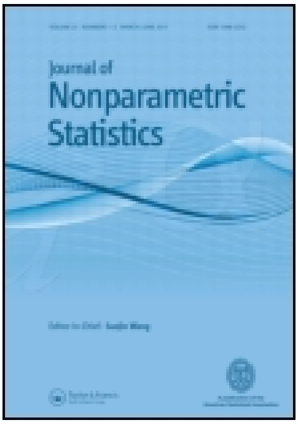

Journal of Nonparametric Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gnst20

Depth-based weighted empirical likelihood and general estimating equations

Yunlu Jiang [a], Shaoli Wang [c], Wenxiu Ge [a] & Xueqin Wang [a b]

[a] Department of Statistical Science, School of Mathematics & Computational Science, Sun Yat-Sen University, Guangzhou, 510275, People's Republic of China

[b] Bioinformatics Lab, Zhongshan Medical School, Sun Yat-Sen University, Guangzhou, 510080, People's Republic of China

[c] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433, People's Republic of China

Published online: 14 Jul 2011.

Journal of Nonparametric Statistics
Vol. 23, No. 4, December 2011, 1051–1062

Depth-based weighted empirical likelihood and general estimating equations

Yunlu Jiang[a], Shaoli Wang[c], Wenxiu Ge[a] and Xueqin Wang[a,b]*

[a]*Department of Statistical Science, School of Mathematics & Computational Science, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China;* [b]*Bioinformatics Lab, Zhongshan Medical School, Sun Yat-Sen University, Guangzhou 510080, People's Republic of China;* [c]*School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, People's Republic of China*

(*Received 7 January 2011; final version received 31 May 2011*)

In this paper, we link the depth-based weighted empirical likelihood (WEL) with general estimating equations to produce a robust estimation of parameters for contaminated data with auxiliary information about the parameters. Such auxiliary information can be expressed through a group of functionally independent general estimating equations. Under general conditions, asymptotic properties of the WEL estimator are established. Furthermore, we prove that the WEL ratio statistic is asymptotically chi-squared distributed. Simulation studies are conducted to test the robustness of the WEL estimator. Finally, we apply the proposed method to analyse the gilgai survey data.

Keywords: weighted empirical likelihood; general estimating equations; depth function; robustness

AMS Subject Classifications: 62G05; 62G20; 62G35; 62H12

1. Introduction

The method of empirical likelihood (EL) was introduced by Owen (1988, 1990, 1991) to construct confidence regions in nonparametric settings. It permits likelihood inference without parametric assumptions on data distribution and without the estimations of the asymptotic covariance. The EL method yields confidence regions that are Bartlett correctable and, therefore, can produce more efficient estimation (DiCiccio, Hall, and Romano 1991; Chen and Hall 1993). This method was applied to contaminated data by Owen (2001) and Zhao (2007).

The auxiliary information expressed through equality constraints can be incorporated into the EL method (Qin and Lawless 1994), an approach which enables applications of the EL method in more general situations, for example, semi-parametric likelihood-based inference for truncated data in survey sample (Li and Qin 1998, Qin 2000), missing response data (Wang and Rao 2002; Wang and Dai 2008), and the element-wise and subject-wise ELs for longitudinal data (Wang, Qian, and Carroll 2010).

*Corresponding author. Email: wangxq88@mail.sysu.edu.cn

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485252.2011.594510
http://www.informaworld.com

The idea of incorporating the auxiliary information into the EL method may be closely linked to the method of general estimating equations. Like the general estimating equations method, the EL method with equality constraints is sensitive to outliers. In this paper, we construct depth-based weighted empirical likelihood (WEL) with finitely many equality constraints or estimating equations. Particularly, our approach is applicable to the situation when the number of estimating equations is larger than the dimension of the parameter space. A similar idea has been employed by Choi, Hall, and Presnell (2000) in a parametric problem and by Glenn and Zhao (2007) in a nonparametric setting without auxiliary information about the parameters, but our methods for weight construction are different. We construct weights based on a data depth function for multivariate samples. Data depth provides a centre-outward ordering of multi-dimensional data. Points deep inside the data are assigned with a high depth and those on the outskirts with a lower depth. A depth function puts smaller weights on outliers and results in a depth WEL estimator that is more robust than the usual one for contaminated data, as indicated by our simulation results. Several data depth functions have been proposed and extensively studied in the literature, for example, Mahalanobis depth (Rao 1988), simplicial depth (Liu 1990), half-space depth (Zuo and Serfling 2000c), spatial depth (Serfling 2002) and projection depth (Zuo 2003).

The remainder of this paper is organised as follows. In Section 2, we review the EL approach linked with general estimating equations and then introduce the WEL estimator and study its asymptotic properties. The asymptotic distribution of WEL ratio statistics for parameters is established. In Section 3, we describe a method of choosing the weights. In Section 4, simulation studies are conducted to demonstrate the performance of the proposed methodology, and a real data analysis is illustrated. Concluding remarks are presented in Section 5. The proofs are given in the appendix.

2. Definition

2.1. *EL with general estimating equations*

Suppose that $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ are i.i.d. d-dimensional random vectors following an unknown distribution F with a p-dimensional parameter $\boldsymbol{\theta}$. We assume that the information about $\boldsymbol{\theta}$ and F is available through $r \geq p$ functionally independent unbiased estimating functions, denoted by $\boldsymbol{g}(\boldsymbol{X},\boldsymbol{\theta}) = (g_1(\boldsymbol{X},\boldsymbol{\theta}),\cdots,g_r(\boldsymbol{X},\boldsymbol{\theta}))^{\mathrm{T}}$, such that $E_F\{\boldsymbol{g}(\boldsymbol{X},\boldsymbol{\theta})\} = \mathbf{0}$. Under some regularity conditions, Qin and Lawless (1994) showed that the EL under the aforementioned constraints was equivalent to the profile EL function for $\boldsymbol{\theta}$:

$$K(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{\left(\frac{1}{n}\right)\frac{1}{1+\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta})}\right\},$$

where $\boldsymbol{t}(\boldsymbol{\theta}) = (t_1(\boldsymbol{\theta}), t_2(\boldsymbol{\theta}),\ldots,t_r(\boldsymbol{\theta}))^{\mathrm{T}}$ are the Lagrange multipliers that satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(X_i,\boldsymbol{\theta})}\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta}) = \mathbf{0}.$$

