*GJMS Volume 3(2023), No 1* 1–8

# An extended empirical likelihood approach to improved estimation

Bong-Jin Choi[1] and Mingao Yuan[1]

[1]North Dakota State University, Department of Statistics
Fargo, ND 58710, USA

**ABSTRACT:** The extended empirical likelihood has been proposed recently to improve the coverage accuracy of the empirical likelihood ratio confidence region. In this paper, we use the extended empirical likelihood(EEL) to incorporate side information to improve the efficiency of the empirical estimator of some linear functions. We get the asymptotic normality of the EEL-weighted estimator, and our simulation study shows that the EEL-weighted estimator performs better than the usual empirical likelihood(EL) weighted estimator and the empirical estimator, especially when the sample size is small.

***Keywords***: empirical likelihood, extended empirical likelihood, side information, U-statistics, Jackknife empirical likelihood.

## 1 INTRODUCTION

The empirical likelihood(EL) was introduced by Owen(1990, 2001) to construct a confidence region in a nonparametric setting. As an analog of the parametric likelihood method, it has been extensively applied to different fields due to some of the nice properties. As nonparametric method, it doesn't require a prespecified distribution for the data. The confidence region respects the range of the data and performs better than that based on asymptotic normality. Later, EL has been applied to many statistical fields(Cheng 2012, Yuan and Zhang 2019). Besides, it is very convenient to use empirical likelihood method to incorporate side information and construct more efficient point estimator. Qin and Lawless(1994) proposed the maximum empirical likelihood estimator(MELE). They adopted the empirical likelihood to combine information about the parameters and showed that the MELE's are asymptotically efficient in some sense. Since then, many researchers worked on MELE. Among them, Zhang(1995, 1997, 1999) used it to improve estimates in M-estimation, quantile process and bootstrapping in the presence of auxiliary information. Yuan et al.(2012) studied U-statistics with side information and Peng (2015) investigated MELE for differentiable statistical functionals, von-Mises functions, L-estimators and U-statistics. The estimators obtained by MELE is called empirical likelihood(EL) weighted estimator.

However, the empirical likelihood method has two practical issues(Tsao 2013). The first one is the rate at which the likelihood ratio converges to the limiting chi-square distribution and the second one is the convex hull constraint confines the confidence region to a bounded region inside the parameter space. To overcome these two issues, many correction approaches have been proposed(DiCiccio, Hall and Romano(1991), Chen(1993), Chen, Variyath and Abraham(2008), Emerson and Owen(2009)). Especially, Tsao(2013) introduced the extended empirical likelihood(EEL) method to address these two issues by expanding the parameter space and making a transformation of the original empirical likelihood ratio. Their simulation shows that the extended empirical likelihood method performs better in both large and small sample situtations.

Due to the good performance of the extended empirical likelihood(Tsao (2013)), in this paper, we use it to incorporate side information to improve the efficiency of the empirical estimator. We call the resulted estimator the extended empirical likelihood weighted(EEL-weighted) estimator. Compared to the usual empirical likelihood weighted estimator in Wang(2015), the EEL-weighted estimator performs better in terms of MSE, especially when the sample size is smaller.

In section 2, we describe the extended empirical likelihood(EEL) weighted estimator and present the main results. In section 3, a simulation study is presented while in section 4 we give the proof of the main results.

## 2 Extended Empirical Likelihood Weighted Estimator

Let $X \in \mathbb{R}^d$ be a random vector with distribution $F$ and $X_1, \dots, X_n$ be $n$ i.i.d. copies of $X$. We want to estimate the linear functional $\delta = E(\psi(X))$, where $\psi$ is an square-integrable univariate or multivariate function. Suppose side information is available and given by

$$E[g(X, \theta_0)] = 0,$$

where $g(X, \theta_0)$ is a given $p$-dimensional function and $\theta_0 \in \mathbb{R}^p$ is known. The empirical estimator of $\delta$ is $\hat{\delta} = \frac{1}{n}\sum_{i=1}^n \psi(X_i)$. Clearly, the empirical estimator doesn't use the side information. Hence, it is not efficient in some sense.

