Maximum Empirical Likelyhood Estimation of Linear Functionals Of A Probability Measure With Infinitely Many Constraints

Shan Wang and Hanxiang Peng

University of San Francisco
Department of Mathematics and Statistics
San Francisco, CA 94117, USA

Indiana University Purdue University Indianapolis
Department of Mathematical Sciences
Indianapolis, IN 46202-3267, USA

Abstract: In this article, we construct semiparametrically efficient estimators of linear functionals of a probability measure when side information is available using an easy empirical likelihood approach. We allow constraint functions to grow with the sample size and the use of estimated constraint functions. We focus on three cases of information which can be characterized by infinitely many constraints: (1) the marginal distributions are known, (2) the marginal distributions are unknown but identical, and (3) distributional symmetry. An improved spatial depth function is defined and its asymptotic properties are studied, simulation results on efficiency gain are reported.

AMS 2000 subject classifications: Primary 62D05; secondary 62J05, 62F12, 62F40.
Keywords and phrases: Empirical likelihood; Infinitely many constraints; Maximum empirical likelihood estimator; Semiparametric efficiency; Spatial median.

1. Introduction

Suppose that $Z_1, \dots, Z_n$ are independent and identically distributed (i.i.d.) random variables with a common distribution Q taking values in a measurable space $\mathcal{Z}$. In this article, 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 a vector function (constraint) $\mathbf{u}$ which satisfies

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

The commonly used sample mean $\bar{\boldsymbol{\psi}} = \frac{1}{n}\sum_{j=1}^n \boldsymbol{\psi}(Z_j)$ of $\boldsymbol{\theta} = E(\boldsymbol{\psi}(Z))$ does not use the information, and is not efficient in the sense of least dispersed regular

*Corresponding author

estimators, see e.g. Bickel, *et al.* (1993). Based on the criterion of maximum empirical likelihood, an improved estimator which utilizes the information is

$$\tilde{\boldsymbol{\theta}} = \frac{1}{n} \sum_{j=1}^{n} \frac{\boldsymbol{\psi}(Z_j)}{1 + \mathbf{u}(Z_j)^\top \tilde{\boldsymbol{\zeta}}}, \tag{1.1}$$

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

$$\sum_{j=1}^{n} \frac{\mathbf{u}(Z_j)}{1 + \mathbf{u}(Z_j)^\top \boldsymbol{\zeta}} = 0. \tag{1.2}$$

We shall refer to $\tilde{\boldsymbol{\theta}}$ as the *EL-weighted* estimator.

There is an extensive amount of literature on the empirical likelihood tests of hypothesis, see e.g. Owen (1988, 2001). Soon it was used to construct point estimators. Qin and Lawless (1994) studied maximum empirical likelihood estimators (MELE) and showed in Corollary 2 that MELE are fully efficient. As a special case of MELE, estimators of the preceding easy form were studied in Zhang (1995, 1997) in M-estimation and quantile processes in the presence of auxiliary information (side information). For a *fixed* number m of *known* constraint functions, the asymptotic normality (ASN) and efficiency of MELE were established.

Hjort, *et al.* (2009) extended the scope of the empirical likelihood hypothesis testing, and developed a general theory for constraints with nuisance parameters and considered the case with infinitely many constraints. Peng and Schick (2013) generalized the empirical likelihood testing to allow for the number of constraints to grow with the sample size and for the constraints to use estimated criteria functions. Tan and Peng (2018) expanded the results of the latter to U-statistics based general estimating equations.

Parente and Smith (2011) studied generalized empirical likelihood estimators for irregular constraints. Peng and Schick (2018) presented a theory of maximum empirical likelihood estimation and empirical likelihood ratio testing with irregular and estimated constraint functions. Wang and Peng (2022) used the easy EL-weighted approach to construct improved estimators of linear functionals of a probability measure when side information is available. Motivated by nuisance parameters common in semiparametric models and the infinite dimension of such models, they studied the use of estimated functions for growing number of constraints with the sample size. They applied the results to improve estimation efficiency in the structural equation models.

We shall rely the results of Wang and Peng (2022) to construct efficient estimators of linear functionals of a probability measure for a few cases of side information which is determined by infinitely many constrants. Bickel, *et al.* (1991) characterized efficient estimation of $E(h(X;Y))$ for known h when the marginal distributions of X and of Y are *known*, and construct an efficient estimator based on the criterion of minimum chisquare-type objective function. Peng and Schick (2005) calculated the information lower bound when the marginal distributions are unknown but identical, and constructed an efficient estimator based on the criterion of least squares objective. Here we construct the EL-weighted estimators and demonstrate the semiparametric efficiency. Note the simple analytic form of our estimators, and the property that it is convenient to incorporate side information to improve efficiency.

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) among others.

In what follows, we will summarize some results from Wang and Peng (2022) for the convenience of our use. Meanwhile, we provide the proof of the semiparametric effiency. In many semiparametric models, the constraint vector function $\mathbf{u} = (u_1, ..., u_m)^\top$ is usually unknown and must be estimated by some measurable function $\hat{\mathbf{u}} = (\hat{u}_1, ..., \hat{u}_m)^\top$. Using it, we now work with the EL-weights,

$$\hat{\pi}_j = \frac{1}{n}\frac{1}{1+\hat{\mathbf{u}}(Z_j)^\top\hat{\boldsymbol{\zeta}}}, \quad j=1,\dots,n, \tag{1.3}$$

where $\hat{\boldsymbol{\zeta}}$ solves Eqt (1.2) with $\mathbf{u} = \hat{\mathbf{u}}$. A natural estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ now is

$$\hat{\boldsymbol{\theta}} = \sum_{j=1}^{n}\hat{\pi}_j\boldsymbol{\psi}(Z_j) = \frac{1}{n}\sum_{j=1}^{n}\frac{\boldsymbol{\psi}(Z_j)}{1+\hat{\mathbf{u}}(Z_j)^\top\hat{\boldsymbol{\zeta}}}. \tag{1.4}$$

We now allow the number of constraints to depend on the sample size n, $m = m_n$, and tend to infinity slowly with n. To stress the dependence, 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,

$$\tilde{\boldsymbol{\theta}}_n = \frac{1}{n}\sum_{j=1}^{n}\frac{\boldsymbol{\psi}(Z_j)}{1+\mathbf{u}_n(Z_j)^\top\tilde{\boldsymbol{\zeta}}_n} \quad \text{and} \quad \hat{\boldsymbol{\theta}}_n = \frac{1}{n}\sum_{j=1}^{n}\frac{\boldsymbol{\psi}(Z_j)}{1+\hat{\mathbf{u}}_n(Z_j)^\top\hat{\boldsymbol{\zeta}}_n}, \tag{1.5}$$

where $\tilde{\boldsymbol{\zeta}}_n$ and $\hat{\boldsymbol{\zeta}}_n$ solve Eqt (1.2) with $\mathbf{u} = \tilde{\mathbf{u}}_n$ and $\mathbf{u} = \hat{\mathbf{u}}_n$, respectively,.

