Statistics and Its Interface Volume 9 (2016) 99–112

Kernel smoothing and jackknife empirical likelihood-based inferences for the generalized Lorenz curve*

Shan Luo and Gengsheng Qin†

Lorenz curve is one of the most commonly used devices for describing the inequality of income distributions. The generalized Lorenz curve is the Lorenz curve scaled by the mean of an income distribution and itself is an interesting object of study. In this paper, we define a smoothed estimator for the generalized Lorenz curve and propose a smoothed jackknife empirical likelihood method to construct confidence intervals for the generalized Lorenz curve. It is shown that the Wilks' theorem still holds for the smoothed jackknife empirical likelihood. Extensive simulation studies are conducted to compare the finite sample performances of the proposed methods with other methods based on simple random samples. Finally, the proposed methods are illustrated with a real example.

Keywords and phrases: Bootstrap, Confidence interval, Empirical likelihood, Generalized Lorenz curve, Jackknife.

1. INTRODUCTION

The Lorenz curve was first introduced by Lorenz (1905) for representing the inequality of a wealth distribution. Shorrocks (1983) and Kakwani (1984) extended the Lorenz curve to the *generalized Lorenz curve* (GLC) by taking into account differences between wealth distributions. Shorrocks (1983) also showed that one income distribution is preferable to another under any increasing and Schur-concave social welfare function if and only if the GLC of the first distribution lies above that of the second distribution. Because the GLC has these flexible properties, the GLC is an interesting object to study.

Consider an income variable $X \in [0, \infty)$ with cumulative distribution function $F(x)$. The generalized Lorenz ordinate is defined as follows:

$$\theta(t) = \int_0^{\xi_t} x dF(x), \quad 0 \le t \le 1, \tag{1}$$

where $\xi_t = F^{-1}(t)$ is the t-th quantile of $F(x)$. For a fixed $t \in [0, 1]$, the generalized Lorenz ordinate $\theta(t)$ is the average wealth owned by the wealth-holders below the bottom t-th percent. The height of the GLC is used to denote income level, and the convexity of the GLC is used to denote the extent of income inequality.

Previous studies have been conducted on the properties of generalized Lorenz curves. Thistle (1989a) showed that the income distribution function is uniquely determined by its GLC. Later, Thistle (1989b) derived the duality between the GLC and the distribution function, and showed that the generalized Lorenz dominance is equivalent to a second-order stochastic dominance. Kleiber and Kramer (2003) decomposed the generalized Lorenz order into two components: size and distribution. Xu (1997) proposed an asymptotic distribution-free test for the GLC. Beach and Davidson (1983) obtained the asymptotic normality of the empirical estimator for the GLC. Zheng (2002) further extended the asymptotic normality of the empirical estimator for the GLC to non-simple random samples. Inferences can be made based on these asymptotic normal distributions for the GLC. However, normal approximation-based inferences may have poor finite sample performances when income data is highly right skewed. Motivated by their previous research, several new non-parametric inferential methods will be developed for the GLC in this paper.

Empirical likelihood (EL), introduced by Owen (1988, 1990), allows researchers to utilize likelihood methods without assuming the data follows a specific parametric distribution. As mentioned in Wood et al. (1996), under mild conditions, the EL ratio statistic converges in distribution to a chi-square distribution. EL has wide applications in many fields. For example, without taking into account the high skewness of the health care cost distribution, the estimated mean cost can have a significant deviation from the actual average cost. Zhou, Qin, Lin and Li (2006) developed a new EL-based inference method in censored cost regression models and showed that the EL-based method outperforms the existing method. In medical diagnostics, Qin and Zhou (2006) proposed an EL-based inference method for the area under the ROC curve. Some recent developments of EL include inferences for: regression models (Feng and Peng, 2012), time series models (Chan, Li and Peng, 2012), risk

*The project was supported in part by a NSA grant. We would like to thank the editor, the associate editor and referees for their helpful comments which led to substantial improvements in this article.
†Corresponding author.

measures (Wei, Wen, and Zhu (2009), Wei and Zhu (2010), Li, Gong and Peng (2011)), and survey data (Rao and Wu, 2010). EL-based methods have been shown to have diverse advantages over other methods. We refer to Owen (2001) for more details.

Most of the previous EL-based methods were successfully applied with linear constraints. But significant computational burden arises with nonlinear functionals, due to the presence of nonlinear constraints in the underlying optimization problem. In order to reduce the computational burden, Jing, Yuan and Zhou (2009) proposed a jackknife empirical likelihood (JEL) method with application to U-statistics. The general idea of the JEL is to construct a jackknife pseudo-sample which is assumed to be asymptotically independent. Then the standard EL method is implemented to the jackknife pseudo-sample. Gong, Peng and Qi (2010) proposed a smoothed JEL for the ROC curve, and observed that the JEL method results in a shorter confidence interval than the naive bootstrap method. JEL method has been extended to risk measures (Peng et al., 2012), and high dimensional data (Wang, Peng and Qi, 2013). Recently, Yang, Qin and Belinga-Hall (2012) developed a plug-in EL method for the interval estimation of the GLC and concluded that the EL-based method outperforms the normal approximation-based methods. However, their EL ratio statistic follows a scaled chi-square distribution, which requires estimation of an unknown scale constant. To avoid estimating the unknown scale constant, a smoothed JEL for the GLC is proposed in this paper.

