}, \boldsymbol{\zeta}) = 0, \quad \mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = 0. \end{aligned} \tag{3.2}$$

Let $\Phi = E(\boldsymbol{\xi}\boldsymbol{\xi}^\top)$, $\Psi = E(\boldsymbol{\zeta}\boldsymbol{\zeta}^\top)$, $\Theta_\epsilon = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top)$ and $\Theta_\delta = E(\boldsymbol{\delta}\boldsymbol{\delta}^\top)$. The parameter vector then is $\boldsymbol{\vartheta} = \mathrm{vecs}(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \mathbf{B}, \Gamma, \Lambda_x, \Lambda_y, \Phi, \Psi, \Theta_\epsilon, \Theta_\delta)$, denoted by q the dimenion. Let $\Sigma_{yy}(\boldsymbol{\vartheta})$ be the structured variance-covariance of $\mathbf{Y}$, and let $\mathbf{A} = \mathbf{I}_d - \mathbf{B}$. Based on the relationships in (3.1) – (3.2), one derives

$$\Sigma_{yy}(\boldsymbol{\vartheta}) = \Lambda_y\mathbf{A}^{-1}(\Gamma\Phi\Gamma^\top + \Psi)\mathbf{A}^{-\top}\Lambda_y^\top + \Theta_\epsilon,$$

assuming that $\mathbf{A}$ is invertible. Similarly, one derives the structured covariance matrix $\Sigma_{yx}(\boldsymbol{\vartheta})$ of $\mathbf{Y}$ and $\mathbf{X}$ and the variance-covariance $\Sigma_{xx}(\boldsymbol{\vartheta})$ of $\mathbf{X}$,

$$\Sigma_{yx}(\boldsymbol{\vartheta}) = \Lambda_y\mathbf{A}^{-1}\Gamma\Phi\Lambda_x^\top = \Sigma_{xy}(\boldsymbol{\vartheta})^\top, \quad \Sigma_{xx}(\boldsymbol{\vartheta}) = \Lambda_x\Phi\Lambda_x^\top + \Theta_\delta.$$

The structured variance-covariance $\Sigma(\boldsymbol{\vartheta})$ of $\mathbf{Z} = (\mathbf{Y}^\top, \mathbf{X}^\top)^\top$ then is

$$\Sigma(\boldsymbol{\vartheta}) = \begin{pmatrix} \Sigma_{yy}(\boldsymbol{\vartheta}) & \Sigma_{yx}(\boldsymbol{\vartheta}) \\ \Sigma_{xy}(\boldsymbol{\vartheta}) & \Sigma_{xx}(\boldsymbol{\vartheta}) \end{pmatrix}.$$

These formulas can be found in literature, but we would mention that they are implied by the structural relationships in (3.1) – (3.2). While the unstructured sample variance-covariance matrix estimator $\mathbb{S}_n$ in (1.1) of the unstructured variance-covariance Σ of $\mathbf{Z}$ ignores the information contained in (3.2), the EL-weighted estimator $\hat{\mathbb{S}}_n$ of Σ in (1.9) utilizes the information, and results in an improved estimator $\hat{\boldsymbol{\vartheta}}$ determined by (1.10).

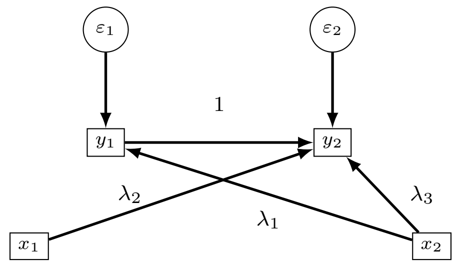

FIG 1. *The path diagram for SEM (3.4)*

Example 2. In the combined model in Example 1, consider $\Lambda_y = \mathbf{I}_d$, $\Lambda_x = \mathbf{I}_c$, $\mathrm{Var}(\boldsymbol{\delta}) = 0$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = 0$. This is a SEM, and (3.1) – (3.2) simplify to

$$\mathbf{Y} = \mathbf{BY} + \Gamma\mathbf{X} + \boldsymbol{\zeta}, \quad E(\boldsymbol{\zeta}) = 0, \quad \mathrm{Cov}(\mathbf{X}, \boldsymbol{\zeta}) = 0. \tag{3.3}$$

Identification is crucial for the consistency and asymptotic normality of the MDF estimators. Necessary and sufficient conditions can be found in the literature, e.g., Bollen (1989), Brito and Pearl (2002) and Drton, *et al.* (2011). In particular, the Null B Rule and the Recursive Rules are sufficient conditions for the identifiability of the parameters. The former states that if $\mathbf{B} = 0$ then the parameters can be identified, while the latter says that if $\mathbf{B}$ can be written as a lower triangular matrix with zero diagonal and the covariance matrix Ψ of the error $\boldsymbol{\zeta}$ is diagonal then the parameters are identifiable. An example for the latter case is the model given by

$$\mathbf{Y} = \mathbf{BY} + \Lambda\mathbf{X} + \boldsymbol{\epsilon}, \tag{3.4}$$

where $\mathbf{B}, \Lambda$ are 2×2 matrices, with $\mathbf{B}$ having all entries equal to 0 except for the $(2, 1)$ entry equal to β, and Λ having the (1, 1) entry equal to 0 and the (1, 2), (2, 1) and (2, 2) entries equal to $\lambda_i, i = 1, 2, 3$, respectively. The path diagram is shown in Fig. 2.

Side information. SEM make use of the information up to the second moments, whereas other information is completely ignored. For example, random errors are modeled as uncorrelated with covariates. It is common that the random error ϵ is modeled as independent of the random covariate $\mathbf{X}$. The information contained in the independence can be utilized by the vector constraint function,

$$\mathbf{g}(\mathbf{Z}) = \boldsymbol{\Phi}_m(F(\varepsilon)) \otimes \boldsymbol{\Phi}_m(G(\mathbf{X})), \tag{3.5}$$

where $\boldsymbol{\Phi}_m(t) = \sqrt{2}(\cos(\pi t), \dots, \cos(m\pi t))^\top$ is a vector of the first m terms of the trigonometric basis, and F and G are the respective distribution functions (DF) of the linear combination $\varepsilon = \mathbf{a}^\top\boldsymbol{\epsilon}$ of $\boldsymbol{\epsilon}$ and $\mathbf{X}$. Here $\mathbf{a}$ is a known constant vector and $\otimes$ denotes the Kronecker product. See Example 1 of Peng and Schick (2013, [29]) for more details. As F, G are unknown, we estimate them by the empirical DF (EDF) F_n, G_n. We replace $\boldsymbol{\epsilon}$ with $\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{B}}\mathbf{Y} - \hat{\Gamma}\mathbf{X}$, where $\hat{\mathbf{B}}$

and $\hat{\Gamma}$ are the MDF estimators of $\mathbf{B}$ and Γ. Substitution of them in (3.5) yields the estimated constraint function,

$$\hat{\mathbf{g}}(\mathbf{Z}) = \mathbf{\Phi}_m(F_n(\hat{\varepsilon})) \otimes \mathbf{\Phi}_m(G_n(\mathbf{X})), \tag{3.6}$$

This is a semiparametric model with (infinite dimensional) nuisance parameters F, G, and the plug-in estimators of F_n, G_n lead to the estimated constraints.

Another example of side information is that the marginal medians (or means) m_{01} and m_{02} of $\mathbf{X}$ are *known*. Such marginal information is often possible such as from the past data. In this case, the constraint function is

$$\mathbf{g}(\mathbf{Z}_j) = (\mathbf{1}[X_{1j} \le m_{01}] - 0.5,\ \mathbf{1}[X_{2j} \le m_{02}] - 0.5)^\top,\ j = 1, \ldots, n. \tag{3.7}$$

3.2. The asymptotic properties

We need some results from Shapiro (2007). Let $\boldsymbol{\vartheta}_0$ be the true value of parameter $\boldsymbol{\vartheta}$ and $\boldsymbol{\xi}_0 = \boldsymbol{\sigma}(\boldsymbol{\vartheta}_0)$. By the Taylor expansion it is not difficult to show that a discrepancy function F satisfies

$$2\mathbf{H}_0 := \frac{\partial^2 F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)}{\partial \mathbf{t} \partial \mathbf{t}^\top} = \frac{\partial^2 F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\top} = -\frac{\partial^2 F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)}{\partial \mathbf{t} \partial \boldsymbol{\xi}^\top}, \tag{3.8}$$

and $\mathbf{H}_0$ is positive definite, see also Shapiro (2007). In particular, for both F_{ML} and F_{GLS} (in the case of $W = \mathbb{S}_\ltimes$), one has

$$\mathbf{H}_0 = \Sigma_0^{-1} \otimes \Sigma_0^{-1}, \tag{3.9}$$

where $\Sigma_0 = \Sigma(\boldsymbol{\vartheta}_0)$. Formally, set $\Delta(\boldsymbol{\vartheta}) = \partial \boldsymbol{\sigma}(\boldsymbol{\vartheta}) / \partial \boldsymbol{\vartheta}^\top$ with $\Delta_0 = \Delta(\boldsymbol{\vartheta}_0)$ and

$$\begin{aligned} &\mathbf{w}(\mathbf{z}) = \operatorname{vecs}\big((\mathbf{z} - \boldsymbol{\mu}_0)(\mathbf{z} - \boldsymbol{\mu}_0)^\top - \Sigma_0\big), \quad \mathbf{z} \in \mathcal{R}^p, \\ &\mathbf{v}(\mathbf{z}) = \mathbf{g}(\mathbf{z}) + E(\dot{\mathbf{g}}(\mathbf{Z}))\Psi(\mathbf{z}), \quad \Psi(\mathbf{z}) = (\Delta_0^\top \mathbf{H}_0 \Delta_0)^{-1} \Delta_0^\top \mathbf{H}_0 \mathbf{w}(\mathbf{z}). \end{aligned} \tag{3.10}$$

Summarizing Shapiro's (2007) results, we have

Lemma 1. *Let $\mathbf{Z}, \mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be i.i.d. random vectors with finite and nonsingular covariance matrix* $\operatorname{Var}(\mathbf{w}(\mathbf{Z}))$. *Assume that $\boldsymbol{\vartheta}_0$ is an interior point of Θ which is compact and can be approximated at $\boldsymbol{\vartheta}_0$ by $\mathcal{R}^q$. Suppose that F is a discrepancy function. Suppose that $\boldsymbol{\sigma}(\boldsymbol{\vartheta})$ is twice continuously differentiable with gradient $\Delta(\boldsymbol{\vartheta})$ of full rank q in a neighborhood of $\boldsymbol{\vartheta}_0$. Suppose that the model is locally identifiable, i.e., $\boldsymbol{\sigma}(\boldsymbol{\vartheta}) = \boldsymbol{\sigma}(\boldsymbol{\vartheta}_0)$ implies $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ for $\boldsymbol{\vartheta}$ in a neighborhood of $\boldsymbol{\vartheta}_0$. Then*

$$\sqrt{n}(\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \Longrightarrow \mathcal{N}(0, \mathbf{V}_0), \tag{3.11}$$

where $\mathbf{V}_0 = (\Delta_0^\top \mathbf{H}_0 \Delta_0)^{-1} \Delta_0^\top \mathbf{H}_0 \operatorname{Var}(\mathbf{w}(\mathbf{Z})) \mathbf{H}_0 \Delta_0 (\Delta_0^\top \mathbf{H}_0 \Delta_0)^{-1}$.