Received August 15, 2019 

To incorporate the side information, consider the empirical likelihood ratio(Owen 1990,2001)

$$\mathcal{R}_n(\theta_0) = \sup\left\{ \prod_{1\le i\le n} n\pi_i \,\middle|\, \sum_{i=1}^n \pi_i g(X_i,\theta_0) = 0, \pi_i \ge 0, \sum_{i=1}^n \pi_i = 1 \right\}.$$

By the method of Lagrange multipliers(Owen 2001), the solution to the above optimization problem is

$$\tilde{\pi}_i = \frac{1}{n}\frac{1}{1+\eta^T g(X_i,\theta_0)},$$

where $\eta \in \mathbb{R}^p$ satisfies the following equation

$$\frac{1}{n}\sum_{i=1}^n \frac{g(X_i,\theta_0)}{1+\eta^T g(X_i,\theta_0)} = 0.$$

Then the EL-weighted estimator(Wang 2015) of $\delta$ is

$$\hat{\delta}_{EL} = \sum_{i=1}^n \tilde{\pi}_i \psi(X_i).$$

Since $\tilde{\pi}_i$ contains the side information, the EL-weighted estimator is more efficient than the empirical estimator(Wang(2015)). The asymptotic normality of the JEL-weighted estimator is given in Wang(2015).

Alternatively, we use the extended empirical likelihood to incorporate the side information. Let $\tilde{\theta}$ be the maximum empirical likelihood estimator(MELE) of $\theta_0$, that is, $\tilde{\theta} = arg\max \mathcal{R}_n(\theta)$. Then the extended empirical likelihood defined by Tsao(2013) is

$$\mathcal{R}_n^*(\theta_0) = \sup\left\{ \prod_{1\le i\le n} n\pi_i \,\middle|\, \sum_{i=1}^n \pi_i g(X_i,\theta_0 - k_n(\theta_0-\tilde{\theta})) = 0, \pi_i \ge 0, \sum_{i=1}^n \pi_i = 1 \right\},$$

where $k_n = 1 - \frac{1}{r_n}$ and $r_n > 0$. Similary, the solution is

$$\pi_i = \frac{1}{n}\frac{1}{1+\lambda^T g(X_i,\theta_0 - k_n(\theta_0-\tilde{\theta}))},$$

with $\lambda$ subject to the equation

$$\frac{1}{n}\sum_{i=1}^n \frac{g(X_i,\theta_0 - k_n(\theta_0-\tilde{\theta}))}{1+\lambda^T g(X_i,\theta_0 - k_n(\theta_0-\tilde{\theta}))} = 0.$$

We define the EEL-weighted estimator as

$$\hat{\delta}_{EEL} = \sum_{i=1}^n \pi_i \psi(X_i).$$

To get the asymptotic distribution of the EEL-weighted estimator, we assume the following conditions hold for $g(X,\theta)$.

(C1) $E(g(X,\theta_0)) = 0$ and $0 < Var(g(X,\theta_0)) < +\infty$ is positive definite.
(C2) The first and second partial derivatives of $g(X,\theta)$ with respect to $\theta$ are all continuous in $\theta$. And they are bounded in norm by an integrable function of $X$ in a neighborhood of $\theta_0$.

Under condition (C1) and (C2) and assuming $k_n \to 0$ as $n \to +\infty$, we have

$$\hat{\delta}_{EEL} = \bar{\psi} - \bar{\phi} + o_p(\frac{1}{\sqrt{n}}),$$

where $\bar{\phi} = \frac{1}{n}\sum_{i=1}^n \phi(X_i)$, $\phi(X) = E[\psi(X)g^T(X,\theta_0)]W^{-1}g(X,\theta_0)$ and $W = Var(g(X,\theta_0))$. If $\Sigma_0 = Var(\psi) - Var(\phi)$ is positive definite, then

$$\sqrt{n}(\hat{\delta}_{EEL} - \delta) \Rightarrow N(0,\Sigma_0), \;\; n \to +\infty.$$

Hence, $\hat{\delta}_{EEL}$ has smaller asymptotic variance than the empirical estimator.

The expansion and asymptotic distribution in Theorem 1 is the same as that of the EL-weighted estimator, Theorem 2.2.1 in Wang(2015). This is due to the fact that when $k_n$ goes to zero, the transformed term $\theta_0 - k_n(\theta_0 - \tilde{\theta})$ will tend to $\theta_0$. Hence for large sample size, the EEL-weighted estimator and the EL-weighted estimator are expected to have similar performance. However, Tsao(2013) showed the extended empirical likelihood has higher coverage accuracy than the usual empirical likelihood for small sample size. Hence the EEL weighted estimator will take more advantage of the side information and be more efficient than the EL-weighted estimator when sample size is small, which is confirmed by our simulation study in Section 3.