Since $\prod_{i=1}^{n} p_i$ is maximised at $p_i = n^{-1}$ in the absence of the parametric constraints, the empirical log-likelihood ratio is

$$k(\boldsymbol{\theta}) = \sum_{i=1}^{n}\log\{1+\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta})\}.$$

The empirical log-likelihood ratio $k(\boldsymbol{\theta})$ can then be minimised to obtain an estimate $\hat{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$, called the EL estimator. They also derived the asymptotic distribution of the estimator $\hat{\boldsymbol{\theta}}$ under mild conditions.

2.2. WEL with general estimating equations

The EL method works well with i.i.d. samples. For contaminated data, Choi et al. (2000) assigned weights to each observation to tilt the parametric likelihood in a parametric setting. This approach usually yields a more robust estimation. Glenn and Zhao (2007) extended the WEL approach to a nonparametric setting. We will explore the same idea for EL combined with additional information about parameters expressed through finitely many functionally independent unbiased estimating functions.

Definition 2.1 *Suppose that $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are i.i.d. d-dimensional random vectors with an unknown distribution function F and a p-dimensional parameter $\boldsymbol{\theta}$ associated with F. Assume that p_i is the probability mass placed on $\boldsymbol{X}_i$. Given a weight vector $\boldsymbol{\omega}_n$, $\sum_{i=1}^{n} \omega_{ni} = 1, \omega_{ni} \geq 0$, the WEL for parameter $\boldsymbol{\theta}$ is then defined as*

$$\mathrm{WEL}(\theta) = \sup \left\{ \prod_{i=1}^{n} p_i^{n\omega_{ni}} : (p_1, \dots, p_n) \in \wp_n, \sum_{i=1}^{n} p_i \boldsymbol{g}(X_i, \boldsymbol{\theta}) = \boldsymbol{0} \right\}, \tag{1}$$

where $\wp_n = \{(p_1, \dots, p_n) \in [0,1]^n, p_1 + \cdots + p_n = 1\}$.

Remark 2.1 When the weights $\omega_{ni} = 1/n, i = 1, 2, \dots, n$, the WEL (1) is reduced to the classical EL. In order to achieve the robustness with EL, the weight ω_{ni} for an outlier $\boldsymbol{X}_i$ should be small.

We next derive the limiting distributions of the WEL estimator. Throughout the remainder of this paper, let $\boldsymbol{\theta}_0$ be the true value of the parameter $\boldsymbol{\theta}$. We maximise the EL $\prod_{i=1}^{n} p_i^{n\omega_{ni}}$ subject to restrictions

$$(p_1, \dots, p_n) \in \wp_n, \quad \sum_{i=1}^{n} p_i \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) = \boldsymbol{0}.$$

For any given $\boldsymbol{\theta}$, suppose that $\boldsymbol{0}$ is inside the convex hull of points $\boldsymbol{g}(\boldsymbol{X}_1, \boldsymbol{\theta}), \dots, \boldsymbol{g}(\boldsymbol{X}_n, \boldsymbol{\theta})$, then a unique maximum exists. The maximum can be found via Lagrange multipliers. Let

$$H = \sum_{i=1}^{n} n\omega_{ni} \log(p_i) + \lambda \left(1 - \sum_{i=1}^{n} p_i\right) - n\boldsymbol{t}^{\mathrm{T}} \sum_{i=1}^{n} p_i \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}),$$

where λ and $\boldsymbol{t} = (t_1, \dots, t_r)^{\mathrm{T}}$ are the Lagrange multipliers. Setting to zero the partial derivatives of H with respect to p_i gives

$$\frac{\partial H}{\partial p_i} = \frac{n\omega_{ni}}{p_i} - \lambda - n\boldsymbol{t}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) = 0.$$

It follows that

$$0 = \sum_{i=1}^{n} p_i \frac{\partial H}{\partial p_i} = n - \lambda,$$

which implies $\lambda = n$. We, therefore, obtain

$$p_i = \frac{\omega_{ni}}{1 + \boldsymbol{t}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})};$$

using this formula for p_i together with the second restriction in Equation (1), we have

$$\boldsymbol{0} = \sum_{i=1}^{n} p_i \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\omega_{ni} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})}{1 + \boldsymbol{t}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})},$$

from which $\boldsymbol{t}$ can be determined in terms of $\boldsymbol{\theta}$.

The WEL function for $\boldsymbol{\theta}$ is then rewritten as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \frac{\omega_{ni}}{1 + \boldsymbol{t}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})} \right\}^{n\omega_{ni}}.$$

Since $\prod_{i=1}^{n} p_i^{n\omega_{ni}}$ is maximised at $p_i = \omega_{ni}$ in the absence of parametric constraints, the empirical log-likelihood ratio is

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} n\omega_{ni} \log\{1 + \boldsymbol{t}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\}. \tag{2}$$

We can minimise Equation (2) to obtain an estimate $\tilde{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$ and call it the WEL estimator. It can be shown that $\tilde{\boldsymbol{\theta}}$ is asymptotically normal. Before presenting theorems on asymptotic properties, we list some regularity conditions:

C1: $Pr\{\mathbf{0} \in \mathrm{ch}(\boldsymbol{g}(\boldsymbol{X}_1, \boldsymbol{\theta}), \ldots, \boldsymbol{g}(\boldsymbol{X}_n, \boldsymbol{\theta}))\} \to 1$, where $\mathrm{ch}(A)$ denotes the convex hull of a set A.
C2: The partial derivatives $\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ are continuous in a neighbourhood of the true value $\boldsymbol{\theta}_0$. Furthermore, $\|\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta})/\partial \boldsymbol{\theta}\|$ and $\|\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta})\|^3$ are bounded by some integrable function $G(X)$ in this neighbourhood.
C3: $E\{\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)^{\mathrm{T}}\}$ is positive definite.
C4: The rank of $E[\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}]$ is p.
C5: The second-order partial derivatives $\partial^2 \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}$ are continuous in $\boldsymbol{\theta}$ and $\|\partial^2 \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}\|$ is bounded by some integrable function $H(X)$ in a neighbourhood of the true value $\boldsymbol{\theta}_0$.
C6: The weight vector $\boldsymbol{\omega}_n = (\omega_{n1}, \omega_{n2}, \ldots, \omega_{nn})^{\mathrm{T}}$, $\omega_{ni} \geq 0$, $\sum_{i=1}^{n} \omega_{ni} = 1$, satisfies that $\lim_{n\to\infty} \sum_{i=1}^{n} n\omega_{ni}^2 = c$, where c is a finite positive real number.

THEOREM 2.1 *Assume that conditions* (C1)–(C6) *hold. Then, with probability* 1 *as* $n \to \infty$, $l(\boldsymbol{\theta})$ *attains its minimum value at a point* $\tilde{\boldsymbol{\theta}}$, *which is in the interior of the ball of* $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq n^{-1/3}$, *and* $\tilde{\boldsymbol{\theta}}$ *and* $\tilde{\boldsymbol{t}} = \boldsymbol{t}(\tilde{\boldsymbol{\theta}})$ *satisfy*

$$\boldsymbol{Q}_{1n}(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{t}}) = \mathbf{0}, \quad \boldsymbol{Q}_{2n}(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{t}}) = \mathbf{0}, \tag{3}$$

where

$$\boldsymbol{Q}_{1n}(\boldsymbol{\theta}, \boldsymbol{t}) = \sum_{i=1}^{n} \frac{\omega_{ni} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})}{1 + \boldsymbol{t}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})},$$

$$\boldsymbol{Q}_{2n}(\boldsymbol{\theta}, \boldsymbol{t}) = \sum_{i=1}^{n} \frac{\omega_{ni}}{1 + \boldsymbol{t}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})} \left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{t}.$$

THEOREM 2.2 *Under conditions* $(C1) - (C6)$, *we have*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, cV), \tag{4}$$

where $V = \{E(\partial \boldsymbol{g}/\partial \boldsymbol{\theta})^{\mathrm{T}} (E\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}})^{-1} E(\partial \boldsymbol{g}/\partial \boldsymbol{\theta})\}^{-1}$, *and c is given in condition* (C6).

The method of WEL not only yields a robust estimation of the parameters under general estimating equation constraints but also enables hypothesis testing via the WEL ratio statistic, as the next theorem states.

THEOREM 2.3 *The WEL ratio statistic for testing* $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *is*

$$W(\boldsymbol{\theta}_0) = 2l(\boldsymbol{\theta}_0) - 2l(\tilde{\boldsymbol{\theta}}),$$

where $l(\boldsymbol{\theta})$ *is given by Equation* (2). *Assume that conditions* (C1)–(C6) *hold. Under* $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, *we have* $W(\boldsymbol{\theta}_0) \xrightarrow{D} \chi^2_{(p)}$ *as* $n \to \infty$, *where p is given in condition* (C4).

3. Depth-based weights

Use of a proper weight vector $\boldsymbol{\omega}_n$ to reduce the influence of outliers is essential to the robustness of a WEL estimator. A method of measuring the extremeness of an observation in a sample is to examine its relative location with respect to the rest of the population. An observation is more likely to be an outlier if it is far away from the centre of the distribution. Depth function provides such a measurement. Points deep inside the data are assigned with a high depth and those on the outskirts with a lower depth. The spatial depth has a less computational complexity ($O(n^2)$, independent of the dimension of the data, Serfling 2002) among competitors and thus is preferred. Therefore, we use the spatial depth function to generate weights $\boldsymbol{\omega}_n$ in this paper.

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent copies of a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ with a common distribution function F, and let F_n be the corresponding empirical distribution, which assigns mass $1/n$ to each sample point $\boldsymbol{X}_i$, $i = 1, 2, \ldots, n$. The spatial depth function or L^2 depth function is defined as

$$D(\boldsymbol{x}; F) = \frac{1}{1 + E||\boldsymbol{x} - \boldsymbol{X}||}, \tag{5}$$

where $||\cdot||$ is the Euclidean norm. The weights based on this depth function are

$$\omega_{ni} = \frac{D(\boldsymbol{X}_i; F_n)}{\sum_{i=1}^{n} D(\boldsymbol{X}_i; F_n)}. \tag{6}$$

Theorem 3.1 *For a weight vector given by Equation* (6), *if* $\int D(\boldsymbol{x}; F)\,\mathrm{d}F(\boldsymbol{x}) > 0$ *and* $0 \le D(\boldsymbol{x}; F) \le 1$, *then*

$$\omega_{ni} \ge 0, \quad \sum_{i=1}^{n} \omega_{ni} = 1,$$

and

$$\sum_{i=1}^{n} n\omega_{ni}^2 \xrightarrow{\text{a.s.}} \frac{\int D^2(\boldsymbol{x}; F)\mathrm{d}F(\boldsymbol{x})}{\{\int D(\boldsymbol{x}; F)\,\mathrm{d}F(\boldsymbol{x})\}^2} \text{ as } n \to \infty.$$