The ASN of $\tilde{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n$ are, respectively, given in Theorems 3 and 4 of Wang and Peng (2022), and we now prove the semiparametric efficiency of $\tilde{\boldsymbol{\theta}}_n$ and quote Theorem 4 in the Appendix for convenience of our use. For $\mathbf{a} \in \mathcal{R}^m$, write $\|\mathbf{a}\|$ the euclidean norm. For $\mathbf{a}, \mathbf{b} \in \mathcal{R}^m$, write $\mathbf{a} \otimes \mathbf{b}$ the kronecker product. Let $L_2^m(Q) = \big\{\mathbf{f} = (f_1, \dots, f_m)^\top : \int \|\mathbf{f}\|^2\, dQ < \infty\big\}$, and let $L_{2,0}^m(Q) = \big\{\mathbf{f} \in L_2^m(Q) : \int \mathbf{f}\, dQ = 0\big\}$. For $\mathbf{f} \in L_2^m(Q)$, write $\bar{\mathbf{f}} = n^{-1}\sum_{j=1}^n \mathbf{f}(Z_j)$ the sample average of $\mathbf{f}(Z_1), \dots, \mathbf{f}(Z_n)$, and $[\mathbf{f}]$ the closed linear span of the components $f_1, \dots, f_r$ in $L_2(Q)$. Let Z be an i.i.d. copy of Z_1. Denote by $[\mathbf{u}_\infty]$ the closed linear span of $\mathbf{u}_\infty = (u_1, u_2, \dots)$ in $L_{2,0}(Q)$. Set

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

Following Peng and Schick (2013), a sequence $\mathbf{W}_n$ of $m_n \times m_n$ dispersion matrices is said to be *regular* if

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

Theorem 1. *Suppose that $\mathbf{u}_n$ satisfies (C) for each $m = m_n$ such that*

$$\max_{1\le j\le n} \|\mathbf{u}_n(Z_j)\| = o_p(m_n^{-3/2} n^{1/2}), \tag{1.6}$$

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

$$|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o = o_p(m_n^{-1}), \tag{1.7}$$

$$\frac{1}{n}\sum_{j=1}^n \big(\boldsymbol{\psi}(Z_j)\otimes \mathbf{u}_n(Z_j) - E\big(\boldsymbol{\psi}(Z_j)\otimes \mathbf{u}_n(Z_j)\big)\big) = o_p(m_n^{-1/2}). \tag{1.8}$$

Then $\tilde{\boldsymbol{\theta}}_n$ is semiparametrically efficient as $m_n \to \infty$. Moroever,

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

where $\Sigma_0 = \mathrm{Var}(\boldsymbol{\psi}(Z)) - \mathrm{Var}(\boldsymbol{\varphi}_0(Z))$ with $\boldsymbol{\varphi}_0 = \Pi(\boldsymbol{\psi}|[\mathbf{u}_\infty])$.

Proof. We only need to show the efficiency. It suffices to prove that the orthonormal complement $\mathcal{T} = [\mathbf{u}_\infty]^\perp$ in $L_{2,0}(Q)$ is the tangent space. To this end, let $Q_t : |t| \le t_0$ with $Q_0 = Q$ be a regular parametric submodel with the score function a. By (C),

$$\int u\, dQ_t = 0, \quad u \in [\mathbf{u}_\infty].$$

Differentiating both sides of the equality with respect to t at $t = 0$ yields

$$\int ua\, dQ = 0, \quad u \in [\mathbf{u}_\infty].$$

This shows $a \in \mathcal{T}$. For any bounded $a \in \mathcal{T}$, consider $q_t = dQ_t/dQ = 1 + at$, $|t| \le t_0$ for sufficient small t_0. It is clear that q_t is a density and the submodel with the density has the score function a which satisfies $\int ua\, dQ = 0$. Since bounded functions in $\mathcal{T}$ are dense, it follows that the above conclusion holds for any $a \in \mathcal{T}$. This shows $\mathcal{T}$ is the tangent space. □

The article is organized as follows. In Section 2, the EL-weighted spatial depth function is constructed, and its ASN and efficiency are established when certain distributional symmetry is available. The ASN and efficiency of the EL-weighted estimators of linear functionals are proved when the marginal distribution functions are known in Section 3, and when the marginal distributions are unknown but equal in Section 4. The simulation results are reported in Section 5. Section 6 contains Theorem 4 of Wang and Peng (2022).

2. The EL-weighted spatial median

In this section, we introduce the EL-weighted spatial depth function, exhibit effciency and give the asymptotic normality.

The statistical depth functions provide a center-outward ordering of a point in $\mathcal{R}^p$ with respect to a distribution. High depth values correspond to centrality while low values to "outlyingness". They can be used to define multivariate medians and a depth-based median consists of the point(s) with the deepest depth. Common depth functions include the Tukey depth (halfspace depth)(Tukey, 1975), the simplicial depth (Liu, 1990), the projection depth (Zuo and Serfling, 2000), and the spatial depth (Chaudhuri, 1996). Depth functions possess robustness property. Here we shall use the empirical likelihood approach to construct more efficient depth functions. We shall illustrate this with the spatial depth function. The (population) depth function $D(\mathbf{x})$ with respect to a distribution F is defined as

$$D(\mathbf{x}) = 1 - \|E\big(\mathbb{S}(\mathbf{x}-\mathbf{X})\big)\|, \quad \mathbf{x} \in \mathcal{R}^p,$$

where $\mathbb{S}(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ if $\mathbf{x} \neq 0$ ($\mathbb{S}(0) = 0$) is the spatial sign function and $\mathbf{X}$ has the distribution function (DF) $F(\mathbf{x})$, denoted by $\mathbf{X} \sim F(\mathbf{x})$. The depth function $D(\mathbf{x})$ can be estimated by the sample depth function given by

$$D_n(\mathbf{x}) = 1 - \Big\|\frac{1}{n}\sum_{i=1}^n \mathbb{S}_{\mathbf{x}}(\mathbf{X}_i)\Big\|.$$

where $\mathbb{S}_{\mathbf{x}}(\mathbf{t}) = \mathbb{S}(\mathbf{t}-\mathbf{x})$. The sample spatial median $\mathbf{m}_n$ is defined as the value which maximizes the depth function, that is,

$$\mathbf{m}_n = \arg\max_{\mathbf{x}\in\mathcal{R}^p} D_n(\mathbf{x}) = \arg\min_{\mathbf{x}\in\mathcal{R}^p} \Big\|\frac{1}{n}\sum_{i=1}^n \mathbb{S}_{\mathbf{x}}(\mathbf{X}_i)\Big\|.$$

Suppose that there is available additional information that can be expressed by a constraint function $\mathbf{u}$. While the sample depth $D_n(\mathbf{x})$ does not utilize the information, the *EL-weighted depth function* $\widetilde{D}_n(\mathbf{x})$ makes use of it and is defined by

$$\widetilde{D}_n(\mathbf{x}) = 1 - \Big\|\frac{1}{n}\sum_{i=1}^n \frac{\mathbb{S}_{\mathbf{x}}(\mathbf{X}_i)}{1+\mathbf{u}(\mathbf{X}_i)^\top\tilde{\boldsymbol{\zeta}}}\Big\|, \quad \mathbf{x} \in \mathcal{R}^p, \tag{2.1}$$

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

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

The EL-weighted spatial median $\widetilde{\mathbf{m}}$ is defined as the value which maximizes the EL-weighted depth function, that is,

$$\widetilde{\mathbf{m}} = \arg\max_{\mathbf{x}\in\mathcal{R}^p} \widetilde{D}_n(\mathbf{x}) = \arg\min_{\mathbf{x}\in\mathcal{R}^p} \Big\|\frac{1}{n}\sum_{i=1}^n \frac{\mathbb{S}(\mathbf{x}-\mathbf{X}_i)}{1+\mathbf{u}(\mathbf{X}_i)^\top\tilde{\boldsymbol{\zeta}}}\Big\|. \tag{2.3}$$

The EL-weighted estimator of $\boldsymbol{\theta}(\mathbf{x}) = E(\mathbb{S}_{\mathbf{x}}(\mathbf{X}))$ is given by

$$\tilde{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{S}_{\mathbf{x}}(\mathbf{X}_i)}{1+\mathbf{u}(\mathbf{X}_i)^\top \tilde{\boldsymbol{\zeta}}}, \quad \mathbf{x} \in \mathbf{R}^p. \tag{2.4}$$

Known marginal medians. In our simulation study, we looked at the side information that the bivariate random vector $\mathbf{X} = (X_1, X_2)^\top$ has *known* marginal medians m_{10} and m_{20}. That is, the componentwise median $(m_{10}, m_{20})^\top$ is known. In this case, $\mathbf{u}(x_1, x_2) = (\mathbf{1}[x_1 \le m_{10}] - 1/2, \mathbf{1}[x_2 \le m_{20}] - 1/2)^\top$. We are motivated as follows. It is well known that the spatial median is a better location estimator than the componentwise median because the former takes into account the correlation of the components while the latter ignores it, see Chen, *et al.* (2009). We are intereted in how much information is lost when the componentwise median is used by looking at how much efficiency of the EL-weighted spatial median $\tilde{\mathbf{m}}$ (when the marginal medians are known) gains over the sample spatial median (when the marginal medians are unknown). In this case,

$$\mathbf{W} = E((\mathbf{u}\mathbf{u}^\top)(\mathbf{X})) = \begin{pmatrix} 1 & 4\alpha - 1 \\ 4\alpha - 1 & 1 \end{pmatrix},$$

where $\alpha = P(X_1 \le m_{10}, X_2 \le m_{20})$. Thus $\mathbf{W}$ is nonsingular iff $\alpha \ne 1/2$.