The remainder of this paper is organized as follows. In Section 2, a smoothed estimator for the GLC is defined, and its asymptotic normality is obtained. In Section 3, jackknife pseudo-values for the GLC are defined, and properties of the JEL are derived. In Section 4, confidence intervals based on normal approximation theory are established, while multiple bootstrap confidence intervals and a JEL-based confidence interval are developed. In Section 5, extensive simulation studies are conducted to evaluate the finite sample performances of the proposed intervals, and a 2012 individual income data for full-time professors from the University System of Georgia is used to illustrate an application of the recommended intervals. The proof of the main theorems for the GLC will be given in the Appendix.

2. THE SMOOTHED GENERALIZED LORENZ CURVE

As mentioned in the previous section, Shorrocks (1983) and Kakwani (1984) defined the generalized Lorenz curve as $\theta(t) = \int_0^{\xi_t} x dF(x),\ 0 \le t \le 1$. Let $X_1, X_2, ..., X_n$ be a simple random sample drawn from the population X with c.d.f. $F(x)$. For a fixed $t \in (0,1)$, the generalized Lorenz ordinate $\theta(t)$ satisfies $E[XI(X \le \xi_t)] - \theta(t) = 0$. An empirical estimate for $\theta(t)$ can be found from the following estimating equation

$$\frac{1}{n}\sum_{i=1}^{n} X_i I(F_n(X_i) \le t) - \theta(t) = 0,$$

where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$.

Therefore, the empirical estimator for the GLC $\theta(t)$ is

$$\hat{\theta}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i I(X_i \le F_n^{-1}(t)) = \frac{1}{n}\sum_{i=1}^{n} X_i I(t \ge F_n(X_i)).$$

However, $\hat{\theta}(t)$ is a non-smoothing estimator for $\theta(t)$. In many applications, $\theta(t)$ is a smoothing function. To find a smoothing estimator for $\theta(t)$, we use the kernel method. Kernel estimation has found wide usage in broad fields. Falk (1983, 1985) concluded that, for a distribution function $F(x)$, or its quantile function $F^{-1}(x)$, their corresponding kernel-based estimators asymptotically dominate their empirical estimators. Lloyd and Yong (1999) proved that the kernel estimator for the ROC curve performs better than the empirical estimator by having a smaller mean-square error, especially when sample size increases. In this paper, we propose a smoothed version of the empirical estimator for the GLC.

Define the kernel function as $K(x) = \int_{-\infty}^{x} \omega(y)dy$, where $\omega(\cdot)$ is a probability density function. The kernel estimator for $\theta(t)$ is defined as follows:

$$\hat{T}_n(t) = \frac{1}{n}\sum_{i=1}^{n} X_i K\left(\frac{t - F_n(X_i)}{h}\right), \tag{2}$$

where h is a bandwidth to be selected.

Theorem 2.1. *Assume $\omega(\cdot)$ is a probability density function with bounded support and its first derivative exists on its supporting set. If $h = h(n) \to 0$, $\sqrt{n}h^2 \to \infty$ as $n \to \infty$, then*

$$\sqrt{n}\{\hat{T}_n(t) - \theta(t)\} \xrightarrow{d} N(0, \sigma^2(t)),$$

where $\sigma^2(t) = \xi_t^2 t(1-t) + \int_0^{\xi_t} x^2 dF(x) - \theta^2(t)$.

In order to apply Theorem 2.1 in practice, the bandwidth h needs to be selected subtly. In this study, a 2-fold cross-validation method with equal sample split is used. The bandwidth h is chosen to be $cn^{-1/4}(\log\log n)^{3/5}$ where c is a constant, based on our simulation experiences. Clearly, the constant c will control the choice of bandwidth h. Here and thereafter, we denote $\hat{T}_{n,c}(t) = \hat{T}_n(t)$ with $h = cn^{-1/4}(\log\log n)^{3/5}$. At a given t, the constant c is chosen by minimizing the Mean Squared Error (MSE): $MSE(c) = E[\hat{T}_{n,c}(t) - \theta(t)]^2$. To estimate $MSE(c)$, we randomly split the original sample into a training sample and a validation sample. A kernel estimate $\hat{T}_{n,c}^{(1)}(t)$ for $\theta(t)$ is obtained based on the training sample, and an empirical estimate $\hat{\theta}^{(2)}(t)$ for $\theta(t)$ is obtained based on the validation sample. By repeating this random split L times (based on our simulation study, $L \geq 30$ is the recommended number of random splits), we obtain a set of kernel estimates and empirical estimates $\{(\hat{T}_{n,c}^{(1,l)}(t), \hat{\theta}^{(2,l)}(t)) : l = 1, \cdots, L\}$ for the generalized Lorenz ordinate, and the following cross-validation estimate of the MSE:

$$CV_c = \frac{1}{L}\sum_{l=1}^{L}[\hat{T}_{n,c}^{(1,l)}(t) - \hat{\theta}^{(2,l)}(t)]^2.$$

Then, c is chosen as the constant that minimizes CV_c.