Remark 3.1 From Theorem 3.1, we observe that condition (C6) holds for weight vector given in Equation (6), and $c = \int D(\boldsymbol{x}; F)^2\,\mathrm{d}F(\boldsymbol{x})/\{\int D(\boldsymbol{x}; F)\,\mathrm{d}F(\boldsymbol{x})\}^2$. From Equation (4), the asymptotic variance of $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ can be consistently estimated by

$$\tilde{V} = \left[\frac{(1/n)\sum_{i=1}^{n} D^2(\boldsymbol{X}_i; F_n)}{\{(1/n)\sum_{i=1}^{n} D(\boldsymbol{X}_i; F_n)\}^2}\right] \times \left[\left\{\sum_{i=1}^{n} \tilde{p}_i \frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right\}^{\mathrm{T}} \left\{\sum_{i=1}^{n} \tilde{p}_i \boldsymbol{g}(\boldsymbol{X}_i, \tilde{\boldsymbol{\theta}}) \boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}_i, \tilde{\boldsymbol{\theta}})\right\}^{-1} \left\{\sum_{i=1}^{n} \tilde{p}_i \frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right\}\right]^{-1},$$

where $\tilde{\boldsymbol{\theta}}$ is the WEL estimator, and $\tilde{p}_i$ is an estimator of p_i, namely,

$$\tilde{p}_i = \frac{\omega_{ni}}{1 + \tilde{\boldsymbol{t}}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{X}_i, \tilde{\boldsymbol{\theta}})},$$

$\tilde{\boldsymbol{t}}$ is given by Equation (3), and $D(\cdot; \cdot)$ is defined in Equation (5). If $\boldsymbol{X}_i$ is an outlier, $\tilde{p}_i$ should be small. Therefore, $\tilde{V}$ is robust against outliers.

4. Simulation and application

In this section, we conduct simulations to examine the robustness of the WEL estimator, and then an illustration is given via a real data analysis.

Example 4.1 We consider a two-sample problem with a common mean. Such problems are common in survey sampling (e.g. Kuk and Mak 1989; Chen and Qin 1993). Suppose that we have i.i.d. observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ and $E(X_i) = \theta$ and $E(Y_i) = \theta$. To estimate θ, we use estimating equations $E\boldsymbol{g} = E(g_1, g_2) = \mathbf{0}$, where $g_1 = X - \theta$ and $g_2 = Y - \theta$, respectively.

In our simulation study, we take three different contamination levels: 0%, 4% and 8%. Suppose that there are two distributions F_1 and F_2:

$$F_1 \sim N\left(\begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right), \quad F_2 \sim N\left(\begin{pmatrix}10\\10\end{pmatrix}, \begin{pmatrix}6 & 2\\2 & 6\end{pmatrix}\right).$$

We use F_2 as the contaminating distribution. For each contamination level, we generate 1000 samples of sizes $n = 50, 100, 150$ and 200, respectively, from the mixture of F_1 and F_2 with corresponding contamination probabilities (levels). For the WEL estimator, the weights are calculated according to Section 3. We compute the WEL estimator and the usual EL estimator and then calculate their mean squared errors (MSEs). We take the nominal level $1 - \alpha = 0.95$ and compute their average lengths (ALs) of confidence intervals and their corresponding empirical coverage probabilities (CPs). The results are presented in Table 1.

Table 1. WELE, ELE, MSEs, ALs and CPs for three different contamination levels.

Contamination levels	Method		$n = 50$	$n = 100$	$n = 150$	$n = 200$
0%	WEL	WELE	2.0022	1.9996	1.9973	1.9999
		(a)	0.0096	0.0053	0.0034	0.0024
		(b)	0.3596	0.2574	0.2111	0.1836
		(c)	93.3	91.8	92.6	94.0
	EL	ELE	2.0018	1.9996	1.9973	1.9997
		(a)	0.0095	0.0052	0.0034	0.0023
		(b)	0.3804	0.2731	0.2242	0.1951
		(c)	94.3	93.7	95.2	95.7
4%	WEL	WELE	2.1010	2.0912	2.0932	2.0915
		(a)	0.0224	0.0140	0.0123	0.0110
		(b)	0.5962	0.4181	0.3423	0.2970
		(c)	96.8	94.4	90.3	87.5
	EL	ELE	2.3246	2.3125	2.3214	2.3205
		(a)	0.1213	0.1050	0.1081	0.1062
		(b)	0.9383	0.6678	0.5556	0.4845
		(c)	90.1	62.0	23.9	6.2
8%	WEL	WELE	2.2191	2.2165	2.2143	2.2145
		(a)	0.0601	0.0529	0.0497	0.0489
		(b)	0.8002	0.5685	0.4639	0.4021
		(c)	95.4	80.5	62.2	40.2
	EL	ELE	2.6284	2.6392	2.6399	2.6422
		(a)	0.4151	0.4182	0.4155	0.4168
		(b)	1.2427	0.9024	0.7413	0.6448
		(c)	46.9	1.2	0	0

Notes: (a), MSEs; (b), ALs of 95% confidence intervals; (c), empirical CPs (%); WELE and ELE denote the WEL estimator and the EL estimator, respectively.

From Table 1, we can see that the WEL estimator is robust against outliers, though it does not perform well with a highly contaminated sample. Meanwhile, in the case of no outliers, we observe that MSEs of the WEL estimator and EL estimator are almost the same. Although CPs of the WEL estimator are somewhat lower than those of the EL estimator, ALs obtained by the WEL are much shorter than those obtained by the usual EL. In the cases of contamination, the MSEs and ALs of the WEL estimator are much smaller than those of the EL estimator. As the contamination level of a sample increases, the MSEs and ALs of both estimators increase, but the magnitude of increase is much smaller for the WEL estimator. The CPs of the EL method are seriously influenced and are much smaller than the nominal level of 95%. In contrast, the CPs of the WEL method are closer to the given nominal level.

Example 4.2 In this example, we take the same setting of Example 4.1, except that $n = 100$, and contamination levels $\gamma \in [0, 0.3]$. For each $\gamma \in [0, 0.3]$, we compute MSEs of the WEL estimator and EL estimator and then plot the MSEs against the contamination levels. The MSE curves for both estimators are depicted in Figure 1. From Figure 1, we can observe that the WEL estimator is more robust against outliers than the EL estimator.