Growing number of constraints. Suppose that there exists some constant vector $\mathbf{c}$ such that $T = \mathbf{c}^\top \mathbf{X}$ is *symmetric* about some known value τ_0. Let $\varepsilon_j = \mathbf{c}^\top \mathbf{X}_j - \tau_0, j = 1, \ldots, n$. Then ε_j's are i.i.d. random variables which are symmetric about zero. Let ε be an i.i.d. copy of ε_j's, and let F be the distribution function of ε. Let $L_{2,0}(F, \text{odd})$ be the subspace of $L_{2,0}(F)$ consisting of the odd functions. Symmetry of ε about 0 implies

$$E(a(\varepsilon)) = 0, \quad a \in L_{2,0}(F, \text{odd}).$$

Let $s_k(t) = \sin(k\pi t), t \in [-1, 1], k = 1, 2, ...$ be the orthonormal trigonometric basis. Define $G(t) = 2F(t) - 1, t \in \mathcal{R}$. Then $G(t)$ is an odd function in $L_{2,0}(F, \text{odd})$, and $s_k(G(t)), k = 1, 2, ...$ form a basis of the space.

In this case, the constraints are $\mathbf{u}_n(\mathbf{X}_j) = (s_1(G(\varepsilon_j)), ..., s_{m_n}(G(\varepsilon_j)))^\top$, where we allow m_n to grow to infinity slowly with n. The EL-weighted depth function is calculated by (2.1) with $\mathbf{u} = \mathbf{u}_n$ and $\tilde{\boldsymbol{\zeta}} = \tilde{\boldsymbol{\zeta}}_n$ which solves Eqt (2.2) with $\mathbf{u} = \mathbf{u}_n$. The EL-weighted estimator of $\boldsymbol{\theta}(\mathbf{x}) = E(\mathbb{S}_{\mathbf{x}}(\mathbf{X}))$ then is

$$\tilde{\boldsymbol{\theta}}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{S}_{\mathbf{x}}(\mathbf{X}_i)}{1+\mathbf{u}(\mathbf{X}_i)^\top \tilde{\boldsymbol{\zeta}}_n}, \quad \mathbf{x} \in \mathbf{R}^p. \tag{2.5}$$

Theorem 2. *Suppose that F is continuous. Then for arbitrary but fixed $\mathbf{x} \in \mathbf{R}^p$, as $m_n \to \infty$ such that $m_n^4/n \to 0$, $\tilde{\boldsymbol{\theta}}_n(\mathbf{x})$ in (2.5) satisfies*

$$\tilde{\boldsymbol{\theta}}_n(\mathbf{x}) = \bar{\mathbb{S}}_{\mathbf{x}} - \bar{\boldsymbol{\varphi}}_{\mathbf{x}0} + o_p(n^{-1/2}),$$

where $\boldsymbol{\varphi}_{\mathbf{x}0} = \Pi(\mathbb{S}_{\mathbf{x}}(\mathbf{X}) | L_{2,0}(F, \text{odd}))$ is the projection of $\mathbb{S}_{\mathbf{x}}(\mathbf{X})$ onto $L_{2,0}(F, \text{odd})$. As a consequence, if $\Sigma_0(\mathbf{x}) = \text{Var}(\mathbb{S}_{\mathbf{x}}(\mathbf{X})) - \text{Var}(\boldsymbol{\varphi}_{\mathbf{x}0}(\mathbf{X}))$ is nonsingular,

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

Proof of Theorem 2. We shall apply Theorem 1 to prove the result. Since $\mathbf{W}_n = E((\mathbf{u}\mathbf{u}^\top)(\mathbf{X})) = \mathbf{I}_{m_n}$ is the identity matrix, it follows that (C) holds and $\mathbf{W}_n$ is regular. As $\|\mathbf{u}_n(\mathbf{X}_j)\| \le m_n^{1/2}$ for each j and $m_n^4/n = o(1)$, (1.6) is satisfied, while (1.7) holds in view of the inequalities

$$nE(|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o^2) \le E(\|\mathbf{u}_n(\mathbf{X})\|^4) \le m_n^2.$$

Let $\mathbf{K}_n$ be the left hand side of (1.8). Then (1.8) follows from

$$nE(\|\mathbf{K}_n\|^2) \le E(\|\mathbb{S}_{\mathbf{x}}(\mathbf{X}) \otimes \mathbf{u}_n(\mathbf{X})\|^2) \le m_n E(\|\mathbb{S}_{\mathbf{x}}(\mathbf{X})\|^2) = m_n.$$

We now apply Theorem 1 to complete the proof. □

Efficiency gain and ASN for $\tilde{\mathbf{m}}$. By the properties of empirical likelihood, one concludes that $\widetilde{D}_n(\mathbf{x})$ is a valid depth function at least for large n as all $1+\mathbf{u}(\mathbf{X}_i)^\top\tilde{\boldsymbol{\zeta}} > 0$. Fix $\mathbf{x} \in \mathcal{R}^p$, let $\mathbf{P}_{\mathbf{x}}$ be the projection of $\mathbb{S}_{\mathbf{x}}(\mathbf{X})$ onto the closed linear span $[\mathbf{u}_\infty] = L_{2,0}(F, odd)$. Then $\Sigma_0(\mathbf{x}) = \mathrm{Var}(\mathbb{S}_{\mathbf{x}}(\mathbf{X})) - \mathrm{Var}(\mathbf{P}_{\mathbf{x}}(\mathbf{X}))$. Clearly,

$$\mathrm{Var}(\mathbf{P}_{\mathbf{x}}(\mathbf{X})) = E\big(\mathbb{S}_{\mathbf{x}}(\mathbf{X})v \otimes \mathbf{u}(\mathbf{X})^\top\big)\mathbf{W}^{-1}E\big(\mathbb{S}_{\mathbf{x}}(\mathbf{X}) \otimes \mathbf{u}(\mathbf{X})\big). \tag{2.6}$$

Let $\mathbb{S}_2(\mathbf{x}) = \mathbb{S}\big(E(\mathbb{S}_{\mathbf{x}}(\mathbf{X}))\big)$. If $\mathbf{W}_0(\mathbf{x}) := \mathbb{S}_2(\mathbf{x})\mathbf{V}_0(\mathbf{x})\mathbb{S}_2(\mathbf{x})^\top$ is nonsingular, then by Theorem 2 for fixed $\mathbf{x} \in \mathbf{R}$,

$$\sqrt{n}(\widetilde{D}_n(\mathbf{x}) - D(\mathbf{x})) \Longrightarrow \mathscr{N}(0, \mathbf{W}_0(\mathbf{x})).$$

Note that the sample depth $D_n(\mathbf{x})$ satisfies

$$\sqrt{n}(D_n(\mathbf{x}) - D(\mathbf{x})) \Longrightarrow \mathscr{N}(0, \mathbf{W}(\mathbf{x})),$$

where $\mathbf{W}(\mathbf{x}) = \mathbb{S}_2(\mathbf{x})\,\mathrm{Var}(\mathbb{S}_{\mathbf{x}}(\mathbf{X}))\mathbb{S}_2(\mathbf{x})^\top$. Thus the reduction of the asymptotic variance-covariance of the EL-weighted depth $\widetilde{D}_n(\mathbf{x})$ is

$$\mathbb{S}_2(\mathbf{x})\,\mathrm{Var}(\mathbf{P}_{\mathbf{x}}(\mathbf{X}))\mathbb{S}_2(\mathbf{x})^\top.$$

We now use the Delta method to drive the ASN of the EL-weighted spatial median $\tilde{\mathbf{m}}$. To this end, we need some results from Chaudhuri (1992) in the case of $m = 1$ for which the spatial median corresponds to his multivariate Hodges-Lehmann type location estimate. The following is his Assumption 3.1.