Alternatively, if we focus on the overall performance of the smoothed estimator across all t, we can use a similar cross-validation procedure for selecting c by minimizing the Average Mean Squared Error (AMSE):

$$AMSE(c) = E\left\{\frac{1}{J}\sum_{j=1}^{J}[\hat{T}_{n,c}(t_j) - \theta(t_j)]^2\right\},\ j = 1, 2, ..., J,$$

where t_j is in a fine grid of (0,1), and J is the number of grid points.

And the cross-validation estimate of the AMSE is

$$ACV_c = \frac{1}{L}\frac{1}{J}\sum_{l=1}^{L}\sum_{j=1}^{J}[\hat{T}_{n,c}^{(1,l)}(t_j) - \hat{\theta}^{(2,l)}(t_j)]^2.$$

Similarly, c is chosen as the constant that minimizes ACV_c.

3. THE SMOOTHED JACKKNIFE EMPIRICAL LIKELIHOOD FOR THE GLC

Using the smoothed estimator for the GLC in the previous section, we can define a smoothed jackknife empirical likelihood for the GLC. Based on $\hat{T}_n(t)$, we define the jackknife pseudo-values as

$$\hat{V}_i(t) = n\hat{T}_n(t) - (n-1)\hat{T}_{n-1,i}(t), \tag{3}$$

where $\hat{T}_{n-1,i}(t) = \frac{1}{(n-1)}\sum_{j\neq i} X_j K(\frac{t - F_{n,i}(X_j)}{h})$ is the given statistics $\hat{T}_{n-1}(t)$ computed on $n-1$ observations $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$, and $F_{n,i}(t) = \frac{1}{n-1}\sum_{j\neq i} I(X_j \leq t)$ is the empirical distribution function based on the $n-1$ observations.

Then, the jackknife empirical likelihood for $\theta = \theta(t)$ can be defined as follows:

$$L(t,\theta) = \sup_{\mathbf{p}}\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i\hat{V}_i(t) = \theta\right\}, \tag{4}$$

where $\mathbf{p} = (p_1, \cdots, p_n)$ is a probability vector.

By using the Lagrange multiplier method, we obtain the maximization for (4) at

$$p_i = \frac{1}{n}\{1 + \lambda[\hat{V}_i(t) - \theta]\}^{-1}, \tag{5}$$

where $\lambda = \lambda(t,\theta)$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{V}_i(t) - \theta}{1 + \lambda(\hat{V}_i(t) - \theta)} = 0. \tag{6}$$

Therefore, the jackknife empirical likelihood ratio for θ can be defined as

$$L_n(\theta) = \prod_{i=1}^{n}(np_i) = \prod_{i=1}^{n}\{1 + \lambda(\hat{V}_i(t) - \theta)\}^{-1}, \tag{7}$$

which gives the log empirical likelihood ratio as

$$l_n(\theta) = -2\log L_n(\theta) = 2\sum_{i=1}^{n}\log\{1 + \lambda(\hat{V}_i(t) - \theta)\} \tag{8}$$

Based on Tukey (1958), the pseudo-values $\hat{V}_i(t)$, $i = 1, \cdots, n$ could be treated as though they were i.i.d, and $\hat{V}_i(t)$'s have approximately the same variance as $\sqrt{n}\hat{T}_n(t)$. Therefore, the variance of $\sqrt{n}\hat{T}_n(t)$, denoted as $var(\sqrt{n}\hat{T}_n(t))$, can be estimated by the sample variance of $\hat{V}_1(t), \cdots, \hat{V}_n(t)$. The jackknife variance estimator of $\hat{T}_n(t)$ is thus defined as follows:

$$\begin{aligned} \upsilon_{JACK}(t) &= \frac{1}{n(n-1)}\sum_{i=1}^{n}[\hat{V}_i(t) - \frac{1}{n}\sum_{j=1}^{n}\hat{V}_j(t)]^2 \\ &= \frac{n-1}{n}\sum_{i=1}^{n}[\hat{T}_{n-1,i}(t) - \frac{1}{n}\sum_{j=1}^{n}\hat{T}_{n-1,j}(t)]^2. \end{aligned}$$

The following theorem shows that this jackknife variance estimator is a consistent estimator for the asymptotic variance $\sigma^2(t)$.

Theorem 3.1. *Under conditions of Theorem 2.1, as $n \to \infty$, we have*

$$\upsilon_{JACK}(t) \xrightarrow{p} \sigma^2(t), \tag{9}$$

where $\sigma^2(t)$ is defined in Theorem 2.1.

The Wilks theorem for $l_n(\theta)$ is obtained in the following theorem.

Theorem 3.2. *Under the conditions of Theorem 2.1, as $n \to \infty$, we have*

$$l_n(\theta) \xrightarrow{d} \chi^2(1). \tag{10}$$

4. CONFIDENCE INTERVALS FOR THE GENERALIZED LORENZ CURVE

4.1 Normal approximation-based confidence intervals

One of the most popular methods to construct a confidence interval for an unknown parameter is normal approximation. To construct a normal approximation-based confidence interval for the generalized Lorenz ordinate θ, we first need to obtain an appropriate estimator for θ, and then derive its asymptotic normal distribution.