Remark 2. Lemma 1 implies $\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = O_p(n^{-1/2})$. Consequently, each residual satisfies $\hat{\boldsymbol{\epsilon}}_i - \boldsymbol{\epsilon}_i = O_p(n^{-1/2})$ for $\mathbf{Z}_i$ of bounded second moment. We shall impose a stronger assumption of $E\|\hat{\boldsymbol{\epsilon}}_i - \boldsymbol{\epsilon}_i\|^2 = O(n^{-1})$ uniformly in i.

Proof of Lemma 1. We shall present the proof based on Theorem 5.5 of Shapiro (2007). To this end, we first verify the conditions of his Proposition 4.2 to show $\tilde{\boldsymbol{\vartheta}}$ is a consistent estimator of $\boldsymbol{\vartheta}_0$. Note that $\mathbf{s}_n = \text{vecs}(\mathbb{S}_n)$ is clearly a (strongly) consistent estimator of $\boldsymbol{\sigma}_0 = \boldsymbol{\sigma}(\boldsymbol{\vartheta}_0)$ since $\mathbb{S}_n$ is a (strongly) consistent estimator $\Sigma_0 = \Sigma(\boldsymbol{\vartheta}_0)$. The local identifiability of $\boldsymbol{\sigma}(\boldsymbol{\vartheta})$ at $\boldsymbol{\vartheta}_0$ implies the uniqueness of the optimal solution (i.e. $\boldsymbol{\vartheta}_0$), hence his (4.4) is proved since Θ is compact, see the last paragraph of his page 238. This establishes the consistency by his Proposition 4.2. It thus follows from his Theorem 5.5, (5.11) and (5.33) that $\tilde{\boldsymbol{\vartheta}}$ satisfies

$$\tilde{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}_0 + (\Delta_0^\top \mathbf{H}_0 \Delta_0)^{-1} \Delta_0^\top \mathbf{H}_0 (\mathbf{t}_n - \boldsymbol{\sigma}_0) + o_p(n^{-1/2}), \tag{3.12}$$

where $\mathbf{t}_n = \text{vecs}(\mathbf{T}_n)$ with $\mathbf{T}_n = n^{-1} \sum_{j=1}^n (\mathbf{Z}_j - \boldsymbol{\mu}_0)(\mathbf{Z}_j - \boldsymbol{\mu}_0)^\top$. Since $\mathbf{Z}$ has finite fourth moment, it follows from the central limit theorem,

$$\sqrt{n}(\mathbf{t}_n - \boldsymbol{\sigma}_0) \Longrightarrow \mathscr{N}(0, \text{Var}(\mathbf{w}(\mathbf{Z}))). \tag{3.13}$$

The preceding two displays yield the desired (3.24) and end the proof. □

Révész (1976) investigated the approximation of the empirical distribution function in two dimension. Unlike in the case of one dimension in which the Kolmogorov-Smirnov statistic is asymptotically distribution free, the test in two dimension is not asymptotically distribution free, as shown in his Theorem 3, which is quoted in Lemma 2 below. Let $\mathbf{Y} = (Y_1, Y_2)^\top$ be a random vector, and let $\mathbf{T}$ be a transformation of $\mathbf{Y}$ on $\mathcal{R}^2$ such that $\mathbf{TY}$ is uniformly distributed. Consider the transformation given by $\mathbf{T}(y_1, y_2) = (H(y_1), G(y_2|y_1))^\top$, where

$$H(y_1) = P(Y_1 \le y_1), \quad G(y_2|y_1) = P(Y_2 \le y_2 | Y_1 = y_1). \tag{3.14}$$

Lemma 2. *Let $\mathbf{Y}_1 = (Y_{11}, Y_{12})$, $\mathbf{Y}_2 = (Y_{21}, Y_{22})$, ... be a sequence of i.i.d. rv's having a common DF $F(\mathbf{y}) = F(y_1, y_2)$. Suppose that $F(y_1, y_2)$ is absolutely continuous and satisfies*

$$\Big|\frac{\partial G(y_2|H^{-1}(y_1))}{\partial y_1}\Big| \le L, \ \Big|\frac{\partial^2 G(y_2|H^{-1}(y_1))}{\partial y_1^2}\Big| \le L, \ \mathbf{y} = (y_1, y_2) \in \mathcal{R}^2, \tag{3.15}$$

for some constant $L > 0$. Then we can define a sequence $\{\bar{B}_n\}$ of Brownian Measures (B.M.) and a Kiefer Measure (K.M.) $\bar{K}$ such that

$$\begin{aligned} &\sup_{\mathbf{y}\in\mathcal{R}^2} |\beta_n(\mathbf{y}) - \bar{B}_n(TD_{\mathbf{y}})| = O(n^{-\frac{1}{19}}), \quad a.s. \\ &\sup_{\mathbf{y}\in\mathcal{R}^2} |n^{\frac{1}{2}}\beta_n(\mathbf{y}) - \bar{K}(TD_{\mathbf{y}}; n)| = O(n^{\frac{1}{2}\frac{2}{5}}), \quad a.s. \end{aligned} \tag{3.16}$$

where $\beta_n(\mathbf{y}) = n^{\frac{1}{2}}(F_n(\mathbf{y}) - F(\mathbf{y}))$ with $F_n(\mathbf{y})$ the EDF and $D_{\mathbf{y}} = [0, y_1] \times [0, y_2]$.

Remark 3. We shall assume $\sup_{\mathbf{y}} |\bar{B}_n(TD_{\mathbf{y}})| = O(1)$ a.s. for the DF F.

Here $\bar{B}$ and $\bar{K}$ are the stochastically equivalent versions of the "measures" B and K, and Révész (1976) remarked that "All the results here will be formulated and proved in the two-dimensional case only; it appears, however, that the generalization to higher dimensions is possible via the methods of this paper". To generalize the theorem to the d-dimensional case, we keep the definition of the Wiener Process $W(\mathbf{x}) = W(x_1, ..., x_d)$ to be a separable Gaussian process from Révész's paper, and define

$$\begin{aligned} B.M.: &\quad B(Q_z) = W(Q_z) - \lambda(Q_z)W(1, ..., 1) \\ K.M.: &\quad K(Q_z; y) = W(Q_z, y) - \lambda(Q_z)W(1, ..., 1, y) \end{aligned}$$

where $\lambda(\cdot)$ is a Lebesgue measure on $\mathscr{X}^d$. The d-dimensional transformation is defined by Rosenblatt (1952): Let $\mathbf{X} = (X_1, ..., X_d)$ be a random vector with DF $F(x_1, ..., x_d)$. Let $\mathbf{z} = (z_1, ..., z_d) = T\mathbf{x} = T(x_1, ..., x_d)$, where T is the transformation given by

$$\begin{aligned} z_1 &= P(X_1 \le x_1) = F_1(x_1), \\ z_2 &= P(X_2 \le x_2 | X_1 = x_1) = F_2(x_2|x_1), \\ &\vdots \\ z_d &= P(X_d \le x_d | X_{d-1} \le x_{d-1}, ..., X_1 = x_1) = F_d(x_d|x_{d-1}, ..., x_1). \end{aligned}$$

We generalize the conditions to the d-dimensional case (S).

(S1) $F(\mathbf{x})$ is absolutely continuous on $\mathbf{x} \in \mathcal{R}^d$.

(S2) For all $\mathbf{x} = (x_1, ..., x_d) \in \mathcal{R}^d$, there exists a constant $L > 0$,

$$\Big|\frac{\partial^2 F_2(x_2|F_1^{-1}(x_1))}{\partial x_1^2}\Big| \le L, \quad \Big|\frac{\partial F_2(x_2|F_1^{-1}(x_1))}{\partial x_1}\Big| \le L,$$

(S3)

$$\begin{aligned} &\Big|\frac{\partial^2 F_3(x_3|F_2^{-1}(x_2|F_1^{-1}(x_1)), F_1^{-1}(x_1))}{\partial x_i \partial x_j}\Big| \le L, \quad i, j = 1, 2, \\ &\Big|\frac{\partial F_3(x_3|F_2^{-1}(x_2|F_1^{-1}(x_1)), F_1^{-1}(x_1))}{\partial x_i}\Big| \le L, \quad i = 1, 2, \end{aligned}$$

$$\vdots$$

(Sd) For $d > 2$,

$$\begin{aligned} &\Big|\frac{\partial^2 F_d(x_d|F_{d-1}^{-1}(x_d|F_{d-2}^{-1}(x_{d-2}|...), ..., F_1^{-1}(x_1))}{\partial x_i \partial x_j}\Big| \le L, \quad i, j = 1, ..., d, \\ &\Big|\frac{\partial F_d(x_d|F_{d-1}^{-1}(x_d|F_{d-2}^{-1}(x_{d-2}|...), ..., F_1^{-1}(x_1))}{\partial x_i}\Big| \le L, \quad i = 1, ..., d. \end{aligned}$$

Theorem 5. *Let $\mathbf{X}_1 = (X_{11}, ..., X_{1d})$, $\mathbf{X}_2 = (X_{21}, ..., X_{2d})$, ... be a sequence of i.i.d. rv's having a common distribution function $F(\mathbf{x})$. Assume (S). Then we* *can define a sequence* $\{\bar{B}_n\}$ *of Brownian Measures (B.M.) and a Kiefer Measure (K.M.)* $\bar{K}$ *such that almost surely,*

$$\sup_{\mathbf{x}\in\mathcal{R}^d} |\beta_n(\mathbf{x}) - \bar{B}_n(TD_{\mathbf{x}})| = O(n^{-\frac{1}{19}}),$$
$$\sup_{\mathbf{x}\in\mathcal{R}^d} |n^{\frac{1}{2}}\beta_n(x) - \bar{K}(TD_{\mathbf{x}}; n)| = O(n^{\frac{1}{2}\frac{2}{5}}), \tag{3.17}$$

where $\beta_n(\mathbf{x}) = n^{\frac{1}{2}}(F_n(\mathbf{x}) - F(\mathbf{x}))$ *with* $F_n(\mathbf{x})$ *the EDF based on the sample* $\mathbf{X}_1$*, ...,* $\mathbf{X}_n$*, and* $D_{\mathbf{x}} = [0, x_1] \times ... \times [0, x_d]$.