(PC) $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d random vectors in $\mathcal{R}^d$ with an absolutely continuous (with respect to the Lebesgue measure) distribution having a density f that is bounded on every bounded subset of $\mathcal{R}^d$.

Assume (PC) and $d \ge 2$. Let $\mathbf{H}(\mathbf{x}) = \|\mathbf{x}\|^{-1}(\mathbf{I}_d - \mathbf{x}\mathbf{x}^\top/\|\mathbf{x}\|^2)$ if $\mathbf{x} \ne 0$ and $\mathbf{H}(0) = 0$. Note that $\mathbb{S}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$ are the first and second order partial derivatives of $\|\mathbf{x}\|$. Under (PC), the underlying distribution is absolutely continuous with respect to the Lebesgue measure on $\mathcal{R}^p (d \ge 2)$, hence the (population) spatial median $\mathbf{m}_0$ uniquely exists and satisfies the equation $E(\mathbb{S}(\mathbf{m}_0 - \mathbf{X})) = 0$. The spatial median $\mathbf{m}_n$ satisfies

$$\sum_{i=1}^{n} \mathbb{S}(\mathbf{m}_n - \mathbf{X}_i) = 0.$$

Let $\mathbf{J} = E\big((\mathbb{S}\mathbb{S}^\top)(\mathbf{m}_0 - \mathbf{X})\big)$ and $\mathbf{K} = E\big(\mathbf{H}(\mathbf{m}_0 - \mathbf{X})\big)$. Chaudhuri (1992) showed in his Theorem 3.3 and its corollary that if (PC) holds then the matrices $\mathbf{J}$ and $\mathbf{K}$ are positive definite and $\mathbf{m}_n$ satisfies

$$\sqrt{n}(\mathbf{m}_n - \mathbf{m}_0) \Longrightarrow \mathscr{N}(0,\, \mathbf{K}^{-1}\mathbf{J}\mathbf{K}^{-\top}).$$

Note that the EL-weighted spatial median $\widetilde{\mathbf{m}}_n$ satisfies the equation,

$$\sum_{i=1}^{n} \frac{\mathbb{S}(\mathbf{m} - \mathbf{X}_i)}{1 + \mathbf{u}(\mathbf{X}_i)^\top \tilde{\boldsymbol{\zeta}}} = 0.$$

Using the Delta method, we derive, with $\mathbf{V}_0(\mathbf{m}_0) = \mathbf{J} - \mathrm{Var}(\mathbf{P}_{\mathbf{m}_0}(\mathbf{X}))$,

$$\sqrt{n}(\widetilde{\mathbf{m}}_n - \mathbf{m}_0) \Longrightarrow \mathscr{N}(0,\, \mathbf{K}^{-1}\mathbf{V}_0(\mathbf{m}_0)\mathbf{K}^{-\top}),$$

where $\mathrm{Var}(\mathbf{P}_{\mathbf{m}_0}(\mathbf{X}))$ is calculated by (2.6).

Growing number of estimated constraints. For unknown $F(x)$, we estimate it by the symmetrized empirical distribution function,

$$\mathbb{F}(x) = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{1}[\varepsilon_j \le x] + \mathbf{1}[-\varepsilon_j \le x]), \quad x \in \mathcal{R}.$$

Let $\mathbb{G}(x) = 2\mathbb{F}(x) - 1$. We thus obtain computable functions $s_j(\mathbb{G}(x))$. Write $\mathbf{u}_n$ for $\mathbf{u}$, and estimate it by $\hat{\mathbf{u}}_n(x) = (s_1(\mathbb{G}(x)), ..., s_{m_n}(\mathbb{G}(x)))^\top$. The EL-weighted estimator of $\boldsymbol{\theta}(\mathbf{x}) = E(\mathbb{S}_{\mathbf{x}}(X))$ is now given by

$$\hat{\boldsymbol{\theta}}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{S}_{\mathbf{x}}(\mathbf{X}_i)}{1 + \hat{\mathbf{u}}_n(\mathbf{X}_i)^\top \hat{\boldsymbol{\zeta}}_n}, \quad \mathbf{x} \in \mathcal{R}^p, \tag{2.7}$$

where $\hat{\boldsymbol{\zeta}}_n$ solves Eqt (2.2) with $\mathbf{u} = \hat{\mathbf{u}}_n$. We have

Theorem 3. *Suppose that F is continuous. Then $\hat{\boldsymbol{\theta}}_n$ defined in (2.7) satisfies the conclusions of Theorem 2 as $m_n \to \infty$ such that $m_n^6/n \to 0$.*

Proof of Theorem 3. We shall use Theorem 6 of Wang and Peng (2022) for the proof. First, (C) is satisfied with $\mathbf{W}_n$ regular as $\mathbf{W}_n = I_{m_n}$. Next, (6.1) follows from $\|\hat{\mathbf{u}}_n(Z_j)\|^2 \le m_n$ and $m_n^4/n = o(1)$. Let

$$\hat{\mathbf{W}}_n = \frac{1}{n} \sum_{j=1}^{n} (\hat{\mathbf{u}}_n \hat{\mathbf{u}}_n^\top)(\mathbf{X}_j), \quad \bar{\mathbf{W}}_n = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{u}_n \mathbf{u}_n^\top)(\mathbf{X}_j)^\top.$$

Then $\bar{\mathbf{W}}_n - \mathbf{W}_n = o_p(m_n^{-1})$ follows from $m_n^4/n = o(1)$ and

$$nE(|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o^2) \le E(\|\mathbf{u}_n(\mathbf{Z}_1)\|^4) \le m_n^2.$$

Let $D_n = n^{-1}\sum_{j=1}^n \|\hat{\mathbf{u}}_n(\mathbf{X}_j) - \mathbf{u}_n(\mathbf{X}_j)\|^2$. It is easy to see

$$|\hat{\mathbf{W}}_n - \bar{\mathbf{W}}_n|_o \le D_n + 2|\bar{\mathbf{W}}_n|_o^{1/2} D_n^{1/2}.$$

Thus (6.2) follows from $D_n = o_p(m_n^{-2})$ to be shown next. To this end, let $\mathbf{s}_n = (s_1, ..., s_{m_n})^\top$. Then $\|\mathbf{s}_n(t)\| \le m_n^{1/2}$. One verifies $\|\boldsymbol{\psi}_n'(t)\| \le a m_n^{3/2}$ for some constant a. Therefore, $D_n = o_p(m_n^{-2})$ follows from $D_n = O_p(m_n^3/n)$ and $m_n^5/n = o(1)$, in view of

$$\frac{1}{n}\sum_{j=1}^n \|\mathbf{s}_n(\mathbb{G}(\mathbf{X}_j)) - \mathbf{s}_n(G(\mathbf{X}_j))\|^2 \le a m_n^3 \sup_{t\in\mathcal{R}} |\mathbb{G}(t) - G(t)|^2 = O_p(m_n^3/n).$$

Denoting $\boldsymbol{\psi}(\mathbf{y}) = \mathbb{S}_{\mathbf{x}}(\mathbf{y})$, we break

$$\frac{1}{n}\sum_{j=1}^n \Big(\boldsymbol{\psi}(\mathbf{X}_j) \otimes \hat{\mathbf{u}}_n(\mathbf{X}_j) - E\big(\boldsymbol{\psi}(\mathbf{X}_j) \otimes \mathbf{u}_n(\mathbf{X}_j)\big)\Big) = \mathbf{J}_n + \mathbf{K}_n, \quad \text{where}$$

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

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

By Cauchy inequality,

$$\begin{aligned} E(\|\mathbf{J}_n\|^2) &\le E(\|\boldsymbol{\psi}(\mathbf{X}_1)\|^2)\frac{1}{n}\sum_{j=1}^n E(\|\hat{\mathbf{u}}_n(\mathbf{X}_j) - \mathbf{u}_n(\mathbf{X}_j)\|^2) \\ &= E(D_n) = O(m_n^3/n) = o(m_n^{-1}), \end{aligned}$$

as $m_n^4/n = o(1)$. We now bound the variance by the second moment to get

$$E(\|\mathbf{K}_n\|^2) \le \frac{1}{n}E(\|\boldsymbol{\psi}(\mathbf{X}_1) \otimes \mathbf{u}_n(\mathbf{X}_1)\|^2) \le 4\frac{m_n}{n} = o(m_n^{-1})$$

as $m_n^2/n = o(1)$. Taken together we prove (6.3) – (6.4).