Sometimes the side information is given by U-statistics(Yuan et al 2012, Wang 2015). Let $h : \mathbb{R}^m \to \mathbb{R}^p$ be a permutation symmetric function with parameter $\theta \in \mathbb{R}^q$. The U-statistics with kernel $h$ of order $m$ is defined as

$$U_n(h) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}; \theta).$$

We assume that $0 < Var(h) < +\infty$ and the side information is given by

$$E[U_n(h)] = E[h(X_1, \ldots, X_m; \theta_0)] = 0.$$

However, we can not construct empirical likelihood directly using the above equation. We employ the jackknife pseudo value method(Jing et al 2009). The jackknife pseudo values of the U-statistics are defined as

$$V_{nj} = nU_n - (n-1)U_{n-1}^{(-j)}, \ \ j = 1, 2, \ldots, n,$$

where $U_{n-1}^{(-j)}$ is the U-statistics based on $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$. It is shown that $V_{nj}$ are asymptotically independent(Lin 2013). Then we can define the jackknife empirical likelihood(JEL)(Wang 2015) as

$$\mathcal{R}_n(\theta_0) = \sup\left\{ \prod_{i=1}^n nw_i : \sum_{i=1}^n w_i V_{ni}(\theta_0) = 0, w_i > 0, \sum_{i=1}^n w_i = 1 \right\}.$$

The solution is

$$\tilde{w}_i = \frac{1}{n} \frac{1}{1 + \tau^T V_{ni}(\theta_0)},$$

where $\tau$ satisfies

$$\frac{1}{n} \sum_{i=1}^n \frac{\tau^T V_{ni}(\theta_0)}{1 + \tau^T V_{ni}(\theta_0)} = 0.$$

The jackknife empirical likelihood(JEL) weighted estimator(Wang 2015) is

$$\hat{\delta}_{JEL} = \sum_{i=1}^n \tilde{w}_i \psi(X_i).$$

The asymptotic normality of the JEL-weighted estimator is given in Wang(2015).

Let $\tilde{\theta}$ be the maximum jackknife empirical likelihood estimator of $\theta_0$. Then we define the jackknife extended empirical likelihood(JEEL) as

$$\mathcal{R}_n^*(\theta_0) = \sup\left\{ \prod_{i=1}^n nw_i : \sum_{i=1}^n w_i V_{ni}(\theta_0 + k_n(\tilde{\theta} - \theta_0)) = 0, w_i > 0, k_n > 0, \sum_{i=1}^n w_i = 1 \right\}.$$

The solution is

$$w_i = \frac{1}{n} \frac{1}{1 + \xi^T V_{ni}(\theta_0 + k_n(\tilde{\theta} - \theta_0))},$$

where $\xi$ satisfies

$$\frac{1}{n} \sum_{i=1}^n \frac{V_{ni}(\theta_0 + k_n(\tilde{\theta} - \theta_0))}{1 + \xi^T V_{ni}(\theta_0 + k_n(\tilde{\theta} - \theta_0))} = 0.$$

Define the jackknife extended empirical likelihood weighted estimator as

$$\hat{\delta}_{JEEL} = \sum_{i=1}^n \pi_i \psi(X_i).$$

To get the asymptotic distribution, let's state the following assumption for $h_1(x) = E[h(X_1, X_2, \ldots, X_m)|X_1 = x]$.
(A1) $h(X_1, \ldots, X_m; \theta)$ is twice continuously differentiable with respect to $\theta$. The norm of the first and second partial derivatives are controlled by some square-integrable function in a neighborhood of $\theta_0$.
(A2) $E||h(X_1, \ldots, X_m; \theta)||^2 \le M < +\infty$ in a neighborhood of $\theta_0$ for $M > 0$.
(A3) $0 < Var(h(X_1, \ldots, X_m; \theta_0)) < +\infty$.

Then we have Under the assumptions (A1)-(A3) and $nk_n = o(1)$, we have

$$\hat{\delta}_{JEEL} = \bar{\psi} - \bar{\phi} + o_p(\frac{1}{\sqrt{n}}),$$

where $\bar{\phi} = \frac{1}{n} \sum_{i=1}^n \phi(X_i)$, $\phi(X) = E[\psi(X) h_1^T(X; \theta_0)] W^{-1} h_1(X; \theta_0)$ and $W = Var(h_1(X; \theta_0))$. If $\Sigma_0 = Var(\psi) - Var(\phi)$ is positive definite, then

$$\sqrt{n}(\hat{\delta}_{JEEL} - \delta) \Rightarrow N(0, \Sigma_0), \ \ n \to +\infty.$$

Hence, $\hat{\delta}_{JEEL}$ has smaller asymptotic variance than the empirical estimator.