We need a property of U-statistics. Let $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ be i.i.d. rv taking values in a measurable space $\mathcal{S}$. Let $\mathbf{h}$ be a measurable function from $\mathcal{S}^2$ to $\mathcal{R}^m$ which is symmetric, i.e., $\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{h}(\mathbf{y}, \mathbf{x}), \mathbf{x}, \mathbf{y} \in \mathcal{S}$. A multivariate U-statistic (of order 2) with kernel $\mathbf{h}$ is defined as

$$\mathbf{U}_n(\mathbf{h}) = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \mathbf{h}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j).$$

Assume that $\mathbf{h}$ is square-integrable. Let $\boldsymbol{\mu}(\mathbf{h}) = E(\mathbf{h}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2))$. Recall that a kernel $\mathbf{k}$ is *degenerate* if $E(\mathbf{k}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)|\boldsymbol{\xi}_2) = 0$ a.s. Let $\bar{\mathbf{h}}(\mathbf{x}) = E(\mathbf{h}(\mathbf{x}, \boldsymbol{\xi}_2))$, and

$$\mathbf{h}^*(\mathbf{x}, \mathbf{y}) = \mathbf{h}(\mathbf{x}, \mathbf{y}) - \bar{\mathbf{h}}(\mathbf{x}) - \bar{\mathbf{h}}(\mathbf{y}) + \boldsymbol{\mu}(\mathbf{h}).$$

Then $\mathbf{h}^*$ is a degenerate kernel. Let $\tilde{\mathbf{h}} = \mathbf{h} - \boldsymbol{\mu}(\mathbf{h})$. Then

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) = \boldsymbol{\mu}(\mathbf{h}) + \tilde{\mathbf{h}}(\mathbf{x}) + \tilde{\mathbf{h}}(\mathbf{y}) + \mathbf{h}^*(\mathbf{x}, \mathbf{y}).$$

One thus obtains the Hoeffding decomposition for a multivariate U-statistic,

$$\mathbf{U}_n(\mathbf{h}) = \boldsymbol{\mu}(\mathbf{h}) + \frac{2}{n}\sum_{j=1}^{n} \tilde{\mathbf{h}}(\boldsymbol{\xi}_j) + \mathbf{U}_n(\mathbf{h}^*) =: \boldsymbol{\mu}(\mathbf{h}) + \hat{\mathbf{U}}_n(\mathbf{h}) + \mathbf{U}_n(\mathbf{h}^*), \quad a.s. \tag{3.18}$$

Let $\mathbf{k}$ be a degenerate kernel with $E(\|\mathbf{k}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)\|^2) < \infty$. For $i < j, k < l$, one has $E(\mathbf{k}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)\mathbf{k}(\boldsymbol{\xi}_k, \boldsymbol{\xi}_l)^\top) = E(\mathbf{k}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^{\otimes 2})$ if $i = l, j = l$, and is equal to zero otherwise. Thus

$$E(\mathbf{U}_n(\mathbf{k})^{\otimes 2}) = \binom{n}{2}^{-1} E(\mathbf{k}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^{\otimes 2}). \tag{3.19}$$

It is easy to see $E(\mathbf{h}^*(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^{\otimes 2}) \preceq E(\mathbf{h}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^{\otimes 2})$. Thus we prove

Lemma 3. *Suppose that* $\mathbf{h}$ *is a kernel with* $E(\|\mathbf{h}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)\|^2) < \infty$*. Then*

$$\mathbf{U}_n(\mathbf{h}) - \boldsymbol{\mu}(\mathbf{h}) - \hat{\mathbf{U}}_n(\mathbf{h}) = O_p(n^{-1}\sqrt{E(\|\mathbf{h}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)\|^2)}).$$

We need a Lipschitz-type property.

(L) Let $\tilde{\boldsymbol\vartheta}^{(i)}$ be the estimator based on the observations with $\mathbf{Z}_i$ left out. Assume that there is a constant L_0 such that

$$\max_i \|\tilde{\boldsymbol\vartheta} - \tilde{\boldsymbol\vartheta}^{(i)}\| \le L_0/n. \tag{3.20}$$

Let $\tilde{\boldsymbol\vartheta}^{(ij)}$ be the estimator based on the observations with $\mathbf{Z}_i, \mathbf{Z}_j$ left out. Applying (L) repeatedly, one has for some constabt L_0',

$$\max_{ij} \|\tilde{\boldsymbol\vartheta} - \tilde{\boldsymbol\vartheta}^{(ij)}\| \le L_0'/n. \tag{3.21}$$

Let $\mathbf{v}_n(\mathbf{z}_1) = \boldsymbol\Phi_{m_n}(F(\varepsilon_1)) \otimes \boldsymbol\Phi_{m_n}(G(\mathbf{x}_1)) + 2(\mathbf{h}_{1,\mathbf{A}}(\mathbf{z}_1) + \mathbf{h}_{1,\mathbf{B}}(\mathbf{z}_1))$, where

$$\mathbf{h}_{1,\mathbf{A}}(\mathbf{z}_1) = E(\dot{\boldsymbol\Phi}_{m_n}(F(\varepsilon_2)) \otimes \boldsymbol\Phi_{m_n}(G(\mathbf{X}_2))(\mathbf{1}[\varepsilon_1 \le \varepsilon_2] - F(\varepsilon_2)|\mathbf{Z}_1 = \mathbf{z}_1),$$
$$\mathbf{h}_{1,\mathbf{B}}(\mathbf{z}_1) = E(\boldsymbol\Phi_{m_n}(F(\varepsilon_2)) \otimes \dot{\Phi}_{m_n}(G(\mathbf{X}_2))(\mathbf{1}[\mathbf{x}_1 \le \mathbf{X}_2] - G(\mathbf{X}_2))).$$

Theorem 6. *Suppose that the assumptions in Lemma 1 hold. Assume (L), (S) and the assumptions in Remark 2 and Remark 3. Suppse that ε has a bounded density. Suppose that $\mathbf{W}_{n2} = E(\boldsymbol\Phi_{m_n}(G(\mathbf{X}))^{\otimes 2})$ is regular and that $\int \mathbf{v}\mathbf{v}^\top\, dQ$ is nonsingular. If both m_n and n tend to infinity such that $m_n^{12}/n = o(1)$, then $\hat{\mathbf{s}}_n$ satisfies the stochastic expansion,*

$$\hat{\mathbf{s}}_n = \mathbf{s}_n - \mathbf{c}\,\mathrm{Var}(\mathbf{v}(\mathbf{Z}))^{-1}\bar{\mathbf{v}} + o_p(n^{-1/2}), \tag{3.22}$$

where $\mathbf{c} = E\big(\mathbf{w}(\mathbf{Z}) \otimes \mathbf{v}^\top(\mathbf{Z})\big)$. Thus, with $\mathbf{D} = \mathrm{Var}(\mathbf{w}(\mathbf{Z})) - \mathbf{c}\,\mathrm{Var}(\mathbf{v}(\mathbf{Z}))^{-1}\mathbf{c}^\top$,

$$\sqrt{n}(\hat{\mathbf{s}}_n - \boldsymbol\sigma(\boldsymbol\vartheta_0)) \Longrightarrow \mathscr{N}(0,\, \mathbf{D}), \tag{3.23}$$

As a consequence, $\hat{\boldsymbol\vartheta}$ given in (1.10) satisfies

$$\sqrt{n}(\hat{\boldsymbol\vartheta} - \boldsymbol\vartheta_0) \Longrightarrow \mathscr{N}(0, \mathbf{V}), \tag{3.24}$$

where $\mathbf{V} = (\Delta_0^\top \mathbf{H}_0 \Delta_0)^{-1} \Delta_0^\top \mathbf{H}_0 \mathbf{D} \mathbf{H}_0 \Delta_0 (\Delta_0^\top \mathbf{H}_0 \Delta_0)^{-1}$.

Proof of Theorem 6. We apply Theorem 4 with $\boldsymbol\psi(\mathbf{z}) = \mathrm{vecs}((\mathbf{z} - \boldsymbol\mu)^{\otimes 2})$. Write $m = m_n$. As $\hat{\mathbf{u}}_n(\mathbf{Z}_j) = \hat{\mathbf{g}}(\mathbf{Z}_j) = \boldsymbol\Phi_m(F_n(\hat\varepsilon)) \otimes \boldsymbol\Phi_m(G_n(\mathbf{X}))$ and $m^7/n = o(1)$,

$$\max_{1\le j\le n} \|\mathbf{u}_n(\mathbf{Z}_j)\| + \max_{1\le j\le n} \|\hat{\mathbf{u}}_n(\mathbf{Z}_j)\| \le 4m^2 = o(m^{-3/2} n^{1/2}).$$

This shows (2.18). Since $\mathbf{W}_n = E(\mathbf{u}_n(\mathbf{Z})\mathbf{u}_n(\mathbf{Z})^\top) = \mathbf{I}_m \otimes \mathbf{W}_{n2}$ and $\bar{\mathbf{W}}_n = \frac{1}{n}\sum_{j=1}^n \mathbf{u}_n(\mathbf{Z}_j)\mathbf{u}_n(\mathbf{Z}_j)^\top$, it follows that $\mathbf{W}_n$ is regular by the regularity of $\mathbf{W}_{n2}$, and that $|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o = O_p(m^2 n^{-1/2})$ as

$$\begin{aligned} E|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o^2 &\le E\|\bar{\mathbf{W}}_n - \mathbf{W}_n\|^2 = \mathrm{trace}(E(\bar{\mathbf{W}}_n - \mathbf{W}_n)^{\otimes 2}) \\ &\le n^{-1} E\|\mathbf{u}_n(\mathbf{Z})\|^4 \le m^4 n^{-1}. \end{aligned}$$

One verifies that there exists some constant $c_0 > 0$ such that for all t.

$$\|\boldsymbol\Phi_m(t)\| \le c_0 m^{1/2}, \quad \|\dot{\boldsymbol\Phi}_m(t)\| \le c_0 m^{3/2}, \quad \|\ddot{\boldsymbol\Phi}_m(t)\| \le c_0 m^{5/2}. \tag{3.25}$$

By the MVT, one thus has $|\hat{\mathbf{W}}_n - \bar{\mathbf{W}}|_o = O_p(m^5 n^{-1/2})$. Taken together one proves $|\hat{\mathbf{W}}_n - \mathbf{W}_n|_o = o_p(m^{-1})$ as $m^{12}/n = o(1)$, yielding (2.19). Moreover, it is not difficult to verify that $\mathbf{U}_n = \mathbf{W}_n^{-1/2} \int \mathbf{v}_n \mathbf{v}_n^\top \, dQ \mathbf{W}_n^{-\top/2} = O(1)$.