We now show that (6.5) holds with $\mathbf{v}_n = \mathbf{u}_n$. To this end, using Taylor expansion we write

$$\frac{1}{n}\sum_{j=1}^n (\hat{\mathbf{u}}_n(\mathbf{X}_j) - \mathbf{u}_n(\mathbf{X}_j)) = \mathbf{L}_n + \mathbf{m}_n, \quad \text{where}$$

$$\mathbf{L}_n = \frac{1}{n}\sum_{j=1}^n \boldsymbol{\psi}_n'(G(\varepsilon_j))\big(\mathbb{G}(\varepsilon_j) - G(\varepsilon_j)\big),\ \mathbf{m}_n = \frac{1}{n}\sum_{j=1}^n \boldsymbol{\psi}_n''(G_{nj}^*)\big(\mathbb{G}(\varepsilon_j) - G(\varepsilon_j)\big)^2,$$

where G^*_{nj} lies in between $\mathbb{G}(\varepsilon_j)$ and $G(\varepsilon_j)$. It thus follows

$$\begin{aligned} E\big(\|\mathbf{L}_n\|^2\big) &\le \frac{1}{n} E\Big(\|\boldsymbol{\psi}'_n(G(\varepsilon_1))\|^2 \big(\mathbb{G}(\varepsilon_j) - G(\varepsilon_j)\big)^2\Big) \\ &\le a\frac{m_n^3}{n} \sup_{t\in\mathcal{R}} \big(\mathbb{G}(t) - G(t)\big)^2 \\ &= O_p(m_n^3/n^2) = o_p((m_n n)^{-1}) \end{aligned}$$

as $m_n^4/n = o(1)$. This shows $\mathbf{L}_n = o_p((m_n n)^{-1/2})$. One has $\|\boldsymbol{\psi}''(t)\| = O_p(m_n^{5/2})$. Using this, we get

$$\|\mathbf{m}_n\| \le O(m_n^{5/2}) \sup_{t\in\mathcal{R}} |\mathbb{G}(t) - G(t)|^2 = O_p(m_n^{5/2}/n) = o_p((m_n n)^{-1/2})$$

as $m_n^6/n = o(1)$. This yields $\mathbf{m}_n = o_p((m_n n)^{-1/2})$. Taken together the desired (6.5) follows. We now apply Theorem 6 to finish the proof. □

3. Efficient estimation of linear functionals with known marginals

Suppose that there is available the information that the marginal distributions F and G of Q are *known*. This can be charaterized by

$$\begin{aligned} \int c(x)\, dQ(x,y) &= \int c(x)\, dF(x) = 0, \quad c \in L_{2,0}(F), \\ \int d(y)\, dQ(x,y) &= \int d(y)\, dG(y) = 0, \quad d \in L_{2,0}(G). \end{aligned}$$

Bickel, *et al.* (1991) and Peng and Schick (2002) constructed efficient estimators of the linear functional $\theta = \int \psi\, dQ$, and proved the ASN under the assumption,

(K) There exists $\rho > 0$ such that for arbitrary measurable sets A and B,

$$P(X \in A, Y \in B) \ge \rho F(A)G(B).$$

Bickel, *et al.* (1991) showed that the project of $\psi \in L_2(Q)$ onto the sum space $L_{2,0}(F) + L_{2,0}(G)$ uniquely exists. They demonstrated that the asymptotic variance of the efficient estimator $\tilde\theta$ of θ can be substantially less than that of the empirical estimator $n^{-1}\sum_{j=1}^n \psi(X_j, Y_j)$. For example, they showed that the empirical DF $n^{-1}\sum_{j=1}^n \mathbf{1}[X_j \le 1/2, Y_j \le 1/2]$ of $\theta = P(X \le 1/2, Y \le 1/2)$ (taking $\psi_{s,t}(x,y) = \mathbf{1}[x \le s, y \le t]$) has three times the asymptotic variance of the efficient estimator $\tilde\theta$ of θ in the case that F and G are uniform distributions over $[0,1]$ and X, Y are independent

Here we propose an efficient estimator based on maximum empirical likelihood. Employing a basis $\{c_k\}$ of $L_{2,0}(F)$ and $\{d_k\}$ of $L_{2,0}(G)$, we can reduce the uncountably many characterizing equations to countably many ones,

$$\int c_k(x)\, dF(x) = 0, \quad \int d_k(y)\, dG(y) = 0, \quad k = 1, 2, \dots. \tag{3.1}$$

Suppose that F and G are continuous. This allows us to take $c_k = b_k(F)$ and $d_k = b_k(G)$, where $b_k(t)$ are the trigonometric basis,

$$b_k(t) = \sqrt{2}\cos(k\pi t), \quad t \in [0,1], k = 1, 2, \dots. \tag{3.2}$$

That is, $\{c_k\}$ and $\{d_k\}$ are bases of $L_{2,0}(F)$ and $L_{2,0}(G)$, respectively. Using the first $2m_n$ terms as constraints, the EL-weighted estimator of $\boldsymbol{\theta}$ is

$$\hat{\theta}_n = \frac{1}{n}\sum_{j=1}^{n} \frac{\psi(\mathbf{Z}_j)}{1 + \boldsymbol{\zeta}_n^\top \mathbf{u}_n(\mathbf{Z}_j)}, \tag{3.3}$$

where $\mathbf{u}_n(x,y) = (\mathbf{b}_n(F(x))^\top, \mathbf{b}_n(G(y))^\top)^\top$ with $\mathbf{b}_n = (b_1, ..., b_{m_n})^\top$. Using Theorem 1, we prove

Theorem 4. *Suppose that F and G are continuous. Assume (K). Then, as $m_n \to \infty$ such that $m_n^4/n \to 0$,*

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

where φ_0 is the projection of ψ onto the sum space $L_{2,0}(F) + L_{2,0}(G)$. Hence,

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

where $\Sigma = \mathrm{Var}(\psi(\mathbf{Z})) - \mathrm{Var}(\varphi_0(\mathbf{Z}))$.

Remark 1. By Bickel, *et al.* (1991) (pages 1328–29), the estimator $\tilde{\theta}_n$ in (3.3) of $\theta = \int \psi(x,y)\, dQ(x,y)$ is semiparametrically efficient.

Proof of Theorem 4. We shall rely on Theorem 1. Since $\|\mathbf{u}_n\| \le 2\sqrt{m_n}$ and $m_n^4/n = o(1)$, it follows that (1.6) holds. Thus

$$nE(|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o^2) \le E(\|\mathbf{u}_n(\mathbf{Z})\|^4) \le 16m_n^2 = o_p(m_n^{-2})$$

as $m_n^4/n = o(1)$. This shows (1.7). Let

$$\mathbf{K}_n = \frac{1}{n}\sum_{j=1}^{n} \Big(\psi(\mathbf{Z}_j) \otimes \mathbf{u}_n(\mathbf{Z}_j) - E\big(\psi(\mathbf{Z}_j) \otimes \mathbf{u}_n(\mathbf{Z}_j)\big)\Big). \tag{3.4}$$

It follows from $m_n^2/n = o(1)$ that (1.8) holds in view of

$$E(\|\mathbf{K}_n\|^2) \le \frac{1}{n}E(\|\psi(\mathbf{Z}_1) \otimes \mathbf{u}_n(\mathbf{Z}_1)\|^2) \le 4\frac{m_n}{n}E(|\psi(\mathbf{Z}_1)|^2) = o(m_n^{-1}).$$