Two estimators for θ are used to build confidence intervals for θ in our study. First of all, the empirical estimate $\hat{\theta} = \hat{\theta}(t)$ for θ is asymptotically normal with variances σ_v^2, i.e., $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma_v^2)$, where $\sigma_v^2 = \int_0^{\xi_t} (x - \xi_t)^2 dF(x) - (\theta - t\xi_t)^2$.

Therefore, a $(1-\alpha)$ level normal approximation (NA1)-based confidence interval for θ can be constructed as

$$(l_1, u_1) = (\hat{\theta} - \frac{z_{1-\frac{\alpha}{2}} \hat{\sigma}_v}{\sqrt{n}}, \hat{\theta} + \frac{z_{1-\frac{\alpha}{2}} \hat{\sigma}_v}{\sqrt{n}}),$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2}) - th$ quantile of the standard normal distribution, and $\hat{\sigma}_v^2 = \int_0^{\hat{\xi}_t} (x - \hat{\xi}_t)^2 dF_n(x) - (\hat{\theta} - t\hat{\xi}_t)^2$ is a consistent estimate for σ_v^2.

Based on Theorem 3.1, we have that $\upsilon_{JACK}(t) \xrightarrow{p} \sigma^2(t)$. So the jackknife variance estimator $\upsilon_{JACK}(t)$ is a consistent estimator for $\sigma^2(t)$. Thus, the second $(1-\alpha)$ level normal approximation (NA2)-based confidence interval for θ can be constructed as

$$(l_2, u_2) = \left(\hat{T}_n(t) - \frac{z_{1-\frac{\alpha}{2}} \sqrt{\upsilon_{JACK}(t)}}{\sqrt{n}}, \hat{T}_n(t) + \frac{z_{1-\frac{\alpha}{2}} \sqrt{\upsilon_{JACK}(t)}}{\sqrt{n}} \right)$$

4.2 Bootstrap-based confidence intervals

Normal approximation-based intervals may have poor performance when the income data is skewed or has outliers. Bootstrap is a powerful non-parametric approach to make statistical inferences when the asymptotic variance of an estimator is unknown and in a complex form. In this section, we apply bootstrap methods to construct confidence intervals for the GLC.

By drawing bootstrap resample $\{X_1^*, X_2^*, ..., X_n^*\}$ with replacement from the original sample $\{X_1, X_2, ..., X_n\}$, the bootstrap version of the empirical estimator $\hat{\theta}(t)$ for the generalized Lorenz ordinate can be defined as

$$\hat{\theta}^*(t) = \frac{1}{n} \sum_{i=1}^{n} X_i^* I(X_i^* \le \xi_t^*).$$

We repeat this bootstrap procedure for B ($B \ge 500$ is recommended) times. Thus, B bootstrap copies of $\hat{\theta}$ are obtained, denoted then as $\{\hat{\theta}_b^*, b = 1, 2, ..., B\}$. Then, the bootstrap sample variance of $\hat{\theta}_b^*$'s

$$V_G^* = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\theta}^*)^2,$$

where $\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^*$, is used to estimate the asymptotic variance of $\hat{\theta}$.

Two bootstrap confidence intervals based on the empirical estimator for θ are constructed as follows:

1. BT1 interval:

$$(l_3, u_3) = (\hat{\theta} - z_{1-\alpha/2} \sqrt{V_G^*}, \hat{\theta} + z_{1-\alpha/2} \sqrt{V_G^*}).$$

2. BT2 interval:

$$(l_4, u_4) = (\bar{\theta}^* - z_{1-\alpha/2} \sqrt{V_G^*}, \bar{\theta}^* + z_{1-\alpha/2} \sqrt{V_G^*}).$$

We can also apply the bootstrap bias correction and acceleration (BCa1) method to construct a confidence interval for θ, which does not need a variance estimate.

3. BCa1 interval:

$$(l_5, u_5) = (\hat{\theta}^*_{([B\beta_1])}, \hat{\theta}^*_{([B\beta_2])}).$$

where

$$\beta_1 = \Phi(b + \frac{b + z_{\alpha/2}}{1 - a(b + z_{\alpha/2})}), \beta_2 = \Phi(b + \frac{b + z_{1-\alpha/2}}{1 - a(b + z_{1-\alpha/2})})$$

with correction constants a and b defined by

$$a = \frac{1}{6} \sum_{i=1}^{n} \varphi_i^3 / (\sum_{i=1}^{n} \varphi_i^2)^{\frac{3}{2}}, b = \Phi^{-1}(\frac{1}{B} \sum_{b=1}^{B} I(\hat{\theta}_b^* \le \hat{\theta})),$$

where $\varphi_i = \hat{\theta}_{(.)} - \hat{\theta}_{(-i)}$, and $\hat{\theta}_{(-i)}$ is the $\hat{\theta}$ computed by deleting the i-th observation in original data, and $\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(-i)}$.

Similarly, confidence intervals for θ can also be built based on the kernel estimator $\hat{T}_n = \hat{T}_n(t)$. Let $\hat{T}^*(t) = \frac{1}{n} \sum_{i=1}^{n} X_i^* K(\frac{t - F_n(X_i^*)}{h})$ be the bootstrap version of $\hat{T}_n$, and $\{\hat{T}_b^*, b = 1, 2, ..., B\}$ be B bootstrap copies of $\hat{T}_n$. The sample variance of $\hat{T}_b^*$'s

$$V_{GT}^* = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{T}_b^* - \bar{T}^*)^2,$$

where $\bar{T}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{T}_b^*$, can be used to estimate the asymptotic variance of $\hat{T}_n$.