Write the left-hand-side average of (2.20) as $\mathbf{J}_n + \mathbf{K}_n - E(\mathbf{J}_n + \mathbf{K}_n)$, where

$$\mathbf{J}_n = \frac{1}{n} \sum_{j=1}^n \boldsymbol{\psi}(\mathbf{Z}_j) \otimes (\hat{\mathbf{u}}_n(\mathbf{Z}_j) - \mathbf{u}_n(\mathbf{Z}_j)),$$

$$\mathbf{K}_n = \frac{1}{n} \sum_{j=1}^n \boldsymbol{\psi}(\mathbf{Z}_j) \otimes (\mathbf{u}_n(\mathbf{Z}_j) - E(\mathbf{u}_n(\mathbf{Z}_j)).$$

Note first that

$$E(\|\mathbf{K}_n\|^2) \le n^{-1} E(|\boldsymbol{\psi}(\mathbf{Z})\|^2 \|\mathbf{u}_n(\mathbf{Z})\|^2) = O(m^4 n^{-1}). \tag{3.26}$$

We shall show next

$$E(\|\mathbf{J}_n\|^2) = O(m^4 n^{-1}). \tag{3.27}$$

Taken together we prove (2.20) as $m^5/n = o(1)$. To show (3.27), using the inequality $\|\mathbf{A} \otimes \mathbf{B}\| \le \|\mathbf{A}\| \, \|\mathbf{B}\|$ and by (3.25), we get

$$\begin{aligned} \|\hat{\mathbf{u}}_n(\mathbf{Z}_j) - \mathbf{u}_n(\mathbf{Z}_j)\| &\le \|\boldsymbol{\Phi}_m(F_n(\hat{\varepsilon}_j)) - \boldsymbol{\Phi}_m(F(\varepsilon_j))\| \cdot \|\boldsymbol{\Phi}_m(G_n(\mathbf{X}_j))\| \\ &\quad + \|\boldsymbol{\Phi}_m(F(\varepsilon_j)\| \cdot \|\boldsymbol{\Phi}_m(G_n(\mathbf{X}_j)) - \boldsymbol{\Phi}_m(G(\mathbf{X}_j))\| \\ &\le c_0 m^2 (|F_n(\hat{\varepsilon}_j) - F(\varepsilon_j)| + |G_n(\mathbf{X}_j) - G(\mathbf{X}_j)|). \end{aligned}$$

Let $D_n = \sup_t |F_n(t) - F(t)| = O_p(n^{-1/2})$ (Kolmogorov-Simirnov's statistic). As F has a bounded density (by c_f), we have

$$|F_n(\hat{\varepsilon}_j) - F(\varepsilon_j)| \le D_n + |F(\hat{\varepsilon}_j) - F(\varepsilon_j)| \le D_n + c_f |\hat{\varepsilon}_j - \varepsilon_j|, \tag{3.28}$$

By (3.33) below and Remark 2, we thus obtain

$$\frac{1}{n} \sum_{j=1}^n \|\hat{\mathbf{u}}(\mathbf{Z}_j) - \mathbf{u}(\mathbf{Z}_j)\|^2 = O(m^4/n). \tag{3.29}$$

Therefore (3.27) follows from

$$E(\|\mathbf{J}_n\|^2) \le E(\|\boldsymbol{\psi}(\mathbf{Z})\|^2) \frac{1}{n} \sum_{j=1}^n E(\|\hat{\mathbf{u}}(\mathbf{Z}_j) - \mathbf{u}(\mathbf{Z}_j)\|^2) = O(m^4/n). \tag{3.30}$$

We shall now show (2.21)–(2.22). Note

$$\begin{aligned} \boldsymbol{\Phi}_m(F_n(\hat{\varepsilon}_j)) &= \boldsymbol{\Phi}_m(F(\varepsilon_j)) + \dot{\boldsymbol{\Phi}}_m(F(\varepsilon_j))(F_n(\hat{\varepsilon}_j) - F(\varepsilon_j)) + \mathbf{R}_{1j} \\ \boldsymbol{\Phi}_m(G_n(\mathbf{X}_j)) &= \boldsymbol{\Phi}_m(G(\mathbf{X}_j)) + \dot{\boldsymbol{\Phi}}_m(G(\mathbf{X}_j))(G_n(\mathbf{X}_j) - G(\mathbf{X}_j)) + \mathbf{R}_{2j}, \end{aligned} \tag{3.31}$$

where, by (3.28) and the assumption in Remark 2, we have

$$\max_{1 \le j \le n} \|\mathbf{R}_{1j}\| = O_p(m^{5/2}/n). \tag{3.32}$$

By (3.16) and the assumption in Remark 3, we have

$$\max_{1\le j\le n} |G_n(\mathbf{X}_j) - G(\mathbf{X}_j)| = O_p(n^{-1/2}). \tag{3.33}$$

Similarly by Remark 3,

$$\max_{j} \|\mathbf{R}_{2j}\| = O_p(m^{5/2}/n). \tag{3.34}$$

By (3.31),

$$\frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{u}}_n(\mathbf{Z}_j) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{u}_n(\mathbf{Z}_j) + \mathbf{A} + \mathbf{B} + \mathbf{R}, \tag{3.35}$$

where

$$\mathbf{A} = \frac{1}{n}\sum_{j=1}^{n} \dot{\mathbf{\Phi}}_m(F(\varepsilon_j)) \otimes \mathbf{\Phi}_m(G(\mathbf{X}_j))(F_n(\hat{\varepsilon}_j) - F(\varepsilon_j)),$$

$$\mathbf{B} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{\Phi}_m(F(\varepsilon_j)) \otimes \dot{\Phi}_m(G(\mathbf{X}_j))(G_n(\mathbf{X}_j) - G(\mathbf{X}_j)),$$

$$\mathbf{R} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{R}_{1j} \otimes \mathbf{\Phi}_m(G_n(\mathbf{X}_j)) + \frac{1}{n}\sum_{j=1}^{n} \mathbf{\Phi}_m(F_n(\hat{\varepsilon}_j) \otimes \mathbf{R}_{2j}.$$

By (3.32) and (3.34) and the first equality in (3.25), $\|\mathbf{R}\| = O(m^3/n)$.

Let $\mathbf{b}(\mathbf{Z}_i, \mathbf{Z}_j) = \mathbf{\Phi}_m(F(\varepsilon_j)) \otimes \dot{\Phi}_m(G(\mathbf{X}_j))(\mathbf{1}[\mathbf{X}_i \le \mathbf{X}_j] - G(\mathbf{X}_j))$. It then follows $E(\kappa(\mathbf{Z}_i, \mathbf{Z}_j)) = 0$ for all i, j from the independence of ε and $\mathbf{X}$, and $\mathbf{B}$ is approximately a multivariate U-statistic, i.e., $\mathbf{B} = \mathbf{U}_n(\mathbf{h_B}) + O(m^2/n)$, where $\mathbf{h_B}(\mathbf{z}_1, \mathbf{z}_2) = \frac{1}{2}(\mathbf{b}(\mathbf{z}_i, \mathbf{z}_j) + \mathbf{b}(\mathbf{z}_j, \mathbf{z}_i))$. Let $\mathbf{h}_1(\mathbf{z}_1) = E(\mathbf{h}(\mathbf{z}_1, \mathbf{Z}_2))$. Then

$$\mathbf{h}_{1,\mathbf{B}}(\mathbf{z}_1) = E(\mathbf{\Phi}_m(F(\varepsilon_2)) \otimes \dot{\Phi}_m(G(\mathbf{X}_2))(\mathbf{1}[\mathbf{x}_1 \le \mathbf{X}_2] - G(\mathbf{X}_2))).$$

By Lemma 3,

$$\mathbf{B} = \frac{1}{n}\sum_{j=1}^{n} 2\mathbf{h}_{1,\mathbf{B}}(\mathbf{Z}_j) + O_p(m^2/n). \tag{3.36}$$

Write $F_n(\hat{\varepsilon}_j) - F(\varepsilon_j) = (F_n(\hat{\varepsilon}_j) - F_n(\varepsilon_j)) + (F_n(\varepsilon_j) - F(\varepsilon_j))$, giving $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$. Likewise, $\mathbf{h}_{1,\mathbf{A}}(\mathbf{z}_1) = E(\dot{\mathbf{\Phi}}_m(F(\varepsilon_2)) \otimes \mathbf{\Phi}_m(G(\mathbf{X}_2))(\mathbf{1}[\varepsilon_1 \le \varepsilon_2] - F(\varepsilon_2)|\mathbf{Z}_1 = \mathbf{z}_1)$,

$$\mathbf{A}_2 = \frac{1}{n}\sum_{j=1}^{n} 2\mathbf{h}_{1,\mathbf{A}}(\mathbf{Z}_j) + O_p(m^2/n). \tag{3.37}$$

Using (3.20), one calculates $E(\|\mathbf{A}_1\|^2) = O(m^4/n^2)$. Hence

$$\mathbf{A}_1 = \frac{1}{n}\sum_{j=1}^{n} \dot{\mathbf{\Phi}}_m(F(\varepsilon_j)) \otimes \mathbf{\Phi}_m(G(\mathbf{X}_j))(F_n(\hat{\varepsilon}_j) - F_n(\varepsilon_j)) = O_p(m^2/n). \tag{3.38}$$

This and (3.37) prove

$$\mathbf{A} = \frac{1}{n}\sum_{j=1}^{n} 2\mathbf{h}_{1,\mathbf{A}}(\mathbf{Z}_j) + O_p(m^2/n). \tag{3.39}$$

Taken together we show that (2.22) holds with $\mathbf{v} = \mathbf{u}_n + 2(\mathbf{h}_{1,\mathbf{A}} + \mathbf{h}_{1,\mathbf{B}})$ as $m_n^7/n = o(1)$. This also proves (2.21). □

4. Simulation results

We used the R package SEM to carry out the simulations based on the SEM (3.4) with $\beta = 1$. The details of the package can be found in Fox (2006)[**?**]. In LISREL notation and using Fig. 2, the SEM can be written as

$$y_1 = \lambda_1 x_2 + \epsilon_1, \quad y_2 = y_1 + \lambda_3 x_1 + \lambda_2 x_2 + \epsilon_2. \tag{4.1}$$

The parameters to be estimated are the coefficients λ_1, λ_2, λ_3, and the variances $v_1 = \mathrm{Var}(\epsilon_1), v_2 = \mathrm{Var}(\epsilon_2)$ of the measurement-errors.