Example 4.3 As an illustration, we apply the proposed method to analyse the gilgai survey data, which were collected on a line transect survey in gilgai territory in New South Wales, Australia. This data set consists of 365 samples, which were taken at depths 0–10, 30–40 and 80–90 cm below the surface. pH, electrical conductivity and chloride content are measured on a 1:5 soil:water extract from each sample. This data set contains the following nine columns: pH at depth 0–10 cm (pH00), pH at depth 30–40 cm (pH30), pH at depth 80–90 cm (pH80), electrical conductivity in mS/cm (0–10 cm) (e00), electrical conductivity in mS/cm (30–40 cm) (e30), electrical conductivity in mS/cm (80–90 cm) (e80), chloride content in ppm (0–10 cm) (c00), chloride content in ppm (30–40 cm) (c30) and chloride content in ppm (80–90 cm) (c80). A histogram of c00 and c30 in Figure 2 indicates that there are unusual points in the data. We use the proposed WEL to estimate the mean of c00 and c30 and then compare it with the EL estimator. We

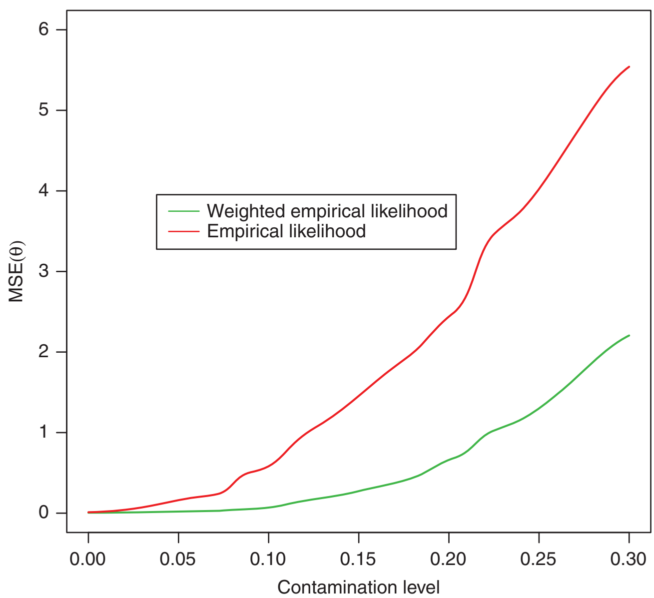

Figure 1. The MSE curves of the contamination level.

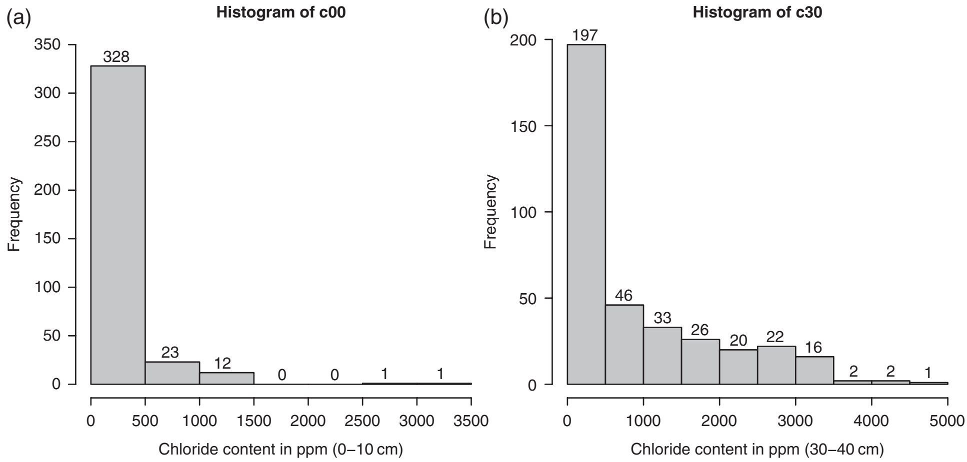

Figure 2. Histogram of c00 (a) and c30 (b).

Table 2. Analysis result for the gilgai survey data.

Variable	WEL estimator	WELCI	EL estimator	ELCI
c00	79.68	[63.39, 95.98]	171.74	[135.41, 208.07]
c30	627.12	[540.29, 713.96]	894.34	[786.26, 1002.40]

Notes: WELCI denotes the 95% confidence intervals obtained by the WEL method. ELCI denotes the 95% confidence intervals obtained by the EL method.

calculate the weights according to Section 3. We take the nominal level $1 - \alpha = 0.95$ and obtain the confidence intervals based on the WEL method and EL method, respectively. The analysis results are shown in Table 2. From Table 2, we can see that the WEL method is robust against outliers.

5. Concluding remarks

In this paper, we extended the method proposed by Glenn and Zhao (2007) to depth-based WEL which links EL and general estimating equations together. This approach enabled us to incorporate auxiliary information about the parameters and yielded robust estimation. Spatial depth function was exploited for an appropriate choice of weights. Simulation studies indicated that the newly proposed WEL estimator was not sensitive to outliers, though it did not perform well with highly contaminated data. Further research on the weight choice is needed.

Acknowledgements

Xueqin Wang was partially supported by NSFC under Grant No. 11001280, Doctoral Fund of Ministry of Education of China under Grant No. 20090171110017, Natural Science Foundation of Guangdong Province under Grant No. 10151027501000066 and SRF for ROCS, SEM. Shaoli Wang's work was sponsored by Shanghai Pujiang Program, and was also partially supported by Shanghai University of Finance and Economics through Project 211 Phase III and Shanghai Leading Academic Discipline Project, Project Number: B803.

References

Chen, S.X., and Hall, P. (1993), 'Smoothed Empirical Likelihood Confidence Intervals for Quantiles', *Annals of Statistics*, 21, 1166–1181.