Thus, three new bootstrap confidence intervals for θ are constructed as follows:

4. BT3 interval:

$$(l_6, u_6) = (\hat{T}_n - z_{1-\alpha/2} \sqrt{V_{GT}^*}, \hat{T}_n + z_{1-\alpha/2} \sqrt{V_{GT}^*}),$$

5. BT4 interval:

$$(l_7, u_7) = (\bar{T}^* - z_{1-\alpha/2} \sqrt{V_{GT}^*}, \bar{T}^* + z_{1-\alpha/2} \sqrt{V_{GT}^*}).$$

6. BCa2 interval:

$$(l_8, u_8) = (\hat{T}^*_{([B\beta_1])}, \hat{T}^*_{([B\beta_2])}).$$

where

$$\beta_1 = \Phi(b + \frac{b + z_{\alpha/2}}{1 - a(b + z_{\alpha/2})}), \beta_2 = \Phi(b + \frac{b + z_{1-\alpha/2}}{1 - a(b + z_{1-\alpha/2})})$$

with correction constants a and b defined by

$$a = \frac{1}{6}\sum_{i=1}^{n}\varphi_i^3/(\sum_{i=1}^{n}\varphi_i^2)^{\frac{3}{2}}, b = \Phi^{-1}(\frac{1}{B}\sum_{b=1}^{B} I(\hat{T}_b^* \le \hat{T}_n))$$

where $\varphi_i = \hat{T}_{(.)} - \hat{T}_{(-i)}$, and $\hat{T}_{(-i)}$ is the $\hat{T}_n$ computed by deleting the i-th observation in original data, and $\hat{T}_{(.)} = \frac{1}{n}\sum_{i=1}^{n}\hat{T}_{(-i)}$.

4.3 The smoothed JEL-based confidence interval

The *Smoothed Jackknife Empirical Likelihood* (SJEL) for the GLC θ is derived in Section 3. Based on Theorem 3.2, the SJEL-based confidence interval for θ can be constructed as follows:

$$(l_e, u_e) = \{\theta : l_n(\theta) \le \chi^2_{1,1-\alpha}\}.$$

where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$-th quantile of χ^2_1.

5. SIMULATION STUDIES AND A REAL EXAMPLE

5.1 Simulation studies

5.1.1 Point estimator evaluation

We first conduct simulation studies to compare finite sample performances of the empirical estimator $\hat{\theta}(t)$ and the kernel estimator $\hat{T}_n(t)$ for θ in terms of Mean Square Error (MSE) and bias when t ranges from 0.2 to 0.8. The Quartic/Triweight kernel density function $\omega(y) = \frac{35}{32}(1 - y^2)^2 I(|y| \le 1)$ is selected for the kernel estimator of the GLC, and the bandwidth $cn^{-1/4}(\log\log n)^{3/5}$ is chosen via the proposed cross-validation method, where c is valued differently based on different t. For each setting, 1,000 random samples are generated from the Weibull distribution with shape parameter = 1 and scale parameter = 2, and the sample sizes are chosen to be 100, 200, and 500 respectively.

Table 1 presents the comparison results, where $Bias_{\hat{\theta}}$ represents the bias of the empirical estimator $\hat{\theta}(t)$, and $Bias_{\hat{T}_n(t)}$ represents the bias of the kernel estimator $\hat{T}_n(t)$. From Table 1, we observe that the MSE of the smoothed estimator is less than that of the empirical estimator, and the bias of the smoothed estimator is smaller than that of the empirical estimator in most cases. Based on this table, we found the kernel estimator $\hat{T}_n(t)$ has better MSE performance than the empirical estimator $\hat{\theta}(t)$.

Table 1. Bias and MSE of the empirical estimator and the kernel estimator for GLCs with F = Weibull distribution

Sample Size	t	$Bias_{\hat{\theta}}$	$Bias_{\hat{T}_n(t)}$	$MSE_{\hat{\theta}}$	$MSE_{\hat{T}_n(t)}$
100	0.2	0.002136	0.000119	0.000134	0.000119
	0.3	0.003137	0.000168	0.000470	0.000437
	0.4	0.004103	0.000701	0.001180	0.001117
	0.5	0.005396	0.001185	0.002447	0.002340
	0.6	0.006597	0.002087	0.004525	0.004358
	0.7	0.008415	0.003024	0.007714	0.007459
	0.8	0.009610	0.005467	0.012748	0.012377
200	0.2	0.000719	0.000212	0.000065	0.000062
	0.3	0.000843	0.000725	0.000228	0.000222
	0.4	0.001254	0.001060	0.000573	0.000561
	0.5	0.001489	0.001677	0.001199	0.001180
	0.6	0.001690	0.002520	0.002224	0.002196
	0.7	0.002011	0.003505	0.003876	0.003833
	0.8	0.002629	0.004683	0.006522	0.006456
500	0.2	0.000378	0.000058	0.000023	0.000023
	0.3	0.000633	0.000065	0.000085	0.000084
	0.4	0.000972	0.000120	0.000213	0.000211
	0.5	0.001289	0.000106	0.000443	0.000438
	0.6	0.001583	0.000005	0.000831	0.000824
	0.7	0.001675	0.000387	0.001455	0.001445
	0.8	0.001983	0.000724	0.002486	0.002470

5.1.2 Interval estimation evaluation

After the evaluation of point estimators, we will evaluate the coverage probabilities and interval lengths of the normal approximation-based confidence intervals, the bootstrap-based confidence intervals and the SJEL-based confidence interval by simulation studies.