Table 1: **Simulated MSE Ratios**

| $(n, r_n)$ | (10, 1.38) | (20, 1.3) | (30, 1.29) | (40, 1.28) | (50, 1.25) | (80, 1.2) | (150, 1.1) |
|---|---|---|---|---|---|---|---|
| | | | | $X \sim \mathcal{N}(1, 2^2)$ | | | |
| EL | 0.7312 | 0.8482 | 0.8428 | 0.8548 | 0.8860 | 0.9148 | 0.8704 |
| EEL | 0.7201 | 0.8015 | 0.8083 | 0.8159 | 0.8452 | 0.8702 | 0.8532 |
| | | | | $X \sim \mathcal{N}(1, 5^2)$ | | | |
| EL | 0.7841 | 0.9578 | 0.9410 | 0.9856 | 0.9907 | 0.9661 | 0.9714 |
| EEL | 0.7658 | 0.9089 | 0.8838 | 0.9348 | 0.9514 | 0.9411 | 0.9603 |
| | | | | $X \sim \mathcal{N}(1, 10^2)$ | | | |
| EL | 0.8684 | 0.9276 | 0.9961 | 0.9824 | 0.9503 | 1.0210 | 1.0019 |
| EEL | 0.8387 | 0.8859 | 0.9498 | 0.9310 | 0.9234 | 0.9875 | 0.9913 |
| | | | | $X \sim 0.5\mathcal{N}(0, 5^2) + 0.5\mathcal{N}(4, 5^2)$ | | | |
| EL | 0.8635 | 0.8777 | 0.8456 | 0.8603 | 0.8538 | 0.8645 | 0.7952 |
| EEL | 0.8154 | 0.8479 | 0.8255 | 0.8305 | 0.8321 | 0.8484 | 0.7943 |
| | | | | $X \sim 0.5\mathcal{N}(0, 4^2) + 0.5\mathcal{N}(4, 15^2)$ | | | |
| EL | 0.9042 | 0.9204 | 0.9277 | 0.9162 | 0.9016 | 0.9127 | 0.8824 |
| EEL | 0.8561 | 0.8839 | 0.8968 | 0.8831 | 0.8784 | 0.8934 | 0.8743 |
| | | | | $X \sim 0.2\mathcal{N}(0, 7^2) + 0.8\mathcal{N}(2, 7^2)$ | | | |
| EL | 0.9688 | 0.9836 | 0.9700 | 0.9783 | 0.9888 | 0.9323 | 0.9214 |
| EEL | 0.9097 | 0.9428 | 0.9379 | 0.9432 | 0.9605 | 0.9200 | 0.9164 |
| | | | | $X \sim Laplace(0, 1)$ | | | |
| EL | 0.8460 | 0.9094 | 0.9516 | 0.9773 | 0.9898 | 1.0383 | 1.0283 |
| EEL | 0.8181 | 0.8818 | 0.9227 | 0.9489 | 0.9637 | 1.0090 | 1.0133 |
| | | | | $X \sim Laplace(0, 5)$ | | | |
| EL | 0.7502 | 0.9185 | 0.9715 | 0.9857 | 0.9672 | 1.0299 | 1.0162 |
| EEL | 0.7227 | 0.8859 | 0.9337 | 0.9526 | 0.9474 | 0.9971 | 1.0011 |
| | | | | $X \sim Laplace(4, 8)$ | | | |
| EL | 0.7391 | 0.8993 | 0.9207 | 0.9204 | 0.9337 | 0.9673 | 0.9734 |
| EEL | 0.7140 | 0.8692 | 0.8900 | 0.8852 | 0.8991 | 0.9339 | 0.9596 |
| | | | | $X \sim Logistic(0, 1)$ | | | |
| EL | 0.9007 | 0.9770 | 0.9836 | 0.9856 | 1.0187 | 1.0342 | 1.0384 |
| EEL | 0.8613 | 0.9398 | 0.9504 | 0.9569 | 0.9897 | 1.0098 | 1.0267 |
| | | | | $X \sim Logistic(2, 4)$ | | | |
| EL | 0.8574 | 0.9036 | 0.9473 | 0.9301 | 0.9614 | 0.9791 | 0.9435 |
| EEL | 0.8144 | 0.8685 | 0.9125 | 0.8992 | 0.9283 | 0.9525 | 0.9336 |

## 3 Simulation Study

In this section, we present simulation results to demonstrate the performance of EEL-weighted estimator for small sample sizes. To do this, we assume the mean value is known, that is, $E(X) = \theta_0$ and estimate the moments as in Example 5.2 in Zhang(1999). In this case, the side information is given by $E[g(X, \theta_0)] = 0$ with $g(x, \theta) = x - \theta$. We estimate the second or fourth moment by the empirical estimator, the EL-weighted estimator and the EEL-weighted estimator. The sample sizes are $n = 10, 20, 30, 40, 50, 80, 150$, and we take $\gamma_n = 1.38, 1.3, 1.29, 1.28, 1.25, 1.2, 1.1$. To compare the performance of the three estimators, we run 2000 simulations under each sample size to compute the mean square error(MSE) of each estimator and report the ratios of MSE of the EL-weighted or EEL-weighted estimator to MSE of the emipirical estimator. The smaller ratio implies the estimator is better compared to the empirical estimator. We draw data from normal distribution, mixture distribution, Laplace distribution and Logistic distribution. For the normal case, we estimate the fourth moment, while the second moment is estimated in the rest cases with the known mean as side information.