Chen, J., and Qin, J. (1993), 'Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information', *Biometrika*, 80, 107–116.
Choi, E., Hall, P., and Presnell, B. (2000), 'Rendering Parametric Procedures More Robust by Empirically Tilting the Model', *Biometrika*, 87, 453–465.
DiCiccio, T.J., Hall, P., and Romano, J.P. (1991), 'Empirical Likelihood is Bartlett-Correctable', *Annals of Statistics*, 19, 1053–1061.
Glenn, N.L., and Zhao, Y. (2007), 'Weighted Empirical Likelihood Estimates and Their Robustness Properties', *Computational Statistics & Data Analysis*, 51, 5130–5141.
Kuk, A.Y.C., and Mak, T.K. (1989), 'Median Estimation in the Presence of Auxiliary Information', *Journal of the Royal Statistical Society Series B*, 51, 261–269.
Li, G., and Qin, J. (1998), 'Semiparametric Likelihood-Based Inference for Biased and Truncated Data when the Total Sample Size is Known', *Journal of the Royal Statistical Society Series B*, 60, 243–254.
Liu, R.Y. (1990), 'On a Notion of Data Depth Based Upon Random Simplices', *Annals of Statistics*, 18, 405–414.
Owen, A.B. (1988), 'Empirical Likelihood Ratio Confidence Intervals for a Single Functional', *Biometrika*, 75, 237–249.
Owen, A.B. (1990), 'Empirical Likelihood Ratio Confidence Regions', *Annals of Statistics*, 18, 90–120.
Owen, A.B. (1991), 'Empirical Likelihood for Linear Models', *Annals of Statistics*, 19, 1725–1747.
Owen, A.B. (2001), *Empirical Likelihood*, New York: Chapman and Hall/CRC.
Peligrad, M., and Utev, S. (1997), 'Central Limits Theorem for Linear Process', *Annals of Applied Probability*, 25, 443–456.
Pruitt, W.E. (1966), 'Summability of Independent Random Variables', *Journal of Mathematics and Mechanics*, 15, 769–776.
Qin, J., and Lawless, J.F. (1994), 'Empirical Likelihood and General Estimating Equations', *Annals of Statistics*, 22, 300–325.
Qin, J. (2000), 'Combining Parametric and Empirical Likelihoods', *Biometrika*, 87, 484–490.
Rao, C.R. (1988), 'Methodology Based on the Norm in Statistical Inference', *Sankhya Series A*, 50, 289–313.
Serfling, R. (2002), 'A Depth Function and a Scale Curve Based on Spatial Quantiles', in *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, ed. Y. Dodge, Basel: Birkhaüser, pp. 25–38. http://www.amazon.com/Statistical-Analysis-Statistics-Industry-Technology/dp/3764369205.
Wang, Q.H., and Dai, P. (2008), Semiparametric Model-Based Inference in the Presence of Missing Responses, *Biometrika*, 95, 721–734.
Wang, S., Qian, L., and Carroll, R.J. (2010), 'Generalized Empirical Likelihood Methods for Analyzing Longitudinal Data', *Biometrika*, 97, 79–93.
Wang, Q.H., and Rao, J.N.K. (2002), 'Empirical Likelihood-Based Inference Under Imputation for Missing Response Data', *Annals of Statistics*, 30, 896–924.
Zuo, Y. (2003), 'Projection-Based Depth Functions and Associated Medians', *Annals of Statistics*, 31, 1460–1490.
Zuo, Y., and Serfling, R. (2000c), 'On the Performance of Some Robust Nonparametric Location Measures Relative to a General Notion of Multivariate Symmetry', *Journal of Statistical Planning and Inference*, 84, 55–79.

Appendix

LEMMA A.1 *Assume that $X_1, \ldots, X_n$ are independent random vectors from a common distribution with mean μ and variance $\sigma^2 < +\infty$. Let $\omega_n = (\omega_{n1}, \omega_{n2}, \ldots, \omega_{nn})^{\mathrm{T}}$ be a sequence of weight vectors satisfying $\omega_{ni} \geq 0$ and $\sum_{i=1}^n \omega_{ni} = 1$. If $\max_{1\leq i\leq n} \omega_{ni}^2 \to 0$ as $n \to \infty$, then when $n \to \infty$, we have*

$$\frac{\sum_{i=1}^n \omega_{ni}(X_i - \mu)}{\sqrt{\sum_{i=1}^n \omega_{ni}^2 \sigma^2}} \xrightarrow{D} N(0, 1).$$

The proof of Lemma A.1 *can be found in Peligrad and Utev* (*1997*).

LEMMA A.2 *Assume that $X_1, \ldots, X_n$ are independent random variables from a common distribution with a finite mean μ. Let $Y_n = \sum_{i=1}^n a_{ni} X_i$, where $\{a_{ni}\}$ is a sequence of real numbers. If, in addition, the following conditions hold*:

(1) *For every $i \in N$, we have* $\lim_{n\to\infty} a_{ni} = 0$.
(2) $\lim_{n\to\infty} \sum_{i=1}^n a_{ni} = 1$.
(3) *There exists a constant $M > 0$ such that for every n, $\sum_{i=1}^\infty |a_{ni}| \leq M$,*

then

$$Y_n \xrightarrow{P} \mu \Leftrightarrow \max_{1\leq i\leq n} |a_{ni}| \to 0, \quad \text{as } n \to \infty.$$

The proof of Lemma A.2 *can be found in Pruitt* (*1966*).

Proof of Theorem 2.1 Let $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{u}n^{-1/3}$, where $\|\boldsymbol{u}\| = 1$. Since

$$0 = \sum_{i=1}^{n} \frac{\omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \theta)}{1 + \boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})} = \sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) \left\{1 - \boldsymbol{t}^{\mathrm{T}}(\theta)\boldsymbol{g}(X_i, \boldsymbol{\theta}) + \frac{[\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(X_i, \boldsymbol{\theta})]^2}{1 + \boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(X_i, \boldsymbol{\theta})}\right\},$$

$$\left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}_i, \boldsymbol{\theta})\right\} \boldsymbol{t}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) + \sum_{i=1}^{n} \frac{\omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})[\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})]^2}{1 + \boldsymbol{t}^{\mathrm{T}}(\theta)\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})}.$$