We again generate 1,000 random samples from the Weibull distribution with shape parameter = 1 and scale parameter = 2. The sample sizes are chosen to be 100, 200, and 500, respectively. We calculate various confidence intervals at 95% confidence level with t = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. For the bootstrap variance estimates, 500 bootstrap re-samples are drawn from the original sample generated from the Weibull distribution.

The coverage probabilities and average lengths of the 95% confidence intervals for generalized Lorenz ordinates are presented in Tables 2–3. Based on these tables, we observe that the normal approximation-based NA1 and NA2 intervals have severe under-coverage problems when $t = 10\%, 20\%$ (located in the low tail region of the GLC, which are of more interest in practice). By contrast, the proposed SJEL intervals and BT1, BT2, BCa1 intervals perform much better than the other intervals. When $t \ge 30\%$, the proposed SJEL intervals, the normal approximation-based NA1 and NA2 intervals, and the bootstrap-based BT1, BT2, BT3, BT4 and BCa1 intervals have similar coverage probabilities although SJEL intervals perform slightly better than other intervals in most cases considered here. Overall, we recommend the SJEL-based interval for the GLC, particularly when t falls in the low tail region of the curve.

Table 2. F = Weibull distribution: Coverage probabilities and interval lengths of 95% level NA1, NA2, BT1, BT2, BT3, BT4, BCa1, BCa2 and SJEL intervals for generalized Lorenz ordinates when t ranges from 0.1 to 0.4

		$t = 10\%$		$t = 20\%$		$t = 30\%$		$t = 40\%$	
n	Method	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
100	NA1	0.824	0.0158	0.935	0.0435	0.942	0.0821	0.955	0.1301
	NA2	0.867	0.0171	0.938	0.0442	0.951	0.0820	0.951	0.1293
	BT1	0.972	0.0186	0.964	0.0488	0.957	0.0888	0.960	0.1397
	BT2	0.957	0.0186	0.963	0.0488	0.964	0.0888	0.955	0.1397
	BT3	0.851	0.0164	0.938	0.0423	0.938	0.0793	0.945	0.1262
	BT4	0.871	0.0164	0.945	0.0423	0.944	0.0793	0.95	0.1262
	BCa1	0.939	0.0144	0.955	0.0434	0.955	0.0821	0.957	0.1317
	BCa2	0.861	0.0157	0.903	0.0383	0.919	0.0766	0.928	0.1247
	SJEL	0.937	0.0170	0.959	0.0414	0.950	0.0818	0.951	0.1256
200	NA1	0.876	0.0108	0.934	0.0306	0.946	0.0575	0.953	0.0911
	NA2	0.885	0.0116	0.925	0.0308	0.947	0.0575	0.947	0.0912
	BT1	0.919	0.0118	0.944	0.0322	0.952	0.0599	0.954	0.0947
	BT2	0.918	0.0118	0.939	0.0322	0.946	0.0599	0.946	0.0947
	BT3	0.862	0.0114	0.931	0.0301	0.949	0.0565	0.949	0.0898
	BT4	0.851	0.0114	0.94	0.0301	0.951	0.0565	0.954	0.0898
	BCa1	0.947	0.0104	0.937	0.0304	0.944	0.0577	0.947	0.0919
	BCa2	0.863	0.0093	0.905	0.0272	0.934	0.0544	0.937	0.0881
	SJEL	0.939	0.0115	0.946	0.0300	0.950	0.0572	0.952	0.0895
500	NA1	0.850	0.0067	0.922	0.0192	0.940	0.0361	0.947	0.0575
	NA2	0.880	0.0071	0.912	0.0194	0.942	0.0364	0.943	0.0577
	BT1	0.945	0.0069	0.943	0.0196	0.933	0.0367	0.936	0.0587
	BT2	0.940	0.0069	0.936	0.0196	0.939	0.0367	0.933	0.0587
	BT3	0.883	0.0070	0.920	0.0191	0.939	0.0359	0.953	0.0571
	BT4	0.888	0.0070	0.922	0.0191	0.939	0.0359	0.957	0.0571
	BCa1	0.937	0.0066	0.930	0.0192	0.929	0.0362	0.933	0.0581
	BCa2	0.839	0.0057	0.909	0.0172	0.926	0.0347	0.929	0.0560
	SJEL	0.925	0.0070	0.937	0.0203	0.946	0.0362	0.950	0.0564

5.2 Georgia public university employee income data example

Georgia Department of Audits and Accounts provides an open resource for annually-updated salary information of Georgia public institute employees. It includes individual demographic information such as name, title, salary and travel reimbursement for each employee in public sectors. This public resource is aimed at strengthening the transparency of the Georgia government. Our study will focus on the income distribution of professors in Georgia public colleges and universities. We only retain annual income data for those who work with a full-time schedule, since a part-time employee and temporary employee's salary information does not satisfy our annual salary definition.