For $n = 30, 50, 100$ and based on 50 repetitions, we calculated the averages and medians of the usual and the EL-weighted MDF estimators $\lambda_{n,k}$ and $\tilde{\lambda}_k$ of λ_k, the usual MDF and the EL-weighted estimators of variances $\tilde{v}_1, \tilde{v}_2$ v_1, v_2, and the ratios $r_1 = \bar{v}_2/\bar{v}_1$ and $r_2 = \mathrm{med}(\tilde{v}_2)/\mathrm{med}(\tilde{v}_1)$. The discrepancy function used is the ML (GLS) given in (1.3)((1.4))???. A value of ratio less than one indicated the variance reduction of the EL-weighted estimator over the usual estimator. The results are reported on Tables 1–2.

For Table 1, the side information is the *independence* of $\mathbf{X}$ and $\boldsymbol{\epsilon}$ utilized via the constraint functions given in (3.6) for $m = 1, 3, 5$, where ε was generated from the normal mixture $0.9 * N(0, \mathbf{I}_2) + 0.1 * N(0, 5\mathbf{I}_2)$, and $\mathbf{X}$ from the bivariate exponential $biexp(1, 3)$.

For Table 2, the side information is *known marginal medians* of $\mathbf{X}$, where $\mathbf{X}$ was generated from the bivariate exponential with scale parameters (γ_1, γ_2). One can see that the efficiency gain is substantial (around 40%). The ratios were stable with a slightly decreasing trend with increasing n, and the values of larger scale parameter had larger efficiency gains.

5. Proofs

In this section, we first give two useful general theorems. As applications, we prove the theorems presented in Section 2.

Let $\mathbf{x}_1, ..., \mathbf{x}_n$ be m-dimensional vectors. Set

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j, \quad x_* = \max_{1\le j\le n}\|\mathbf{x}_j\|, \quad \mathbb{S} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j\mathbf{x}_j^\top,$$

$$x^{(\nu)} = \sup_{\|\mathbf{u}\|=1}\left\|\frac{1}{n}\sum_{j=1}^{n}(\mathbf{u}^\top\mathbf{x}_j)^\nu\right\|, \quad \nu = 3, 4,$$

Table 1

Simulated efficiency gain of the EL-weighted estimators in the SEM (3.4) using the side information in (3.6) of independence of ε and $\mathbf{X}$ for a few values of n and number m of bases. $r_1(r_2)$ are the ratios of the averages (medians) of the variances of the EL-weighted estimators to the usual ones. $\varepsilon \sim 0.9 * N(0, I_2) + 0.1 * N(0, 5I_2), \mathbf{X} \sim \text{biexp}(1, 3)$.

$n = 30$											
m	λ	$\bar{b}_1$	$\bar{b}_2$	$m(b_1)$	$m(b_2)$	$\bar{v}_1$	$\bar{v}_2$	r_1	$m(v_1)$	$m(v_2)$	r_2
1	λ_1	0.2181	-0.2218	0.2876	-0.2437	2.4766	2.1962	0.8868	1.8529	1.3835	0.7467
	λ_2	-0.0783	-0.0544	-0.1018	-0.0586	2.1048	1.9226	0.9134	1.6773	1.3444	0.8015
	λ_3	0.0333	-0.1002	-0.0855	-0.1314	0.5620	0.5660	1.0071	0.3729	0.3550	0.9520
3	λ_1	-0.0893	-0.0688	0.0027	-0.0607	2.3394	1.7818	0.7616	1.5748	1.0300	0.6541
	λ_2	0.1005	-0.1172	-0.3671	-0.3255	2.3664	1.8948	0.8007	1.8519	1.3011	0.7026
	λ_3	0.1303	-0.0938	0.0153	-0.0941	0.6303	0.5588	0.8866	0.5221	0.4538	0.8692
5	λ_1	0.3742	-0.0233	0.2705	0.0381	1.7745	1.5109	0.8515	1.0320	0.8891	0.8615
	λ_2	-0.1992	-0.1774	-0.1107	-0.1846	2.2048	1.5480	0.7021	1.0474	0.7961	0.7601
	λ_3	-0.0555	-0.1333	-0.0797	-0.0429	0.4901	0.3838	0.7831	0.2615	0.2143	0.8195
$n = 50$											
m	λ	$\bar{b}_1$	$\bar{b}_2$	$m(b_1)$	$m(b_2)$	$\bar{v}_1$	$\bar{v}_2$	r_1	$m(v_1)$	$m(v_2)$	r_2
1	λ_1	0.0579	-0.0702	0.0963	-0.0546	1.4331	1.2840	0.8960	1.2614	0.9242	0.7327
	λ_2	-0.1007	-0.1425	-0.1802	-0.1571	1.3882	1.2730	0.9170	1.1588	1.0470	0.9035
	λ_3	-0.2388	-0.1862	-0.1546	-0.1966	0.3278	0.3130	0.9549	0.2709	0.2552	0.9420
3	λ_1	-0.0217	-0.0245	0.0697	-0.0398	1.2074	0.9621	0.7968	1.0255	0.7716	0.7524
	λ_2	0.1368	-0.1279	0.1578	-0.1023	1.3264	1.0642	0.8023	1.0684	0.9074	0.8493
	λ_3	0.0039	-0.0187	-0.0181	-0.0118	0.2586	0.2127	0.8225	0.2367	0.1823	0.7702
5	λ_1	0.1038	-0.0706	0.0412	-0.0716	1.1384	0.9720	0.8538	1.0185	0.8304	0.8153
	λ_2	0.2014	-0.2178	0.1338	-0.1635	1.3568	1.1220	0.8269	1.1851	0.9699	0.8184
	λ_3	-0.0049	-0.0213	0.0461	-0.0224	0.3525	0.3362	0.9538	0.2649	0.2500	0.9438
$n = 100$											
m	λ	$\bar{b}_1$	$\bar{b}_2$	$m(b_1)$	$m(b_2)$	$\bar{v}_1$	$\bar{v}_2$	r_1	$m(v_1)$	$m(v_2)$	r_2
1	λ_1	-0.0780	-0.0863	-0.0389	-0.0345	0.5978	0.5240	0.8765	0.5418	0.4167	0.7691
	λ_2	0.0550	-0.0943	0.0390	-0.0564	0.6421	0.5687	0.8857	0.6456	0.4629	0.7170
	λ_3	0.0102	-0.0581	0.0033	-0.0133	0.1508	0.1472	0.9761	0.1365	0.1281	0.9385
3	λ_1	0.1536	-0.0925	0.2299	-0.1106	0.5943	0.4759	0.8008	0.5237	0.4228	0.8073
	λ_2	-0.0769	-0.0637	-0.1581	-0.1517	0.6522	0.6210	0.9522	0.5397	0.5038	0.9335
	λ_3	0.0029	-0.0779	-0.0991	-0.0129	0.1621	0.1642	1.0130	0.1420	0.1410	0.9930
5	λ_1	0.0736	-0.0769	0.1407	-0.1056	0.5228	0.4403	0.8422	0.4625	0.3472	0.7507
	λ_2	-0.0601	-0.0712	-0.1616	-0.1752	0.5515	0.4704	0.8529	0.4872	0.3448	0.7077
	λ_3	-0.0090	-0.0242	-0.0339	-0.0770	0.1590	0.1465	0.9214	0.1299	0.1138	0.8761

TABLE 2

Same as Table 1 except for the side Information of known marginal medians of $\mathbf{X}$ *with* $\mathbf{X}$ *generated from the bivariate exponential with scale parameters* $(\gamma_1,\ \gamma_2)$.

$n = 30, m = 1$											
(γ_1, γ_2)	λ	$\bar{b}_1$	$\bar{b}_2$	$m(b_1)$	$m(b_2)$	$\bar{v}_1$	$\bar{v}_2$	r_1	$m(v_1)$	$m(v_2)$	r_2
$(2, 2)$	λ_1	-0.0720	-0.0942	-0.1217	-0.1600	0.1731	0.1531	0.8845	0.1478	0.1333	0.9019
	λ_2	0.0413	0.0298	-0.0104	0.0495	0.1934	0.1657	0.8568	0.1796	0.1425	0.7934
	λ_3	0.0453	0.0365	0.0885	0.0687	0.2067	0.1808	0.8747	0.1841	0.1640	0.8908
$(2, 3)$	λ_1	0.0424	0.0087	-0.0198	-0.0177	0.3424	0.2126	0.6209	0.3016	0.1918	0.6359
	λ_2	-0.0937	-0.0527	-0.1107	-0.0952	0.4198	0.2852	0.6794	0.3813	0.2478	0.6499
	λ_3	-0.0600	-0.1512	-0.1224	-0.1159	0.2004	0.1819	0.9077	0.1626	0.1518	0.9336
$(2, 4)$	λ_1	0.0209	0.0535	-0.0327	-0.1286	0.7113	0.2411	0.3390	0.6064	0.2367	0.3903
	λ_2	0.0860	0.1354	0.1475	0.1280	0.6681	0.3136	0.4694	0.5949	0.3068	0.5157
	λ_3	0.0825	0.0080	0.0064	-0.1096	0.1741	0.1917	1.1011	0.1312	0.1411	1.0755
$n = 50$											
(γ_1, γ_2)	λ	$\bar{b}_1$	$\bar{b}_2$	$m(b_1)$	$m(b_2)$	$\bar{v}_1$	$\bar{v}_2$	r_1	$m(v_1)$	$m(v_2)$	r_2
$(2, 2)$	λ_1	0.0196	0.0318	0.0746	0.0918	0.0874	0.0721	0.8249	0.0797	0.0678	0.8507
	λ_2	-0.0547	-0.0467	-0.0646	-0.0760	0.1021	0.0876	0.8580	0.0953	0.0830	0.8709
	λ_3	-0.0847	-0.0766	-0.0894	-0.1072	0.0997	0.0876	0.8786	0.0877	0.0811	0.9247
$(2, 3)$	λ_1	0.0059	0.0069	0.0058	0.0237	0.2121	0.1190	0.5611	0.1757	0.1072	0.6101
	λ_2	-0.0544	-0.0326	-0.0447	-0.0061	0.2422	0.1489	0.6148	0.2288	0.1299	0.5677
	λ_3	0.0422	0.0342	0.0937	0.0515	0.1068	0.1026	0.9607	0.0954	0.0827	0.8669
$(2, 4)$	λ_1	0.1288	0.1121	-0.0007	0.0403	0.3427	0.1501	0.4380	0.3068	0.1376	0.4485
	λ_2	0.1900	0.1183	0.0930	0.1413	0.3781	0.1822	0.4819	0.3453	0.1700	0.4923
	λ_3	-0.0182	0.0027	-0.0527	0.0671	0.0978	0.0994	1.0164	0.0895	0.0867	0.9687
$n = 100$											
(γ_1, γ_2)	λ	$\bar{b}_1$	$\bar{b}_2$	$m(b_1)$	$m(b_2)$	$\bar{v}_1$	$\bar{v}_2$	r_1	$m(v_1)$	$m(v_2)$	r_2
$(2, 2)$	λ_1	0.0353	0.0346	0.0460	0.0478	0.0369	0.0326	0.8837	0.0360	0.0326	0.9050
	λ_2	-0.0511	-0.0585	-0.0298	-0.0316	0.0428	0.0378	0.8836	0.0423	0.0358	0.8456
	λ_3	-0.0482	-0.0453	-0.0882	-0.0691	0.0431	0.0390	0.9045	0.0407	0.0361	0.8877
$(2, 3)$	λ_1	-0.0809	-0.0784	-0.1066	-0.0858	0.0985	0.0677	0.6878	0.0941	0.0642	0.6819
	λ_2	0.0191	0.0262	0.0175	0.0007	0.1037	0.0608	0.5865	0.1009	0.0673	0.6673
	λ_3	-0.0104	-0.0128	-0.0179	-0.0125	0.0444	0.0436	0.9820	0.0395	0.0391	0.9899
$(2, 4)$	λ_1	-0.0641	-0.0655	-0.0194	-0.0238	0.1732	0.1011	0.5838	0.1688	0.0973	0.5763
	λ_2	0.0173	0.0186	-0.0162	0.0175	0.1747	0.1025	0.5868	0.1716	0.0848	0.4942
	λ_3	-0.0484	-0.0464	-0.0168	-0.0152	0.0413	0.0409	0.9903	0.0385	0.0371	0.9636