Similar to the proof of Theorem 1 in Owen (1990), when $E[\|\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta})\|^3] < +\infty$, $0 \le \omega_{ni} \le 1$, and $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le n^{-1/3}$, then for $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le n^{-1/3}\}$, we uniformly have

$$t(\theta) = \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}_i, \boldsymbol{\theta})\right\}^{-1} \sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) + o(n^{-1/3}) = O(n^{-1/3}), \text{ a.s.} \tag{A1}$$

By Equation (A1), Lemmas A.1 and A.2 and Taylor's expansion, we have

$$\begin{aligned}
l(\boldsymbol{\theta}) &= n \sum_{i=1}^{n} \omega_{ni}\boldsymbol{t}^{\mathrm{T}}(\theta)\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) - \frac{n}{2} \sum_{i=1}^{n} \omega_{ni}\{\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\}^2 + o(n^{1/3}) \text{ (a.s.)} \\
&= \frac{n}{2} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\right\}^{\mathrm{T}} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}_i, \boldsymbol{\theta})\right\}^{-1} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\right\} + o(n^{1/3}) \text{ (a.s.)} \\
&= \frac{n}{2} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}_0) + \sum_{i=1}^{n} \omega_{ni} \frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{u}n^{-1/3}\right\}^{\mathrm{T}} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(X_i, \boldsymbol{\theta})\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}_i, \boldsymbol{\theta})\right\}^{-1} \\
&\quad \times \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}_0) + \sum_{i=1}^{n} \omega_{ni} \frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{u}n^{-1/3}\right\} + o(n^{1/3}) \text{ (a.s.)} \\
&= \frac{n}{2} \left\{O(n^{-1/2}(\log\log n)^{1/2}) + E\left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) \boldsymbol{u}n^{-1/3}\right\}^{\mathrm{T}} [E\{\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}, \boldsymbol{\theta}_0)\}]^{-1} \\
&\quad \times \left\{O(n^{-1/2}(\log\log n)^{1/2}) + E\left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) \boldsymbol{u}n^{-1/3}\right\} + o(n^{1/3}) \text{ (a.s.)} \\
&\ge (\lambda - \epsilon)n^{1/3} \text{ (a.s.)},
\end{aligned}$$

where $\lambda - \epsilon > 0$ and λ is the smallest eigenvalue of

$$\left\{E\left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)\right\}^{\mathrm{T}} [E\{\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}, \boldsymbol{\theta}_0)\}]^{-1} \left\{E\left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)\right\}.$$

Similarly,

$$\begin{aligned}
l(\boldsymbol{\theta}) &= \frac{n}{2} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}_0)\right\}^{\mathrm{T}} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}_0)\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}_i, \boldsymbol{\theta}_0)\right\}^{-1} \left\{\sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}_0)\right\} + o(1) \text{ (a.s.)}, \\
&= O(\log\log n) \text{ (a.s.)}.
\end{aligned}$$

Since $l(\boldsymbol{\theta})$ is a continuous function in $\boldsymbol{\theta}$ as $\boldsymbol{\theta}$ belongs to the ball $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le n^{-1/3}$, $l(\boldsymbol{\theta})$ has the minimum value in the interior of this ball, and $\tilde{\boldsymbol{\theta}}$ satisfies $(\partial l(\boldsymbol{\theta})/\partial \boldsymbol{\theta})|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = 0$. This completes the proof of Theorem 2.1. ■

Proof of Theorem 2.2 Taking derivatives with respect to $\boldsymbol{\theta}$ and $\boldsymbol{t}^{\mathrm{T}}$, we have

$$\frac{\partial \boldsymbol{Q}_{1n}(\boldsymbol{\theta}, \boldsymbol{0})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \omega_{ni} \frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad \frac{\partial \boldsymbol{Q}_{1n}(\boldsymbol{\theta}, \boldsymbol{0})}{\partial \boldsymbol{t}^{\mathrm{T}}} = \sum_{i=1}^{n} \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})^{\mathrm{T}},$$

$$\frac{\partial Q_{2n}(\boldsymbol{\theta}, \boldsymbol{0})}{\partial \boldsymbol{\theta}} = \boldsymbol{0}, \quad \frac{\partial \boldsymbol{Q}_{2n}(\boldsymbol{\theta}, \boldsymbol{0})}{\partial \boldsymbol{t}^{\mathrm{T}}} = \sum_{i=1}^{n} \omega_{ni}\left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}}.$$

Expanding $\boldsymbol{Q}_{1n}(\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{t}})$ and $\boldsymbol{Q}_{2n}(\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{t}})$ at $(\boldsymbol{\theta}_0,\mathbf{0})$, and using the conditions in Theorem 2.2, we obtain

$$\mathbf{0} = \boldsymbol{Q}_{1n}(\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{t}}) = \boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + \frac{\partial \boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0})}{\partial \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + \frac{\partial \boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0})}{\partial \boldsymbol{t}^{\mathrm{T}}}(\tilde{\boldsymbol{t}}-\mathbf{0}) + o_p(\delta_n),$$

$$\mathbf{0} = \boldsymbol{Q}_{2n}(\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{t}}) = \boldsymbol{Q}_{2n}(\boldsymbol{\theta}_0,\mathbf{0}) + \frac{\partial Q_{2n}(\boldsymbol{\theta}_0,\mathbf{0})}{\partial \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + \frac{\partial Q_{2n}(\boldsymbol{\theta}_0,\mathbf{0})}{\partial \boldsymbol{t}^{\mathrm{T}}}(\tilde{\boldsymbol{t}}-\mathbf{0}) + o_p(\delta_n),$$

where $\delta_n = ||\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0|| + ||\tilde{\boldsymbol{t}}||$. It follows that

$$\begin{pmatrix} \tilde{\boldsymbol{t}} \\ \tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0 \end{pmatrix} = S_n^{-1}\begin{pmatrix} -\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix},$$

where

$$S_n = \begin{pmatrix} \dfrac{\partial \boldsymbol{Q}_{1n}}{\partial \boldsymbol{t}^{\mathrm{T}}} & \dfrac{\partial \boldsymbol{Q}_{1n}}{\partial \boldsymbol{\theta}} \\ \dfrac{\partial \boldsymbol{Q}_{2n}}{\partial \boldsymbol{t}^{\mathrm{T}}} & \mathbf{0} \end{pmatrix}_{(\boldsymbol{\theta}_0,\mathbf{0})}.$$