The table 1 below shows the simulation result. For smaller sample sizes 10, 20, 30, 40, 50, all the ratios are smaller than 1 implying that both the EEL-weighted and the EL-weighted estimator are better than the empirical estimator. Besides, the MSE ratio of EEL-weighted estimator to the empirical estimator is smaller than the MSE ratio of the EL-weighted estimator to the empirical estimator, hence the EEL-weighted estimator is better than the El-weighed estimator. For larger sample sizes 80 and 150, the EEL-weighted estimator and the EL-estimator have approximately the same MSE with the ratios close to 1, which implies that for larger sample size, the three estimators performs very similarly as expected. Hence, even though the EEL-weighted has the same asymptotic distributon as the EL-weighed estimator, for smaller sample size, the EEL-weighted estimator is better than the EL-weighted estimator and the empirical estimator.

# 4 Proofs of Main Results

We prove Theorem 1 for the case when $g(X,\theta)=u(X)-\theta$. The general case can be proved in a similar way. Notice under the condition of the theorem, $||\eta||=O_p(\frac{1}{\sqrt{n}})$ and $\max_{1\le i\le n}|\eta^T(u(X_i)-\theta_0)|=o_p(1)$ by Owen (2001).

Let $Z_i=u(X_i)-\bar{u}-\frac{1}{\gamma_n}(\theta_0-\bar{u})$ and $Y_i=u(X_i)-\theta_0$. Suppose $0<Var(u(X_1))<+\infty$ and $r_n\to 1$ as $n\to+\infty$. Then we have

(1) $Z_i=Y_i+o_p(\frac{1}{\sqrt{n}})$.

(2) $\max_{1\le i\le n}|Z_i|=o_p(\frac{1}{\sqrt{n}})$.

(3) $\frac{1}{n}\sum_{i=1}^n|Z_i|^3=o_p(\frac{1}{\sqrt{n}})$.

**Proof:** (1) By the central limit theorem, we have

$$\begin{aligned} Z_i &= u(X_i)-\bar{u}-\frac{1}{\gamma_n}(\theta_0-\bar{u}) \\ &= u(X_i)-\theta_0+(\frac{1}{\gamma_n}-1)(\bar{u}-\theta_0) \\ &= u(X_i)-\theta_0+o_p(\frac{1}{\sqrt{n}}). \end{aligned}$$

(2) By Lemma 11.2 in Owen(2001), it's easy to get

$$\begin{aligned} \max_{1\le i\le n}|Z_i| &= \max_{1\le i\le n}|Y_i+o_p(\frac{1}{\sqrt{n}})| \\ &\le \max_{1\le i\le n}|Y_i|+o_p(\frac{1}{\sqrt{n}}) \\ &= o_p(\sqrt{n}). \end{aligned}$$

(3) By Lemma 11.3 in Owen(2001), we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n|Z_i|^3 &= \frac{1}{n}\sum_{i=1}^n|Y_i+o_p(\frac{1}{\sqrt{n}})|^3 \\ &\le \frac{1}{n}\sum_{i=1}^n|Y_i|^3+o_p(\frac{1}{\sqrt{n}}) \\ &= o_p(\sqrt{n}). \end{aligned}$$

□

Under the conditions of Theorem 1, $||\lambda||=O_p(\frac{1}{\sqrt{n}})$ and $\max_{1\le i\le n}|\lambda^TZ_i|=o_p(1)$, and $\lambda=\eta+o_p(\frac{1}{\sqrt{n}})$.

**Proof:** Let $\tilde{Y}_i=\lambda^TZ_i$, $Z_n^*=\max_{1\le i\le n}||Z_i||$, $\lambda=||\lambda||\theta$ with $||\theta||=1$. Note that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n Z_i-\frac{1}{n}\sum_{i=1}^n\frac{Z_i\tilde{Y}_i}{1+\tilde{Y}_i} &= \frac{1}{n}\sum_{i=1}^n Z_i(1-\frac{\tilde{Y}_i}{1+\tilde{Y}_i}) \\ &= \frac{1}{n}\sum_{i=1}^n\frac{Z_i}{1+\tilde{Y}_i}=0. \end{aligned}$$