and let λ and Λ denote the smallest and largest eigen value of the matrix $\mathbb{S}$,

$$\lambda = \inf_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbb{S}\mathbf{u} \quad \text{and} \quad \Lambda = \sup_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbb{S}\mathbf{u}.$$

With these we associate the empirical likelihood

$$\mathscr{R} = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \mathbf{x}_j = 0 \Big\}.$$

Peng and Schick [26] carefully examined the above maximization as a numeric problem and detailed some very useful properties. We quote their Lemma 5.2 below for our application.

Lemma 4. *The inequality $\lambda > 5\|\bar{\mathbf{x}}\|x_*$ implies that there is a unique $\boldsymbol{\zeta}$ in $\mathcal{R}^m$ satisfying the below (5.1) to (5.7).*

$$1 + \boldsymbol{\zeta}^\top \mathbf{x}_j > 0, \quad j = 1, \dots, n, \tag{5.1}$$

$$\sum_{j=1}^n \frac{\mathbf{x}_j}{1 + \boldsymbol{\zeta}^\top \mathbf{x}_j} = 0, \tag{5.2}$$

$$\|\boldsymbol{\zeta}\| \le \frac{\|\bar{\mathbf{x}}\|}{\lambda - \|\bar{\mathbf{x}}\|x_*}, \tag{5.3}$$

$$\|\boldsymbol{\zeta}\|x_* \le \frac{\|\bar{\mathbf{x}}\|x_*}{\lambda - \|\bar{\mathbf{x}}\|x_*} < \frac{1}{4}, \tag{5.4}$$

$$\frac{1}{n}\sum_{j=1}^n (\boldsymbol{\zeta}^\top \mathbf{x}_j)^2 = \boldsymbol{\zeta}^\top \mathbb{S}\boldsymbol{\zeta} \le \Lambda\|\boldsymbol{\zeta}\|^2 \le \frac{\Lambda\|\bar{\mathbf{x}}\|^2}{(\lambda - \|\bar{\mathbf{x}}\|x_*)^2}, \tag{5.5}$$

$$\Big\|\frac{1}{n}\sum_{j=1}^n \frac{\mathbf{r}_i}{1 + \boldsymbol{\zeta}^\top \mathbf{x}_j} - \mathbf{r}_j + \mathbf{r}_j \mathbf{x}_j^\top \boldsymbol{\zeta}\Big\| \le \Big\|\frac{1}{n}\sum_{j=1}^n \mathbf{r}_j (\boldsymbol{\zeta}^\top \mathbf{x}_j)^2\Big\| + \frac{4}{3}\frac{1}{n}\sum_{j=1}^n \|\mathbf{r}_j\|\ |\boldsymbol{\zeta}^\top \mathbf{x}_j|^3, \tag{5.6}$$

for vectors $\mathbf{r}_1, \dots, \mathbf{r}_n$ of the same dimension, and

$$\|\boldsymbol{\zeta} - \mathbb{S}^{-1}\bar{\mathbf{x}}\|^2 \le 2\Big(\frac{1}{\lambda} + \frac{\Lambda}{9\lambda^2}\Big)\|\zeta\|^4 x^{(4)}. \tag{5.7}$$

Now use the fact that $\|\mathbf{x}\| = \sup_{\|\mathbf{v}\|=1} \mathbf{v}^\top \mathbf{x}$, the Cauchy-Schwartz inequality, (5.3), (5.4) and (5.5) to bound the square of the first term of the right-hand side of (5.6) by

$$\frac{1}{n}\sum_{j=1}^n (\boldsymbol{\zeta}^\top \mathbf{x}_j)^4 \sup_{\|\mathbf{v}\|=1} \mathbf{v}^\top \Big(\frac{1}{n}\sum_{j=1}^n \mathbf{r}_j \mathbf{r}_j^\top\Big)\mathbf{v} \le \|\boldsymbol{\zeta}\|^4 x^{(4)} \Big|\frac{1}{n}\sum_{j=1}^n \mathbf{r}_j \mathbf{r}_j^\top\Big|_o$$

and the square of the second term by

$$\frac{16}{9} x_*^2 \|\boldsymbol{\zeta}\|^2 \frac{1}{n}\sum_{j=1}^n \|\mathbf{r}_j\|^2 \frac{1}{n}\sum_{j=1}^n (\boldsymbol{\zeta}^\top \mathbf{x}_j)^4 \le \frac{16d}{9}(x_*\|\boldsymbol{\zeta}\|)^2 \|\boldsymbol{\zeta}\|^4 \mathbf{x}^{(4)} \Big|\frac{1}{n}\sum_{j=1}^n \mathbf{r}_j \mathbf{r}_j^\top\Big|_o,$$

where d is the dimension of r_j. Combining the above we obtain

$$\Big\|\frac{1}{n}\sum_{j=1}^{n}\frac{\mathbf{r}_i}{1+\boldsymbol{\zeta}^\top\mathbf{x}_j}-\mathbf{r}_j+\mathbf{r}_j\mathbf{x}_j^\top\boldsymbol{\zeta}\Big\|^2\le\|\boldsymbol{\zeta}\|^4x^{(4)}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{r}_j\mathbf{r}_j^\top\Big|_o\big[1+\frac{16d}{9}(x_*\|\boldsymbol{\zeta}\|)^2\big]. \quad (5.8)$$

We now apply the above results to random vectors. Let $\mathbf{T}_{n1},\dots,\mathbf{T}_{nn}$ be m_n-dimensional random vectors. With these random vectors we associate the empirical likelihood

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

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

$$T_n^*=\max_{1\le j\le n}\|\mathbf{T}_{nj}\|,\quad \bar{\mathbf{T}}_n=n^{-1}\sum_{j=1}^{n}\mathbf{T}_{nj},\quad T_n^{(\nu)}=\sup_{\|\mathbf{u}\|=1}\Big\|\frac{1}{n}\sum_{j=1}^{n}(\mathbf{u}^\top\mathbf{T}_{nj})^\nu\Big\|,$$

and the matrix

$$\mathbb{S}_n=\frac{1}{n}\sum_{j=1}^{n}\mathbf{T}_{nj}\mathbf{T}_{nj}^\top,$$

and let λ_n and Λ_n denote the smallest and largest eigen values of $\mathbb{S}_n$,

$$\lambda_n=\inf_{\|\mathbf{u}\|=1}\mathbf{u}^\top\mathbb{S}_n\mathbf{u}\quad\text{and}\quad\Lambda_n=\sup_{\|\mathbf{u}\|=1}\mathbf{u}^\top\mathbb{S}_n\mathbf{u}.$$

We impose the following conditions on $\mathbf{T}_{nj}$.

(A1) $T_n^*=o_p(m_n^{-3/2}n^{1/2})$.
(A2) $\|\bar{\mathbf{T}}_n\|=O_p(m_n^{1/2}n^{-1/2})$.
(A3) There is a sequence of regular $m_n\times m_n$ dispersion matrices $\mathbf{W}_n$ such that

$$|\mathbb{S}_n-\mathbf{W}_n|_o=o_p(m_n^{-1}).$$

We impose the following conditions on $\boldsymbol{\psi}$ and $\mathbf{T}_{nj}$.

(B1) $n^{-1}\sum_{j=1}^{n}\big(\boldsymbol{\psi}(Z_j)\otimes\mathbf{T}_{nj}^\top-E\big(\boldsymbol{\psi}(Z_j)\otimes\mathbf{T}_{nj}^\top\big)\big)=o_p(m_n^{-1/2})$.
(B2) There exists some measurable function $\boldsymbol{\chi}$ from $\mathcal{Z}$ into $\mathcal{R}^d$ such that $\int\boldsymbol{\chi}\,dQ=0$, $\int\|\boldsymbol{\chi}\|^2\,dQ<\infty$ and

$$\frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{A}_n\mathbf{W}_n^{-1}\mathbf{T}_{ni}-\boldsymbol{\chi}(Z_i)\big)=o_p(n^{-1/2}),$$

where $\mathbf{A}_n=:n^{-1}\sum_{j=1}^{n}E\big(\boldsymbol{\psi}(Z_j)\otimes\mathbf{T}_{nj}^\top\big)$.

Let us first consider the case that m_n tends to infinity with the sample size. We have the following result.