We are now left to prove the regularity of $\mathbf{W}_n$. Since $\mathbf{b}_n$ are the first m_n terms of the orthonormal basis $\{b_k\}$, it follows that $E(\mathbf{b}_n(F(X))\mathbf{b}_n(F(X))^\top) = \mathbf{I}_{m_n}$. The same holds for G. Let $\mathbf{C}_n = E(\mathbf{b}_n(F(X))\mathbf{b}_n(G(Y))^\top)$. Then $\mathbf{W}_n$ is the $2m_n \times 2m_n$ dispersion matrix whose (1,1)- and (2, 2)-blocks are equal to $\mathbb{I}_{m_n}$

and the (1,2)-block equal to $\mathbf{C}_n$. For $\mathbf{s},\mathbf{t}\in\mathcal{R}^{m_n}$ with $\|\mathbf{s}\|^2+\|\mathbf{t}\|^2=1$, set $\mathbf{r}=(\mathbf{s}^\top,\mathbf{t}^\top)^\top$. We have

$$\mathbf{r}^\top\mathbf{W}_n\mathbf{r}=\|\mathbf{s}\|^2+\|\mathbf{t}\|^2+2\mathbf{s}^\top\mathbf{C}_n\mathbf{t}. \tag{3.5}$$

By Cauchy inequality,

$$\begin{aligned}(\mathbf{s}^\top\mathbf{C}_n\mathbf{t})^2 &\le \mathbf{s}^\top E(\mathbf{b}_n(F(X))\mathbf{b}_n(F(X))^\top)\mathbf{s}\,\mathbf{t}^\top E(\mathbf{b}_n(G(Y))\mathbf{b}_n(G(Y))^\top)\mathbf{t}\\ &= \|\mathbf{s}\|^2\|\mathbf{t}\|^2\le 1.\end{aligned}$$

It thus follows from (3.5) that $\mathbf{r}^\top\mathbf{W}_n\mathbf{r}\le 4$ uniformly in n and the above $\mathbf{r}$. For $a\in L_{2,0}(F)$ and $b\in L_{2,0}(G)$, (K) implies

$$\begin{aligned}\int (a(x)-b(y))^2\,dQ(x,y) &\ge \rho\int (a(x)-b(y))^2 dF(x)dG(y)\\ &= \rho\big(\int a^2\,dF+\int b^2\,dG\big).\end{aligned}$$

Thus

$$2\int ab\,dQ\le (1-\rho)\big(\int a^2\,dF+\int b^2\,dG\big).$$

Replacing a with $-a$ yields

$$2\int ab\,dQ\ge -(1-\rho)\big(\int a^2\,dF+\int b^2\,dG\big)$$

Taking $a=\mathbf{s}^\top\mathbf{b}_n(F)$ and $b=\mathbf{b}_n(G)^\top\mathbf{t}$ and noticing

$$\int a^2\,dF=\|\mathbf{s}\|^2,\quad \int b^2\,dG=\|\mathbf{t}\|^2,$$

we derive

$$2\mathbf{s}^\top\mathbf{C}_n\mathbf{t}=2\int \mathbf{s}^\top\mathbf{b}_n(F(x))\mathbf{b}_n(G(y))^\top\mathbf{t}\,dQ(x,y)\ge -(1-\rho)(\|\mathbf{s}\|^2+\|\mathbf{t}\|^2).$$

By (3.5), we thus arrive at

$$\mathbf{r}^\top\mathbf{W}_n\mathbf{r}\ge \|\mathbf{s}\|^2+\|\mathbf{t}\|^2-(1-\rho)(\|\mathbf{s}\|^2+\|\mathbf{t}\|^2)=\rho>0.$$

Taken together we prove the regularity of $\mathbf{W}_n$, and apply Theorem 1 to complete the proof. □

4. Efficient estimation of linear functionals with equal marginals

Suppose that the marginal distributions F and G of X and Y are *equal but unknown*. This is equivalent to the assertion that

$$E(a_k(X)-a_k(Y))=0,\quad k=1,2,\dots, \tag{4.1}$$

where $\{a_k\}$ is an orthonormal basis of $L_{2,0}(H)$ with $H = (F+G)/2$. Assume that F and G are continuous. This allows us take $a_k(x) = b_k(H(x))$ under the assumption $F = G = H$, where $\{b_k\}$ is the trigonometric basis in (3.2). As H is unknown, we estimate it by the pooled empirical distribution function,

$$\mathbb{H}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{2}(\mathbf{1}[X_j \le x] + \mathbf{1}[Y_j \le x]), \quad x \in \mathcal{R}.$$

This gives us computable functions $b_k(\mathbb{H}(x))$. Let $\mathbf{u}_n(x,y) = \mathbf{b}_n(H(x)) - \mathbf{b}_n(H(y)), x, y \in \mathcal{R}$. This is unknown and can be estimated by $\hat{\mathbf{u}}_n(x,y) = \mathbf{b}_n(\mathbb{H}(x)) - \mathbf{b}_n(\mathbb{H}(y))$. Using the first m_n terms as constraints, the EL-weighted estimator of $\theta = E(\psi(X,Y))$ is given by

$$\hat{\theta}_n = \frac{1}{n}\sum_{j=1}^{n}\frac{\psi(X_j,Y_j)}{1+\hat{\boldsymbol{\zeta}}_n^\top \hat{\mathbf{u}}_n(X_j,Y_j)}, \tag{4.2}$$

where $\hat{\boldsymbol{\zeta}}_n$ is the solution to Eqt (1.2) with $\mathbf{u} = \hat{\mathbf{u}}_n$.

Peng and Schick (2005) constructed efficient estimators of linear functionals of a bivariate distribution with equal marginals under the condition,

$$\inf_{a\in A} E[(a(X) - a(Y))^2] > 0, \tag{4.3}$$

where $\mathbb{A} = \{a \in L_{2,0}(H) : \int a^2\, dH = 1\}$ is the unit sphere in $L_{2,0}(H)$. They exhibited that the asymptotic variance of an efficient estimator of θ is about 1/3 of that of the empirical estimator or smaller.

Applying Theorem 6, we show that $\hat{\theta}_n$ is efficient.

Theorem 5. *Suppose that the distribution functions F and G are equal and continuous. Assume (4.3). Then, as $m_n \to \infty$ such that $m_n^6/n \to 0$, $\hat{\theta}_n$ given in (4.2) satisfies*

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

where φ is the projection of ψ onto $\mathbb{A}$. Thus

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

where $\Sigma = \mathrm{Var}(\psi) - \mathrm{Var}(\varphi)$.

Remark 2. By Theorem 3 of Peng and Schick (2005), the estimator $\hat{\theta}_n$ given in (4.2) of $\theta = \int \psi(x,y)\, dQ(x,y)$ is semiparametrically efficient.

Proof of Theorem 5. We shall apply Theorem 6. Recalling the trigonometric basis $\{b_k\}$ in (3.2), one readily verifies that $\mathbf{b}_n = (b_1, \dots, b_{m_n})^\top$ has the properties,

$$\|\mathbf{b}_n\| \le (2m_n)^{1/2}, \quad \|\mathbf{b}_n'\| \le \sqrt{2}\pi m_n^{3/2}, \quad \|\mathbf{b}_n''\| \le \sqrt{2}\pi^2 m_n^{5/2}, \tag{4.4}$$

where $\mathbf{b}_n'$ and $\mathbf{b}_n''$ denote the first and second order derivatives of $\mathbf{b}$.