To create a relatively homogeneous income group, we limit our analysis to the income of full-time assistant professors, associate professors, and full professors from Units of University System and Georgia Military College in the 2012 fiscal year. We initially observed that a few individuals have abnormally low wages, which may be due to the following reasons: some of the 2012 newly-hired professors did not have working records for the whole 2012 fiscal year; part-time professors possibly either took leave or transferred to another organization during the 2012 fiscal year. To filter out these subjects, we first excluded professors who didn't have salary record in the 2011 fiscal year. Then we dropped out those with 2012 fiscal year income significantly lower than 2011 fiscal year income. Individuals with salaries less than $20,000 were also removed. Finally, there remained 5,921 observations in the analysis. Table 4 displays various 95% level confidence intervals for generalized Lorenz ordinates. Based on our simulation studies, we would use SJEL intervals for generalized Lorenz ordinates. From Table 4, we can see that the least wealthy 80% professors have an average annual salary from $60,973.24 to $62,282.91.

6. DISCUSSION

In this paper, a kernel smoothing estimator has been proposed for the generalized Lorenz curve. The new estimator has better finite sample performance than the traditional empirical estimator in terms of mean squared error. Meanwhile, a SJEL method has been developed for inferences on the generalized Lorenz curve. The proposed SJEL-based method has the property that the Wilks theorem still holds for the proposed jackknife empirical likelihood ratio statistic. In practice, the tail regions of the generalized Lorenz

Table 3. F = Weibull distribution: Coverage probabilities and interval lengths of 95% level NA1, NA2, BT1, BT2, BT3, BT4, BCa1, BCa2 and SJEL intervals for generalized Lorenz ordinates when t ranges t from 0.5 to 0.9

		$t=50\%$		$t=60\%$		$t=70\%$		$t=80\%$		$t=90\%$	
n	Method	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
100	NA1	0.953	0.1881	0.939	0.2577	0.947	0.3400	0.948	0.4419	0.951	0.5696
	NA2	0.944	0.1864	0.943	0.2542	0.950	0.3352	0.950	0.4349	0.952	0.5619
	BT1	0.962	0.1995	0.963	0.2724	0.960	0.3604	0.961	0.4720	0.965	0.6215
	BT2	0.961	0.1995	0.962	0.2724	0.959	0.3604	0.965	0.4720	0.967	0.6215
	BT3	0.944	0.1832	0.939	0.2509	0.942	0.3313	0.946	0.4298	0.945	0.5492
	BT4	0.949	0.1832	0.944	0.2509	0.940	0.3313	0.947	0.4298	0.942	0.5492
	BCa1	0.956	0.1897	0.956	0.2617	0.957	0.3490	0.952	0.4571	0.962	0.6020
	BCa2	0.934	0.1824	0.928	0.2507	0.929	0.3271	0.929	0.4268	0.933	0.5300
	SJEL	0.953	0.1849	0.955	0.2510	0.950	0.3305	0.953	0.4275	0.951	0.5698
200	NA1	0.953	0.1320	0.945	0.1810	0.949	0.2390	0.946	0.3103	0.946	0.4019
	NA2	0.939	0.1320	0.938	0.1816	0.941	0.2410	0.940	0.3125	0.936	0.4034
	BT1	0.951	0.1361	0.957	0.1865	0.956	0.2462	0.953	0.3203	0.960	0.4188
	BT2	0.949	0.1361	0.955	0.1865	0.952	0.2462	0.954	0.3203	0.959	0.4188
	BT3	0.947	0.1302	0.951	0.1785	0.947	0.2364	0.941	0.3058	0.940	0.3951
	BT4	0.950	0.1302	0.950	0.1785	0.947	0.2364	0.938	0.3058	0.944	0.3951
	BCa1	0.946	0.1329	0.951	0.1828	0.958	0.2420	0.946	0.3158	0.955	0.4124
	BCa2	0.940	0.1288	0.937	0.1769	0.938	0.2353	0.938	0.3035	0.926	0.3885
	SJEL	0.946	0.1318	0.950	0.1802	0.946	0.2360	0.949	0.3044	0.946	0.4005
500	NA1	0.942	0.0834	0.936	0.1146	0.936	0.1520	0.939	0.1972	0.933	0.2551
	NA2	0.945	0.0836	0.948	0.1147	0.946	0.1518	0.953	0.1970	0.945	0.2555
	BT1	0.932	0.0847	0.938	0.1161	0.938	0.1541	0.942	0.1996	0.941	0.2595
	BT2	0.942	0.0847	0.945	0.1161	0.943	0.1541	0.941	0.1996	0.943	0.2595
	BT3	0.939	0.0830	0.935	0.1139	0.935	0.1511	0.934	0.1962	0.927	0.2536
	BT4	0.940	0.0830	0.935	0.1139	0.935	0.1511	0.936	0.1962	0.931	0.2536
	BCa1	0.930	0.0839	0.930	0.1152	0.938	0.1531	0.937	0.1985	0.937	0.2579
	BCa2	0.930	0.0818	0.925	0.1125	0.929	0.1497	0.933	0.1940	0.916	0.2478
	SJEL	0.939	0.0845	0.942	0.1155	0.942	0.1508	0.945	0.1958	0.948	0.2559