Theorem 7. *Suppose (A1)-(A3) and (B1)-(B2) hold. Then there exists unique* $\boldsymbol{\zeta}_n$ *which satisfies*

$$1+\boldsymbol{\zeta}_n^\top \mathbf{T}_{nj} > 0, \quad \frac{1}{n}\sum_{j=1}^n \frac{\mathbf{T}_{nj}}{1+\boldsymbol{\zeta}_n^\top \mathbf{T}_{nj}} = 0, \tag{5.9}$$

such that as m_n *tends to infinity,*

$$\boldsymbol{\theta}_n =: \frac{1}{n}\sum_{j=1}^n \frac{\boldsymbol{\psi}(Z_j)}{1+\boldsymbol{\zeta}_n^\top \mathbf{T}_{nj}} = \bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\chi}} + o_p(n^{-1/2}), \tag{5.10}$$

where $\bar{\boldsymbol{\chi}} = n^{-1}\sum_{j=1}^n \boldsymbol{\chi}(Z_j)$ *with* $\boldsymbol{\chi}$ *given in (B2).*

Proof. It follows from (A1) and (A2) that $T_n^*\|\bar{\mathbf{T}}_n\| = o_p(1)$, and from (A3) that there are positive numbers $a < b$ such that $P(a \le \lambda_n \le \Lambda_n \le b) \to 1$. Thus all three conditions imply that the probability of the event $\{\lambda_n > 5T_n^*\|\bar{\mathbf{T}}_n\|\}$ tends to one. Consequently, by Lemma 4, there exists an m_n-dimensional random vector $\boldsymbol{\zeta}$ which is uniquely determined on this event by the properties (5.1)–(5.8) including (5.9). To prove (5.10), we apply (5.8) with $\mathbf{r}_j = \boldsymbol{\psi}(Z_j)$. Note first that

$$T_n^{(4)} \le \Lambda_n (T_n^*)^2. \tag{5.11}$$

This, (5.3) and (A1)-(A2) imply that the right side of (5.8) is bounded by

$$\Big(1+\frac{d}{9}\Big)\frac{\|\bar{\mathbf{T}}_n\|^4}{(\lambda_n - \|\bar{\mathbf{T}}_n\|T_n^*)^4}\Lambda_n (T_n^*)^2 \Big|\frac{1}{n}\sum_{j=1}^n \boldsymbol{\psi}(Z_j)\boldsymbol{\psi}(Z_j)^\top\Big|_o = o_p(m_n^{-1}n^{-1}),$$

where the equality holds since the spectral norm of the average is bounded due to the square-integrability of $\boldsymbol{\psi}$. Thus from (5.8) it follows

$$\boldsymbol{\theta}_n = \frac{1}{n}\sum_{j=1}^n \boldsymbol{\psi}(Z_j) - \frac{1}{n}\sum_{j=1}^n \boldsymbol{\psi}(Z_j)\otimes \mathbf{T}_{nj}^\top \boldsymbol{\zeta}_n + o_p(n^{-1/2}). \tag{5.12}$$

In view of (B2) the desired (5.10) now follows from (5.13)-(5.15) below.

$$\frac{1}{n}\sum_{j=1}^n \big(\boldsymbol{\psi}(Z_j)\otimes \mathbf{T}_{nj}^\top - E\big(\boldsymbol{\psi}(Z_j)\otimes \mathbf{T}_{nj}^\top\big)\big)\,\boldsymbol{\zeta}_n = o_p(n^{-1/2}), \tag{5.13}$$

$$\frac{1}{n}\sum_{j=1}^n E\big(\boldsymbol{\psi}(Z_j)\otimes \mathbf{T}_{nj}^\top\big)\big(\boldsymbol{\zeta}_n - \mathbb{S}_n^{-1}\bar{\mathbf{T}}_n\big) = o_p(n^{-1/2}), \tag{5.14}$$

$$\frac{1}{n}\sum_{j=1}^n E\big(\boldsymbol{\psi}(Z_j)\otimes \mathbf{T}_{nj}^\top\big)\big(\mathbb{S}_n^{-1} - \mathbf{W}_n^{-1}\big)\bar{\mathbf{T}}_n = o_p(n^{-1/2}). \tag{5.15}$$

Note first that (B1), (A2) and (5.3) imply (5.13). Next we show

$$\mathbf{A}_n = \frac{1}{n}\sum_{j=1}^n E\big(\boldsymbol{\psi}(Z_j)\otimes \mathbf{T}_{nj}^\top\big) = O(m_n^{1/2}).$$

Indeed, by Cauchy inequality,

$$\|\mathbf{A}_n\|^2 \le \frac{1}{n}\sum_{j=1}^{n}\|E\big(\boldsymbol{\psi}(Z_j)\otimes\mathbf{T}_{nj}^\top\big)\|^2 \le E\big(\|\boldsymbol{\psi}(Z_1)\|^2\big)\frac{1}{n}\sum_{j=1}^{n}E(\|\mathbf{T}_{nj}\|^2)$$
$$= E\big(\|\boldsymbol{\psi}(Z_1)\|^2\big)\mathrm{trace}\,(E(\mathbb{S}_n))\,.$$

But by (A3), the above trace is bounded by

$$\|\mathrm{trace}\,(E(\mathbb{S}_n-\mathbf{W}_n))\,\| + \mathrm{trace}\,(E(\mathbf{W}_n)) \le m_n E\,(|\mathbb{S}_n-\mathbf{W}_n|_o) + \Lambda_n m_n,$$

Thus $\|\mathbf{A}_n\|^2 = O(m_n)$. This, the regularity of $\mathbf{W}_n$ in (A3), (5.7), (5.11) and (A1) imply that the square of the right side of (5.14) is bounded by

$$O(m_n)O_p(\|\boldsymbol{\zeta}_n\|^4 T_n^{(4)}) = O(m_n)O_p(\|\bar{\mathbf{T}}_n\|^4 (T_n^*)^2)$$
$$= O(m_n)o_p(m_n^2 n^{-2} m_n^{-3} n) = o_p(m_n^{-1}n^{-1}),$$

hence (5.14) is proved. Again the rate of A_n and (A2)-(A3) imply (5.15). This completes the proof. □

Examining the proof of Theorem 7 one can see the following holds.

Theorem 8. *Suppose (A1)-(A3) and (B1)-(B2) are met for fixed $m_n = m$. Then the results in Theorem 7 hold as n tends to infinity.*

Proof of Theorem 1. We verify the conditions of Theorem 8 with $\mathbf{T}_{nj} = \mathbf{u}(Z_j)$. Since $\mathbf{u}$ is square-integrable, conditions (A1) – (A3) are satisfied with $\mathbf{W}_n = \mathbf{W} = E(\mathbf{u}\mathbf{u}^\top(Z))$. The square-integrability of $\boldsymbol{\psi}$ implies that (B1) – (B2) are met with $\mathbf{A}_n = \mathbf{A} = E(\boldsymbol{\psi}(Z)\otimes\mathbf{u}(Z)^\top)$ and $\boldsymbol{\chi} = \mathbf{A}\mathbf{W}^{-1}\mathbf{u}$, by the weak law of large numbers. We now apply the result of Theorem 8 to complete the proof. □

Proof of Theorem 2. We shall apply Theorem 8 to prove the result with $\mathbf{T}_{nj} = \hat{\mathbf{u}}(Z_j)$. Clearly conditions (A1), (A3) and (B1) follows from (2.7) – (2.9) respectively, whereas (A2) is implied by (2.11) in view of the fact that the right-hand-side average of (2.11) is $O_p(n^{-1/2})$. By Cauchy inequality,

$$\Big\|\frac{1}{n}\sum_{j=1}^{n}E\big(\boldsymbol{\psi}(Z_j)\otimes(\hat{\mathbf{u}}(Z_j)-\mathbf{v}(Z_j))\big)\Big\|^2 \le E(\|\boldsymbol{\psi}(Z_1)\|^2)\frac{1}{n}\sum_{j=1}^{n}E\big(\|\hat{\mathbf{u}}(Z_j)-\mathbf{v}(Z_j)\|^2\big),$$

which is $o(1)$ by (2.10). Hence the $\mathbf{A}_n$ in (B2) is given by

$$\mathbf{A}_n = \frac{1}{n}\sum_{j=1}^{n}E\big(\boldsymbol{\psi}(Z_j)\otimes\hat{\mathbf{u}}(Z_j)\big) = E(\boldsymbol{\psi}(Z_1)\mathbf{v}(Z_1)) + o(1).$$

Thus by (2.11), (B2) holds with $\boldsymbol{\chi} = E(\boldsymbol{\psi}(Z_1)\mathbf{v}(Z_1))\mathbf{W}^{-1}\mathbf{v}$. We now apply Theorem 8 to complete the proof. □

Proof of Theorem 3. We apply Theorem 7 with $\mathbf{T}_{nj} = \mathbf{u}_n(Z_j)$ to prove the result, i.e. verify its conditions (A1)-(A3) and (B1)-(B2). Obviously (2.14), (2.15) and (2.16) correspond to (A1), (A3) and (B1) respectively. It follows from the regularity of $\mathbf{W}_n$ that trace$(\mathbf{W}_n) \le Bm_n$ for some constant B. Thus from $nE[\|\bar{\mathbf{T}}_n\|^2] = \text{trace}(\mathbf{W}_n) = O(m_n)$ it yields (A2). We are now left to prove (B2). Noticing $\mathbf{A}_n = E(\boldsymbol{\psi}(Z) \otimes \mathbf{u}_n^\top(Z))$ and $\mathbf{W}_n = E(\mathbf{u}_n \mathbf{u}_n(Z)^\top)$, and $\mathbf{A}_n \mathbf{W}_n^{-1} \mathbf{u}_n$ is the projection of $\boldsymbol{\psi}(Z)$ onto the closed linear span $[\mathbf{u}_n]$, so that $\mathbf{A}_n \mathbf{W}_n^{-1} \mathbf{u}_n(Z)$ is the conditional expectation of $\boldsymbol{\psi}(Z)$ given $\mathbf{u}_n(Z)$, i.e.,

$$\mathbf{A}_n \mathbf{W}_n^{-1} \mathbf{u}_n(Z) = E(\boldsymbol{\psi}(Z)|\mathbf{u}_n(Z)).$$

Since $E(\boldsymbol{\psi}(Z)|\mathbf{u}_n(Z)), n \ge 1$ forms a martingale with respect to the sigma algebras, $\sigma(\mathbf{u}_n(Z)), n \ge 1$, generated by $\mathbf{u}_n(Z)$, it follows from Lévy's martingale convergence theorem (see e.g. page 510, Shiryaev [34]) that

$$E(\boldsymbol{\psi}(Z)|\sigma(\mathbf{u}_n(Z))) \to E(\boldsymbol{\psi}(Z)|\sigma(\mathbf{u}_\infty(Z))), \quad a.s. \quad n \to \infty.$$