Recalling $\mathbf{u}_n(x,y) = \mathbf{b}_n(H(x)) - \mathbf{b}_n(H(y))$ and $\hat{\mathbf{u}}_n(x,y) = \mathbf{b}_n(\mathbb{H}(x)) - \mathbf{b}_n(\mathbb{H}(y))$, one gets by the first inequality in (4.4) that

$$\|\mathbf{u}_n\| \le 2\sqrt{2}\sqrt{m_n}, \quad \|\hat{\mathbf{u}}_n\| \le 2\sqrt{2}\sqrt{m_n}. \tag{4.5}$$

Hence (6.1) holds as $m^4/n = o(1)$. Noting $\mathbf{W}_n = E(\mathbf{u}_n(\mathbf{Z})\mathbf{u}_n(\mathbf{Z})^\top)$, one has by (4.3) that

$$\boldsymbol{\lambda}^\top \mathbf{W}_n \boldsymbol{\lambda} = E\big((\boldsymbol{\lambda}^\top \mathbf{b}_n(H(X)) - \boldsymbol{\lambda}^\top \mathbf{b}_n(H(Y)))^2\big) \ge \inf_{a\in A} E\big((a(X) - a(Y))^2\big) > 0$$

uniformly in n and $\|\boldsymbol{\lambda}\| = 1$ as both $\boldsymbol{\lambda}^\top \mathbf{b}_n(H(X))$ and $\boldsymbol{\lambda}^\top \mathbf{b}_n(H(Y))$ live in A. Moreover,

$$\boldsymbol{\lambda}^\top \mathbf{W}_n \lambda \le 4E\big((\boldsymbol{\lambda}^\top \mathbf{b}_n(H(X)))^2\big) = 4.$$

Thus $\mathbf{W}_n$ is regular. Let

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

Then by the first equality in (4.5),

$$nE(|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o^2) \le E(|\mathbf{u}_n(\mathbf{Z}_1)|^4) \le 64 m_n^2.$$

Hence $\bar{\mathbf{W}}_n - \mathbf{W}_n = o_p(m_n^{-1})$ as $m_n^4/n = o(1)$. It can be seen

$$|\bar{\mathbf{W}}_n - \mathbf{W}_n|_o \le D_n + 2|\bar{\mathbf{W}}_n|_o^{1/2} D_n^{1/2},$$

where $D_n = n^{-1}\sum_{j=1}^n \|\hat{\mathbf{u}}_n(\mathbf{Z}_j) - \mathbf{u}_n(\mathbf{Z}_j)\|^2$. Thus (6.2) is implied by

$$D_n = o_p(m_n^{-2}). \tag{4.6}$$

Using the second inequality in (4.4), we derive

$$\frac{1}{n}\sum_{j=1}^n |\mathbf{b}_n(\mathbb{H}(\mathbf{Z}_j)) - \mathbf{b}_n(H(\mathbf{Z}_j))|^2 \le 2\pi^2 m_n^3 \sup_{t\in\mathcal{R}} |\mathbb{H}(t) - H(t)| = O_p(m_n^3/n).$$

Hence $D_n = O_p(m_n^3/n)$ and (4.6) holds as $m_n^5/n = o(1)$. We break

$$\frac{1}{n}\sum_{j=1}^n \Big(\psi(\mathbf{Z}_j)\otimes \hat{\mathbf{u}}_n(\mathbf{Z}_j) - E\big(\psi(\mathbf{Z}_j)\otimes \mathbf{u}_n(\mathbf{Z}_j)\big)\Big) = J_n + K_n,$$

where

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

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

By Cauchy inequality,

$$E(\|\mathbf{J}_n\|^2) \le E(|\psi(\mathbf{Z}_1)|^2)\frac{1}{n}\sum_{j=1}^n E(\|\hat{\mathbf{u}}_n(\mathbf{Z}_j) - \mathbf{u}_n(\mathbf{Z}_j)\|^2) = E(\|\mathbf{J}_n\|^2)E(D_n)$$
$$= O(m_n^3/n) = o(m_n^{-1})$$

where the last equality holds as $m_n^4/n = o(1)$. We now bound the variance by the second moment and by the first equality in (4.5) to get

$$E(\|\mathbf{K}_n\|^2) \le \frac{1}{n}E(|\psi(Z_1)\otimes \mathbf{u}_n(Z_1)|^2) \le 8\frac{m_n}{n}E(|\psi(Z_1)|^2) = o(m_n^{-1})$$

as $m_n^2/n = o(1)$. Taken together (6.3) follows. We now show (6.5) holds with $\mathbf{v}_n = \mathbf{u}_n$. Using Taylor's expansion, we write

$$\frac{1}{n}\sum_{j=1}^n \big(\mathbf{b}_n(\mathbb{H}(X_j)) - \mathbf{b}_n(H(X_j))\big) = \mathbf{L}_n + \mathbf{M}_n,$$

where

$$\mathbf{L}_n = \frac{1}{n}\sum_{j=1}^n \mathbf{b}_n'(H(X_j))\big(\mathbb{H}(X_j) - H(X_j)\big),$$
$$\mathbf{M}_n = \frac{1}{n}\sum_{j=1}^n \mathbf{b}_n''(H_{nj}^*)\big(\mathbb{H}(X_j) - H(X_j)\big)^2,$$

where H_{nj}^* lies in between $\mathbb{H}(X_j)$ and $H(X_j)$. Using the second inequality in (4.4), we get

$$E\big(\|\mathbf{L}_n\|^2\big) \le \frac{1}{n}E\Big(\|\mathbf{b}_n'(H(X_1))\|^2\big(\mathbb{H}(X_1) - H(X_1)\big)^2\Big)$$
$$\le 2\pi^2\frac{m_n^3}{n}\sup_{t\in\mathcal{R}}\big(\mathbb{H}(X_1) - H(X_1)\big)^2$$
$$= O_p(m_n^3/n^2) = o_p((m_n n)^{-1})$$

as $m_n^4/n = o(1)$. This shows $\mathbf{L}_n = o_p((m_n n)^{-1/2})$. Using the third inequality in (4.4), one has as $m_n^6/n = o(1)$ that

$$\|\mathbf{M}_n\| \le \sqrt{2}\pi^2 m_n^{5/2}\sup_{t\in\mathcal{R}}|\mathbb{H}(t) - H(t)|^2 = O_p(m_n^{5/2}/n) = o_p((m_n n)^{-1/2}).$$

This yields $\mathbf{M}_n = o_p((m_n n)^{-1/2})$. Taken together one proves (6.5). This and (4.6) imply (6.4) as $m_n^4/n = o(1)$. Clearly, $\mathbf{U}_n = \mathbf{I}_{m_n}$ satisfies $|\mathbf{U}_n|_o = 1 = O(1)$. Peng and Schick (2005) showed that the projection of any $h \in L_2(Q)$ onto $\mathbb{A}$ uniquely exists under the assumption (4.3). Moreover, it is clear that $b_k(H(x)) - b_k(H(y)), k = 1, 2, \dots$ is a basis of $\mathbb{A}$, so that $[\mathbf{u}_\infty] = \mathbb{A}$. We now apply Theorem 6 to complete the proof. □

5. Simulations

We conducted a simulation study to compare the efficiency of the EL-weighted spatial median $\widetilde{\mathbf{m}}_n$ with the usual sample spatial median $\mathbf{m}_n$. Reported on Tables 1–5 are the maximum eigenvalues of the two variance-covariance matrices and their ratios. A ratio less than one indicates a reduction of the variance of the EL-weighted spatial median than that of the usual sample spatial median. We simulated data from 2-dimensional and 3-dimensional Cauchy distributions, t distribution with $df = 3$, the copula distribution and asymptotic Laplace distribution for sample sizes $n = 10, 50, 100, 200$ and repetitions 2000. Observe that the ratios simulated are substantially small than one, indicating substantial efficiency gains of the EL-weighted over the sample spatial depth. This also confirms the fact that the spatial median, as a location estimator, is much more efficient than the componentwise median.

TABLE 1

Cauchy and t distribution with known marginal medians $(0,0)$ and $(0,0,0)$. $\tilde{\lambda}$ and λ are the maximal eigenvalues of the variance-covariance matrices of the EL-weighted spatial median and the sample spatial median.