Then we have

$$\begin{aligned} \frac{1}{\gamma_n}(\bar{u}-\theta_0) &= \frac{1}{n}\sum_{i=1}^n Z_i \\ &= \frac{1}{n}\sum_{i=1}^n\frac{Z_i\tilde{Y}_i}{1+\tilde{Y}_i} \\ &= \frac{1}{n}\sum_{i=1}^n\frac{Z_iZ_i^T\theta}{1+\tilde{Y}_i}||\lambda||, \end{aligned}$$

from which yields

$$\begin{aligned} \frac{1}{\gamma_n}\theta^T(\bar{u}-\theta_0) &= \frac{1}{n}\sum_{i=1}^n\frac{\theta^TZ_iZ_i^T\theta}{1+\tilde{Y}_i}||\lambda|| \\ &= \theta^T\left(\frac{1}{n}\sum_{i=1}^n\frac{Z_iZ_i^T}{1+\tilde{Y}_i}\right)\theta||\lambda||. \end{aligned}$$

Note that

$$\theta^T\left(\frac{1}{n}\sum_{i=1}^n\frac{Z_iZ_i^T}{1+\tilde{Y}_i}[1+\max_{1\le i\le n}\tilde{Y}_i]\right)\theta||\lambda||\ge\theta^T\left(\frac{1}{n}\sum_{i=1}^n Z_iZ_i^T\right)\theta||\lambda||,$$

then

$$\begin{aligned}\frac{1}{\gamma_n}\theta^T(\bar{u}-\theta_0)(1+||\lambda||Z_n^*) &\geq \frac{1}{\gamma_n}\theta^T(\bar{u}-\theta_0)[1+\max_{1\leq i\leq n}\tilde{Y}_i] \\ &\geq \theta^T\left(\frac{1}{n}\sum_{i=1}^n Z_i Z_i^T\right)\theta||\lambda||.\end{aligned}$$

Hence,

$$||\lambda||[\theta^T S_1\theta - \frac{1}{\gamma_n}\theta^T(\bar{u}-\theta_0)Z_n^*] \leq \frac{1}{\gamma_n}\theta^T(\bar{u}-\theta_0),$$

where

$$\begin{aligned}S_1 &= \frac{1}{n}\sum_{i=1}^n Z_i Z_i^T \\ &= \frac{1}{n}\sum_{i=1}^n\left(u(X_i)-\theta_0+o_p(\frac{1}{\sqrt{n}})\right)\left(u(X_i)-\theta_0+o_p(\frac{1}{\sqrt{n}})\right)^T \\ &= S+o_p(\frac{1}{n^\delta}),\ \ (1<\delta<1).\end{aligned}$$

Then we conclude that $||\lambda|| = O_p(\frac{1}{\sqrt{n}})$ and

$$\max_{1\leq i\leq n}|\tilde{Y}_i| = O_p(\frac{1}{\sqrt{n}})o_p(\sqrt{n}) = o_p(1).$$

Then one has

$$\begin{aligned}0 &= \frac{1}{n}\sum_{i=1}^n Z_i(1-\tilde{Y}_i+\frac{\tilde{Y}_i^2}{1+\tilde{Y}_i}) \\ &= u(X_i)-\theta_0-S_1\lambda+\frac{1}{n}\sum_{i=1}^n\frac{Z_i\tilde{Y}_i^2}{1+\tilde{Y}_i}.\end{aligned}$$

Note that

$$\frac{1}{n}\sum_{i=1}^n\frac{||Z_i||||\tilde{Y}_i||^2}{|1+\tilde{Y}_i|} = o_p(\frac{1}{\sqrt{n}}).$$

Then we have

$$\lambda = S(u(X_i)-\theta_0)+o_p(\frac{1}{\sqrt{n}}) = \eta+o_p(\frac{1}{\sqrt{n}}).$$

□

Under the conditions of Theorem 1, $\pi_i = \tilde{\pi}_i - \frac{1}{n}o_p(\frac{1}{\sqrt{n}})(u(X_i-\theta_0))+\frac{1}{n}o_p(\frac{1}{\sqrt{n}})$.
**Proof:** By Lemma 2, we have