By the property of conditional expectation (see e.g. Proposition 1, page 430, Bickel, *et al.* [1]), the last conditional expectation is the projection of $\boldsymbol{\psi}(Z)$ onto the closed linear span $[\mathbf{u}_\infty(Z)]$, i.e., $E(\boldsymbol{\psi}(Z)|\sigma(\mathbf{u}_\infty(Z))) = \Pi(\boldsymbol{\psi}(Z)|[\mathbf{u}_\infty(Z)])$, hence

$$\boldsymbol{\varphi}(Z) = \Pi(\boldsymbol{\psi}(Z)|[\mathbf{u}_\infty(Z)]) = E(\boldsymbol{\psi}(Z)|\mathbf{u}_\infty(Z)).$$

Thus that (B2) is satisfied with $\boldsymbol{\chi} = \boldsymbol{\varphi}$ follows from

$$nE\Big(\|\frac{1}{n}\sum_{i=1}^n \big(\mathbf{A}_n \mathbf{W}_n^{-1} \mathbf{u}_n(Z_i) - \boldsymbol{\varphi}(Z_i)\big)\|^2\Big) = E\Big(\|\mathbf{A}_n \mathbf{W}_n^{-1} \mathbf{u}_n(Z) - \boldsymbol{\chi}(Z)\|^2\Big),$$

which converges to zero as n tends to infinity by the property of the convergence of Fourier series. This completes the proof. □

Proof of Theorem 4. We prove the result by verifying conditions (A1)-(A3) and (B1)-(B2) of Theorem 7 with $\mathbf{T}_{nj} = \hat{\mathbf{u}}_n(Z_j)$. Clearly $(A1)$, $(A3)$ and (B1) correspond to (2.18), (2.19) and (2.20) respectively, while (A2) follows from (2.21) and (5.16) below. We are left to verify (B2). First, by Cauchy inequality and (2.21),

$$\begin{aligned}&\Big\|\frac{1}{n}\sum_{j=1}^n E\left(\boldsymbol{\psi}(Z_j) \otimes (\hat{\mathbf{u}}_n(Z_j) - \mathbf{v}_n(Z_j))\right)\Big\|^2 \\ &\le E(\|\boldsymbol{\psi}(Z)\|^2)\frac{1}{n}\sum_{j=1}^n E\left(\|\hat{\mathbf{u}}_n(Z_j) - \mathbf{v}_n(Z_j)\|^2\right) = o(m_n^{-1}),\end{aligned}$$

so that the $\mathbf{A}_n$ in (B2) satisfies

$$\mathbf{A}_n = E\left(\boldsymbol{\psi}(Z) \otimes \mathbf{v}_n(Z)\right) + o(m_n^{-1/2}).$$

Note that $\mathrm{trace}(\mathbf{U}_n) \le m_n |\mathbf{U}_n|_o = O(m_n)$ and

$$\begin{aligned} nE(\|\bar{\mathbf{v}}_n\|^2) &= E(\|\mathbf{v}_n(Z)\|^2) \le |\mathbf{W}_n^{1/2}|_o^2 E(\|\mathbf{W}_n^{-1/2}\mathbf{v}_n(Z)\|^2) \\ &= |\mathbf{W}_n^{1/2}|_o^2 \mathrm{trace}(\mathbf{U}_n). \end{aligned}$$

This shows

$$\|\bar{\mathbf{v}}_n\| = O_p(n^{-1/2} m_n^{1/2}). \tag{5.16}$$

By Cauchy inequality,

$$\|E\left(\boldsymbol{\psi}(Z) \otimes \mathbf{v}_n(Z)\right)\|^2 \le E\big(\|\boldsymbol{\psi}(Z)\|^2\big) E\big(\|\mathbf{v}_n(Z)\|^2\big).$$

But

$$\begin{aligned} E\big(\|\mathbf{v}_n(Z)\|^2\big) &\le |\mathbf{W}_n^{1/2}|_o^2 E\big(\|\mathbf{W}_n^{-1/2}\mathbf{v}_n(Z)\|^2\big) \\ &= |\mathbf{W}_n^{1/2}|_o^2 \mathrm{trace}(\mathbf{U}_n) = O(m_n). \end{aligned}$$

Hence

$$E\left(\boldsymbol{\psi}(Z) \otimes \mathbf{v}_n(Z)\right) = O(m_n^{1/2}).$$

Therefore combining the above and in view of (2.22) we arrive at

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} \mathbf{A}_n \mathbf{W}_n^{-1} \hat{\mathbf{u}}_n(Z_j) &= \big(E\left(\boldsymbol{\psi}(Z) \otimes \mathbf{v}_n(Z)\right) + o(m_n^{-1/2})\big) \mathbf{W}_n^{-1} \\ &\qquad \times \big(\bar{\mathbf{v}}_n + o_p(m_n^{-1/2} n^{-1/2})\big) \\ &= E\left(\boldsymbol{\psi}(Z) \otimes \mathbf{v}_n(Z)\right) \mathbf{W}_n^{-1} \bar{\mathbf{v}}_n + o_p(n^{-1/2}), \end{aligned}$$

Analogous to the proof of (B2) in Theorem 3 (or applying $\mathbf{u}_n = \mathbf{v}_n$), we have

$$E\left(\boldsymbol{\psi}(Z) \otimes \mathbf{v}_n(Z)\right) \mathbf{W}_n^{-1} \bar{\mathbf{v}}_n = \bar{\boldsymbol{\chi}} + o_p(n^{-1/2}),$$

where $\boldsymbol{\chi} = \Pi(\boldsymbol{\psi}|[\mathbf{v}_\infty])$ is the projection of $\boldsymbol{\psi}(Z)$ onto the closed linear span $[\mathbf{v}_\infty]$. Clearly $\int \boldsymbol{\chi}\, dQ = 0$ and $\int |\boldsymbol{\chi}|^2\, dQ < \infty$. Thus (B2) is proved with $\boldsymbol{\varphi} = \boldsymbol{\chi} = \Pi(\boldsymbol{\psi}|[\mathbf{v}_\infty])$. This finishes the proof. □

References

[1] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation in Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore.

[2] Bickel, P. J., Ritov, Y. and Wellner, J.A. (1991). Efficient estimation of linear functionals of a probability measure P with known marginal distributions *Ann. Statist.* **19**: 1316–1346.

[3] Bravo, F. (2010). 'Efficient M-estimators with auxiliary information'. *Journal of Statistical Planning and Inference* **140** 11, 3326–3342.

[4] Bollen, K. (1989). *Structural Equations with Latent Variables.* WHILEY/ JOHN WILEY & SONS, New York, Chichester, Brisbane, Toronto, Singapore.

[5] Brito, C. and Pearl, J (2002). A Graphical Criterion for the Identification of Causal Effects in Linear Models. *AAAI/IAAI.* **2002**: 533–539.

[6] Chatterjee, N. and Carroll R.J. (2005). Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* **92**: 399–418.

[7] Chaudhuri, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Ann. Statist.* **20**: 897 – 916.

[8] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91**: 862 – 872.

[9] Chen, Y., Dang, X., Peng, H, and Bart, H. (2009). Outlier Detection with the Kernelized Spatial Depth Function. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**: 288– 305.

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

[11] Drton, M., Foygel, R. and sullivant S. (2011). Global identifiability of linear structural equation models. *Ann. Statist.* **39** 865 – 886.

[12] Hellerstein, J. and Imbens, G. W. (1999). 'Imposing moment restrictions from auxiliary data by weighting'. *Review of Economics and Statistics* **81**: 1–14.

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

[14] Imbens, G. W. and Lancaster, T. (1994). Combining Micro and Macro Data in Microeconometric Models. *Review of Economic Studies* **61**: 655–680.

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

[16] Jöreskog, K.G. (1963). *Statistical estimation to factor analysis: A new technique and its foundation.*. Stockholm: Almqvist & Wiksell

[17] Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34**: 183 – 202.

[18] Jöreskog, K.G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Seminar

[19] Lin, Q. (2013). A Jakknife Empirical Likelihood Approach To Goodness

of Fit U-Statistic Testing With Side Information. *Ph.D. Dissertation.*
[20] Milasevic, P. and Ducharme, G.R. (1987). Uniqueness of The Spatial Median. *Ann. Statist.* **27**: 783–785.
[21] Mardia, K V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57(3)**: 519-530.
[22] Mardia, K V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *The Indian Journal of Statistics* **Series B, 1974**: 115-128.
[23] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**: 90–120.
[24] Owen, A. (2001). *Empirical Likelihood.* Chapman & Hall/CRC, London.
[25] Parente, P. M. D. C. and Smith, R. J. (2011). GEL methods for nonsmooth moment indicators. *Econometric Theory* **27**: 74–113. MR2771012
[26] Peng, H. and Schick, A. (2012). An empirical likelihood approach of goodness of fit testing. *Bernoulli* **19**: 954–981
[27] Peng, H. and Schick, A. (2005). Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown marginals: the least-squares approach. *J. Multiv. Anal.* **95**: 385–409.
[28] Peng, H. (2014). On A class of easy maximum empirical likelihood estimation. The *Manuscript* can be found at the URL: `http://www.math.iupui.edu/~hpeng/preprint.html.`
[29] Peng, H. and Schick, A. (2013). Maximum empirical likelihood estimation and related topics. The *Manuscript* can be found at the URL: `http://www.math.binghamton.edu/anton/preprint.html.`
[30] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**: 30–325.
[31] Révész, P.(1976). On Strong Approximation of The Multidimensional Empirical Process. *Ann. Probab.* **4**, 729–743.
[32] Rosenblatt, M.(1952). Remarks On A Multivariate Transformation. *Ann. Math. Statist.* **23**, 470–472.
[33] Shapiro, A.(2007). Statistical inference of moment structures. *Handbook of Latent Variable and Related Models (S.-Y. Lee, ed.)* , 229-260.
[34] Shiryaev, A.N. (1996). *Probability.* Second edition. Springer-Verlag New York Inc.
[35] Schumacker, R. E.and Lomax, R. G. (1989). *A beginner's guide to structural euqation modeling.* Lawerence Erlbaum Associates, Inc., Publishers. NJ.
[36] Tang, C. Y. and Leng, C. (2012). An empirical likelihood approach to quantile regression with auxiliary information. *Statist. & Probabil. Lett.* **82**: 29–36.
[37] Yuan, A., He, W., Wang, B. and Qin, G. (2012). U-statistics with side information. *J. Multiv. Analy.* **111**: 20 – 38.
[38] Zhang, B. (1995). M-estimation and quantile estimation in the presence of auxiliary information. *J. Statist. Plann. Infer.* **44**: 77 – 94.
[39] Zhang, B. (1997). Quantile processes in the presence of auxiliary information. *Ann. Inst. Statist. Math.* **49**: 35 – 55.

[40] Zuo, Y. and Serfling, Y. (2000). General notions of statistical depth function. *Ann. Statist.* **28**: 461 – 482.