By Lemmas A.1 and A.2, we have

$$S_n \to \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & 0 \end{pmatrix} = \begin{pmatrix} -E(\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}) & E(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\theta}}) \\ E(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\theta}})^{\mathrm{T}} & \mathbf{0} \end{pmatrix}.$$

From this and $\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) = \sum_{i=1}^n \omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta}_0) = O_p(n^{-1/2})$, we have that $\delta_n = O_p(n^{-1/2})$. Thus,

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = S_{22.1}^{-1}S_{21}S_{11}^{-1}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + o_p(1), \quad \text{as } n\to\infty, \tag{A2}$$

where $S_{22.1} = E(\partial \boldsymbol{g}/\partial\boldsymbol{\theta})^{\mathrm{T}}(E\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}})^{-1}E(\partial \boldsymbol{g}/\partial\boldsymbol{\theta})$. By Equation (A2), Lemmas A.1 and A.2, we obtain

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \xrightarrow{D} N(0,cV), \quad \text{as } n\to\infty,$$

where $V = S_{22.1}^{-1} = \{E(\partial \boldsymbol{g}/\partial\boldsymbol{\theta})^{\mathrm{T}}(E\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}})^{-1}E(\partial \boldsymbol{g}/\partial\boldsymbol{\theta})\}^{-1}$ and c is given in condition (C6). ■

Proof of Theorem 2.3 The WEL ratio test statistic is

$$W(\boldsymbol{\theta}_0) = 2\left[\sum_{i=1}^n n\omega_{ni}\log\{1+\boldsymbol{t}_0^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta}_0)\} - \sum_{i=1}^n n\omega_{ni}\log\{1+\tilde{\boldsymbol{t}}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{X}_i,\tilde{\boldsymbol{\theta}})\}\right].$$

Note that

$$l(\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{t}}) = \sum_{i=1}^n n\omega_{ni}\log\{1+\tilde{\boldsymbol{t}}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{X}_i,\tilde{\boldsymbol{\theta}})\} = -\frac{n}{2}\boldsymbol{Q}_{1n}^{\mathrm{T}}(\boldsymbol{\theta}_0,\mathbf{0})A\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + o_p(1),$$

where $A = S_{11}^{-1}\{I + S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\}$. Under $H_0:\ \boldsymbol{\theta}=\boldsymbol{\theta}_0$, we obtain

$$\sum_{i=1}^n \frac{\omega_{ni}\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta}_0)}{1+\boldsymbol{t}_0^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta}_0)} = 0 \Longrightarrow \boldsymbol{t}_0 = -S_{11}^{-1}\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + o_p(1),$$

and

$$\sum_{i=1}^n n\omega_{ni}\log\{1+\boldsymbol{t}_0^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{X}_i,\boldsymbol{\theta}_0)\} = -\frac{n}{2}\boldsymbol{Q}_{1n}^{\mathrm{T}}(\boldsymbol{\theta}_0,\mathbf{0})S_{11}^{-1}\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + o_p(1).$$

From condition (C6), we have

$$\frac{1}{n} \sim \sum_{i=1}^n \omega_{ni}^2 \geq \max_{1\leq i\leq n}\omega_{ni}^2.$$

It follows that $\max_{1\leq i\leq n}\omega_{ni}^2 \to 0$, as $n\to\infty$. By Lemmas A.1 and A.2, we obtain

$$\begin{aligned} W(\boldsymbol{\theta}_0) &= n\boldsymbol{Q}_{1n}^{\mathrm{T}}(\boldsymbol{\theta}_0,\mathbf{0})(A - S_{11}^{-1})\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + o_p(1) \\ &= n\boldsymbol{Q}_{1n}^{\mathrm{T}}(\boldsymbol{\theta}_0,\mathbf{0})S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0}) + o_p(1) \\ &= \{(-S_{11})^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0})\}^{\mathrm{T}}\{(-S_{11})^{-1/2}S_{12}S_{12}S_{22.1}^{-1}S_{21}(-S_{11})^{-1/2}\} \\ &\quad \times \{(-S_{11})^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0,\mathbf{0})\} + o_p(1). \end{aligned}$$

Note that $(-S_{11})^{-1/2}\sqrt{n}\boldsymbol{Q}_{1n}(\boldsymbol{\theta}_0, \boldsymbol{0})$ converges in distribution to a standard multivariate normal distribution and that $(-S_{11})^{-1/2}S_{12}S_{12}S_{22.1}^{-1}S_{21}(-S_{11})^{-1/2}$ is symmetric and idempotent with trace equal to p. Hence, the EL ratio statistic $W(\boldsymbol{\theta}_0)$ converges in distribution to χ_p^2. ■

Proof of Theorem 3.1 For a weight vector given by Equation (6), we have

$$\omega_{ni} \geq 0, \quad \sum_{i=1}^{n} \omega_{ni} = 1,$$

and since $\int D(\boldsymbol{x}; F)\, \mathrm{d}F(\boldsymbol{x}) > 0$ and $0 \leq D(\boldsymbol{x}; F) \leq 1$,

$$\sum_{i=1}^{n} n\omega_{ni}^2 = \frac{(1/n)\sum_{i=1}^{n} D^2(\boldsymbol{X}_i; F_n)}{\{(1/n)\sum_{i=1}^{n} D(\boldsymbol{X}_i; F_n)\}^2} \xrightarrow{\text{a.s.}} \frac{\int D^2(\boldsymbol{x}; F)\mathrm{d}F(\boldsymbol{x})}{\{\int D(\boldsymbol{x}; F)\, \mathrm{d}F(\boldsymbol{x})\}^2} \text{ as } n \longrightarrow \infty. \qquad ■$$