$$\begin{aligned}\pi_i &= \frac{1}{n}\frac{1}{1+\lambda^T[u(X_i)-\bar{u}-\frac{1}{\gamma_n}(\theta_0-\bar{u})]} \\ &= \frac{1}{n}\frac{1}{1+(\eta^T+o_p(\frac{1}{\sqrt{n}}))[u(X_i)-\bar{u}-\frac{1}{\gamma_n}(\theta_0-\bar{u})]} \\ &= \frac{1}{n}\frac{1}{1+\eta^T(u(X_i)-\theta_0)+o_p(\frac{1}{\sqrt{n}})(u(X_i)-\theta_0)+o_p(\frac{1}{\sqrt{n}})} \\ &= \frac{1}{n}\frac{1}{1+\eta^T(u(X_i)-\theta_0)} - \frac{1}{n}\frac{1}{1+\eta^T(u(X_i)-\theta_0)}\frac{o_p(\frac{1}{\sqrt{n}})(u(X_i)-\theta_0)+o_p(\frac{1}{\sqrt{n}})}{1+\lambda^T[u(X_i)-\bar{u}-\frac{1}{\gamma_n}(\theta_0-\bar{u})]} \\ &= \tilde{\pi}_i-\frac{1}{n}o_p(\frac{1}{\sqrt{n}})(u(X_i-\theta_0))+\frac{1}{n}o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

□

Then we are ready to prove Theorem 1.
**Proof of Theorem 1:** By Lemma 3 we have

$$\begin{aligned}\hat{\delta}_{EEL} &= \sum_{i=1}^n\pi_i\psi(X_i) \\ &= \sum_{i=1}^n[\tilde{\pi}_i-\frac{1}{n}o_p(\frac{1}{\sqrt{n}})(u(X_i-\theta_0))+\frac{1}{n}o_p(\frac{1}{\sqrt{n}})]\psi(X_i) \\ &= \sum_{i=1}^n\tilde{\pi}_i\psi(X_i)-o_p(\frac{1}{\sqrt{n}})\frac{1}{n}\sum_{i=1}^n\psi(X_i)(u(X_i)-\theta_0)+o_p(\frac{1}{\sqrt{n}}) \\ &= \bar{\psi}-\bar{\phi}+o_p(\frac{1}{\sqrt{n}}),\end{aligned}$$

from which yields the desired result by Theorem 2.2.1 in Wang (2015).

□

**Proof of Theorem 2:** Let $\theta_n = \theta_0 + k_n(\tilde{\theta} - \theta_0)$. Then by Taylor expansion,

$$\begin{aligned} V_{nj}(\theta_n) &= nU_n - (n-1)U_{n-1}^{(-j)} \\ &= n\Big(U_n(\theta_0) + \frac{\partial U_n(\theta_0^*)}{\partial \theta} k_n(\tilde{\theta} - \theta_0)\Big) \\ &\quad -(n-1)\Big(U_n^{(-j)}(\theta_0) + \frac{\partial U_{n-1}^{(-j)}(\theta_0^{**})}{\partial \theta} k_n(\tilde{\theta} - \theta_0)\Big) \\ &= V_{nj}(\theta_0) + nk_n(\tilde{\theta} - \theta_0)\Big(\frac{\partial U_n(\theta_0^*)}{\partial \theta} - \frac{n-1}{n}\frac{\partial U_{n-1}^{(-j)}(\theta_0^{**})}{\partial \theta}\Big), \end{aligned}$$

where $\theta_0^*$ and $\theta_0^{**}$ are between $\theta_0$ and $\theta_n$. By Theorem 3.1.2 in Li (2016), $\tilde{\theta} - \theta_0 = O_p(\frac{1}{\sqrt{n}})$. By (A1), $\frac{\partial U_n(\theta_0^*)}{\partial \theta}$ and $\frac{\partial U_{n-1}^{(-j)}(\theta_0^{**})}{\partial \theta}$ are bounded in probability. If $nk_n \to 0$ as $n \to +\infty$, then we have

$$V_{nj}(\theta_n) = V_{nj}(\theta_0) + o_p(\frac{1}{\sqrt{n}}).$$

Secondly, note that by Lemma 1.2.1 in Lin(2013),

$$V_{nj}(\theta_0) = mh_1(X_j;\theta_0) + O_p(\frac{1}{\sqrt{n}}).$$

Hence we have

$$V_{nj}(\theta_n) = mh_1(X_j;\theta_0) + O_p(\frac{1}{\sqrt{n}}).$$

Then we verify the conditions of Theorem 6.1 in Peng(2015) below

$$\begin{aligned} \max_{1\le j\le n} |V_{nj}(\theta_n)| &\le \max_{1\le j\le n} |mh_1(X_j;\theta_0)| + 1 = o_p(\sqrt{n}), \\ \frac{1}{n}\sum_{i=1}^n V_{nj}(\theta_n) &= \frac{1}{n}\sum_{i=1}^n mh_1(X_j;\theta_0) + O_p(\frac{1}{\sqrt{n}}) = O_p(\frac{1}{\sqrt{n}}) \\ \frac{1}{n}\sum_{i=1}^n V_{nj}(\theta_n)V_{nj}^T(\theta_n) &= \frac{m^2}{n}\sum_{j=1}^n h_1(X_j;\theta_0)h_1^T(X_j;\theta_0) + O_p(\frac{1}{\sqrt{n}})\frac{m}{n}\sum_{j=1}^n h_1(X_j;\theta_0) + O_p(\frac{1}{n}) \\ &= Var(h_1(X_1;\theta_0)) + o_p(1). \end{aligned}$$