Cauchy						
	$dim = 2$			$dim = 3$		
n	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$
10	.00278	.01427	.19513	.00051	.01287	.03949
50	.00024	.00151	.16113	.00029	.00116	.24600
100	.00023	.00109	.21479	.00031	.00159	.19525
200	.00016	.00072	.21495	.00027	.00130	.20943
Student t (df=3)						
10	0.168	0.290	0.580	0.009	0.022	0.411
50	3.300	4.879	0.676	1.391	1.769	0.786
100	2.790	3.975	0.702	1.332	1.624	0.820
200	2.515	3.536	0.711	1.313	1.565	0.839
Copula distribution						
n	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$
10	0.0013	0.0054	0.2434	0.0007	0.0172	0.0417
50	0.0608	4.5110	0.0135	0.0018	0.5578	0.0033
100	0.0015	0.2297	0.0067	0.0361	2.6387	0.0137
200	0.0004	0.0045	0.0884	0.1601	11.6161	0.0138
Asymptotic Laplace						
10	0.0510	0.5213	0.0978	0.0057	0.0450	0.1273
50	0.1155	0.6779	0.1704	0.1119	0.5036	0.2221
100	0.0977	0.5474	0.1785	0.1043	0.4401	0.2371
200	0.0906	0.4940	0.1834	0.0995	0.4107	0.2423

6. Appendix

Theorem 6. *Suppose $\mathbf{u}_n = (u_1, \ldots, u_{m_n})^\top$ satisfies (C) for each $m = m_n$. Let $\hat{\mathbf{u}}_n$ be an estimator of $\mathbf{u}_n$ such that*

$$\max_{1 \le j \le n} \|\hat{\mathbf{u}}_n(Z_j)\| = o_p(m_n^{-3/2} n^{1/2}), \tag{6.1}$$

TABLE 2
Same as Table 1 except for simulating from 3-dimensional Cauchy.

Known F									
	$m = 1$			$m = 3$			$m = 5$		
n	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$
10	1.455	1.719	0.847	1.503	1.743	0.862	1.523	1.743	0.873
50	1.171	1.451	0.807	1.244	1.471	0.846	1.151	1.433	0.804
100	1.019	1.294	0.788	1.104	1.288	0.858	1.075	1.282	0.838
200	0.988	1.197	0.825	1.027	1.153	0.891	0.978	1.178	0.830
Unknown F									
10	1.499	1.765	0.849	1.460	1.760	0.829	1.416	1.792	0.790
50	1.161	1.373	0.846	1.227	1.411	0.870	1.260	1.406	0.896
100	1.076	1.296	0.830	1.047	1.294	0.809	1.064	1.314	0.810
200	1.008	1.209	0.834	0.977	1.215	0.804	0.972	1.200	0.810

TABLE 3
Same as Table 1 except for simulating from 3-dimensional t Distribution with df = 3

Known F									
	$m = 1$			$m = 3$			$m = 5$		
n	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$
10	1.420	1.712	0.830	1.337	1.700	0.786	1.303	1.695	0.769
50	1.222	1.416	0.863	1.132	1.472	0.769	1.082	1.420	0.762
100	1.142	1.468	0.778	1.157	1.398	0.828	1.113	1.429	0.779
200	1.085	1.396	0.777	1.058	1.317	0.803	1.017	1.254	0.811
Unknown F									
10	1.422	1.765	0.806	1.408	1.761	0.799	1.303	1.749	0.745
50	1.228	1.516	0.810	1.212	1.531	0.792	1.126	1.454	0.774
100	1.192	1.469	0.811	1.131	1.428	0.792	1.119	1.387	0.807
200	1.064	1.329	0.801	1.046	1.312	0.797	1.079	1.320	0.817

TABLE 4
Same as Table 1 except for simulating from 3-dimensional Copula Distribution

Known F									
	$m = 1$			$m = 3$			$m = 5$		
n	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$
10	.0005	0.0059	.0761	.0049	0.0719	.0684	.0099	0.1584	.0625
50	.3377	3.1555	.1070	.3422	3.1870	.1074	.3413	3.1200	.1094
100	.3382	2.9354	.1152	.3379	2.9423	.1148	.3414	2.9613	.1153
200	.3365	2.8107	.1197	.3379	2.8264	.1196	.3338	2.7950	.1194
Unknown F									
10	0.0107	0.1758	0.0610	0.0002	0.0055	0.0394	0.0026	0.0253	0.1035
50	0.3384	3.1378	0.1078	0.3409	3.1904	0.1068	0.3395	3.1714	0.1070
100	0.3401	2.9329	0.1160	0.3396	2.9585	0.1148	0.3388	2.9174	0.1161
200	0.3353	2.8192	0.1189	0.3372	2.8217	0.1195	0.3380	2.8395	0.1191

$$|\hat{\mathbf{W}}_n - \mathbf{W}_n|_o = o_p(m_n^{-1}) \tag{6.2}$$

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

$$\frac{1}{n}\sum_{j=1}^{n}\big(\boldsymbol{\psi}(Z_j)\otimes\hat{\mathbf{u}}_n(Z_j) - E\big(\boldsymbol{\psi}(Z_j)\otimes\hat{\mathbf{u}}_n(Z_j)\big)\big) = o_p(m_n^{-1/2}), \tag{6.3}$$

TABLE 5

Same as Table 1 except for simulating from 3-dimensional Asymmetric Laplace Distribution

Known F									
	$m=1$			$m=3$			$m=5$		
n	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$	$\tilde{\lambda}$	λ	$\tilde{\lambda}/\lambda$
10	.0010	.2135	.0046	.0001	.0011	.1010	.0001	.0020	.0622
50	.1795	5.996	.0299	.0009	.0137	.0643	.0001	.0004	.2244
100	.5964	16.02	.0372	.0019	.0618	.0304	.0001	.0004	.2864
200	.0018	.0877	.0200	.0002	.0010	.1740	.0025	.0675	.0372
Unknown F									
10	.00017	.00204	.08083	.00013	.00092	.10103	.00005	.00181	.02931
50	.00016	.00053	.30456	.00103	.03409	.03018	.00011	.00042	.25820
100	.00011	.00041	.25683	.00007	.00035	.18830	.00008	.00028	.28512
200	.00017	.00132	.12674	.00043	.00447	.09697	.00430	.30016	.01431

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* $\mathbf{U}_n = O(1)$*,*

$$\frac{1}{n}\sum_{j=1}^{n} E\left(\|\hat{\mathbf{u}}_n(Z_j) - \mathbf{v}_n(Z_j)\|^2\right) = o(m_n^{-1}), \quad \textit{and} \tag{6.4}$$

$$\frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{u}}_n(Z_j) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{v}_n(Z_j) + o_p(m_n^{-1/2} n^{-1/2}). \tag{6.5}$$

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

$$\hat{\boldsymbol{\theta}}_n = \bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\varphi}} + o_p(n^{-1/2}), \tag{6.6}$$

where $\boldsymbol{\varphi} = \Pi(\boldsymbol{\psi}|[\mathbf{v}_\infty])$ *is the projection of* $\boldsymbol{\psi}$ *onto the closed linear span* $[\mathbf{v}_\infty]$*. Thus*

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

where $\Sigma = \mathrm{Var}(\boldsymbol{\psi}(Z)) - \mathrm{Var}(\boldsymbol{\varphi}(Z))$*.*

References

[1] 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.

[2] 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.*

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

[4] HJORT, N.L., MCKEAGUE, I.W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**: 1079–1111.

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

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

[7] Parente, P. M. D. C. and Smith, R. J. (2011). GEL methods for nonsmooth moment indicators. *Econometric Theory* **27**: 74–113. MR2771012

[8] Peng, H. and Schick, A. (2018). Maximum empirical likelihood estimation and related topics. *Electron J. Statist.* **12**(2): 2962-2994.

[9] Peng, H. and Schick, A. (2013). An empirical likelihood approach of goodness of fit testing. *Bernoulli* **19**: 954–981

[10] Peng, H. and Tan, F. (2018). Jackknife empirical likelihood goodness-of-fit tests for U-statistics based general estimating equations. *Bernoulli* **24** (1): 449–464. DOI: 10.3150/16-BEJ884.

[11] 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.

[12] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**: 30–325.

[13] Wang, H. and Peng, H. (2022). Improving Efficiency in Structural Equation Models By An Easy Empirical Likelihood Approach. The manuscript can be found at `https://iupui.edu/~hpeng/preprints_hp.html`.

[14] Zhang, B. (1995). M-estimation and quantile estimation in the presence of auxiliary information. *J. Statist. Plann. Infer.* **44**, 77 – 94.

[15] Zhang, B. (1997). Quantile processes in the presence of auxiliary information. *Ann. Inst. Statist. Math.* **49**, 35 – 55.