Hence, by Theorem 6.1 in Peng(2015), we have $\xi = o_p(\frac{1}{\sqrt{n}})$. Similary $\tau = o_p(\frac{1}{\sqrt{n}})$. Hence in this case, we have

$$\begin{aligned} w_i &= \frac{1}{n}\frac{1}{1+\xi^T V_{ni}(\theta_n)} \\ &= \frac{1}{n}\frac{1}{1+(\tau^T + o_p(\frac{1}{\sqrt{n}}))(V_{ni}(\theta_0) + o_p(\frac{1}{\sqrt{n}}))} \\ &= \frac{1}{n}\frac{1}{1+\tau^T V_{ni}(\theta_0)} - \frac{1}{n}\frac{V_{ni}(\theta_0)o_p(\frac{1}{\sqrt{n}}) + o_p(\frac{1}{\sqrt{n}}))}{\big(1+\tau^T V_{ni}(\theta_0)\big)\big(1+\xi^T V_{ni}(\theta_n)\big)} \\ &= \tilde{w}_i - \frac{1}{n}\Big(V_{ni}(\theta_0)o_p(\frac{1}{\sqrt{n}}) + o_p(\frac{1}{\sqrt{n}})\Big), \end{aligned}$$

Then the JEEL-weighted estimator

$$\begin{aligned} \hat{\delta}_{JEEL} &= \sum_{i=1}^n w_i \psi(X_i) \\ &= \hat{\delta}_{JEL} - o_p(\frac{1}{\sqrt{n}})\frac{1}{n}\sum_{i=1}^n \psi(X_i)V_{ni}^T + o_p(\frac{1}{\sqrt{n}}) \end{aligned}$$

By Theorem 2.2.3 in Wang (2015), we get the desired result.

□

## References

## REFERENCES


[1] Chen, S.X.(1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Annals of Institute of Statistical Mathematics*,**93**, 215-220.

[2] Chen,J., Variyath, A.M. and Abraham,B.(2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, **3**, 426-443.

[3] Cheng,G., Zhao,Y. and Li,B.(2012). Empirical likelihood inferences for the semiparametric additive isotonic regression. *Journal of Multivariate Analysis*.**112** :172-182.

[4] DiCiccio, T.J., Hall, P. and Romano, J.P.(1990). Empirical likelihood is Bartlett correctable. *Annals of Statistics*, **19**: 1053-1061.

[5] Emerson, S.C. and Owen, A.B.(2009). Calibration of the empirical likelihood method for a vector mean.*Electronic Journal of Statistics*,**3**: 1161-1192.

[6] Jing, B.Y., Yuan, J.Q. and Zhou, W.(2009). Jackknife empirical likelihood.*Journal of the American Statistical Association*.**104(487)**:1224-1232.

[7] Li,L.N.(2016). Maximum empirical likelihood estimation in U-statistics based general estimating equations. *Ph.D. Dissertation*.

[8] Lin, Q. (2013). A jackknife empirical likelihood approach to goodness of fit U-statistics testing with side information. *Ph.D. Dissertation*.

[9] Owen, A.B.(1990). Empirical likelihood confidence region. *Annals of Statistics*,**18**: 90-120.

[10] Owen, A.B.(2001). *Empirical likelihood*, Chapman Hall/CRC, London.

[11] Peng, H.X.(2015). On a class of easy maximum empirical likelihood estimation. preprint.

[12] Qin,J. and Lawless, J.(1994). Empirical likelihood and general estimating equations. *Annal of Statistics*, **22**: 300-325.

[13] Tsao, M.(2013). Extending the empirical likelihood by domain expansion. *The Canadian Journal of Statistics*. **42(2)**: 257-274.

[14] Wang, S. (2015). An easy empirical likelihood approach to improved estimation. *Ph.D. Dissertation*.

[15] Yuan, A., He, W., Wang, B. and Qin, G.(2012). U-statistics with side information. *J. Multiv. Analy.* **111**: 20-38.

[16] Yuan, M. and Zhang, Y.(2019). Empirical Likelihood Inference for Partial Functional Linear Regression Models Based on B-spline, *International Journal of Statistics and Probability*, **8**: 135-142.

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

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

[19] Zhang, B. (1999). Bootstrapping with auxiliary information. *The Canadian Journal of Statistics*.**27(2)**:237-249.