curve are of great interest in the study of income distribution. Our simulation studies also indicate that the proposed SJEL-based interval performs better than other intervals in most cases considered in this paper, particularly when t falls in the low tail region of the curve. While the bootstrap-based BT1, BT2, BT3, BT4 and BCa1 intervals could have good coverage probabilities, they are computationally expensive, particularly when sample sizes get larger. The proposed SJEL-based method combines the power of both jackknife and empirical likelihood methods. It can be directly calculated by implementing the algorithm for computing the standard empirical likelihood interval (Hall and La Scala, 1990). Another interesting research topic is the overall properties of the smoothing estimator $\hat{T}_n(t)$ and the SJEL method, we will study them elsewhere.

APPENDIX: PROOFS

Proof of Theorem 2.1. We have the following decomposition

(11)

$$
\begin{aligned}
&\sqrt{n}[\hat{T}_n(t)-\theta(t)]\\
&= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i K(\frac{t-F_n(X_i)}{h}) - \int_0^{\xi_t} x dF(x)\right]\\
&= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i K(\frac{t-F_n(X_i)}{h}) - \frac{1}{n}\sum_{i=1}^{n} X_i K(\frac{t-F(X_i)}{h})\right]\\
&\quad + \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i K(\frac{t-F(X_i)}{h}) - \int_0^{\xi_t} x dF(x)\right]\\
&\equiv I_1 + I_2.
\end{aligned}
$$

Term I_1 of (11) can be written as

(12)

$$
\begin{aligned}
I_1 &= \frac{\sqrt{n}}{n}\sum_{i=1}^{n} X_i\left[K(\frac{t-F_n(X_i)}{h}) - K(\frac{t-F(X_i)}{h})\right]\\
&= \int_{-\infty}^{\infty} x\left[K(\frac{t-F_n(x)}{h})\right.\\
&\quad \left. - K(\frac{t-F(x)}{h})\right] d[\sqrt{n}(F_n(x)-F(x))]\\
&\quad + \sqrt{n}\int_{-\infty}^{\infty} x\left[K(\frac{t-F_n(x)}{h}) - K(\frac{t-F(x)}{h})\right] dF(x)\\
&\equiv I_{11} + I_{12}.
\end{aligned}
$$

Table 4. Georgia professor's real income data example: 95% *NA1, NA2, BT1, BT2, BT3, BT4, BCa1, BCa2, SJEL intervals for GLC when* $t = 0.5, 0.6, 0.7, 0.8$

t	Method	Confidence Interval	Length
0.5	NA1	(31698.77, 32310.15)	611.38
	NA2	(31684.31, 32299.43)	615.12
	BT1	(31694.22, 32314.70)	620.48
	BT2	(31695.82, 32316.30)	620.48
	BT3	(31696.16, 32287.59)	591.43
	BT4	(31690.49, 32281.93)	591.43
	BCa1	(31664.83, 32306.79)	641.96
	BCa2	(31686.21, 32306.68)	641.96
	SJEL	(31717.33, 32332.47)	615.14
0.6	NA1	(40268.37, 41044.96)	776.59
	NA2	(40249.51, 41045.61)	796.10
	BT1	(40274.97, 41038.37)	763.40
	BT2	(40281.74, 41045.14)	763.40
	BT3	(40244.99, 41050.12)	805.13
	BT4	(40230.78, 41035.92)	805.13
	BCa1	(40271.12, 41030.95)	759.83
	BCa2	(40272.49, 41105.52)	759.83
	SJEL	(40257.26, 41023.32)	766.06
0.7	NA1	(49906.65, 50875.11)	968.46
	NA2	(49878.86, 50886.52)	1007.6
	BT1	(49913.24, 50868.52)	955.28
	BT2	(49904.21, 50859.49)	955.28
	BT3	(49899.60, 50865.78)	966.18
	BT4	(49883.95, 50850.13)	966.18
	BCa1	(49956.38, 50911.49)	955.11
	BCa2	(49955.02, 51020.14)	1065.1
	SJEL	(49989.57, 50924.68)	935.11
0.8	NA1	(60921.64, 62284.70)	1363.06
	NA2	(60941.11, 62250.93)	1309.82
	BT1	(60936.02, 62270.33)	1334.31
	BT2	(60969.15, 62303.46)	1334.31
	BT3	(60961.48, 62230.56)	1269.08
	BT4	(60932.65, 62201.74)	1269.08
	BCa1	(60857.96, 62272.20)	1414.24
	BCa2	(60979.48, 62249.00)	1269.52
	SJEL	(60973.24, 62282.91)	1309.67

By Taylor expansion, under conditions in Theorem 2.1, I_{12} of (12) can be written as

(13)
$$\begin{aligned}
I_{12} &= \sqrt{n}\int_{-\infty}^{\infty} x\left[K(\frac{t-F_n(x)}{h}) - K(\frac{t-F(x)}{h})\right]dF(x)\\
&= \sqrt{n}\int_{-\infty}^{\infty} x\left[\omega(\frac{t-F(x)}{h})\frac{F(x)-F_n(x)}{h}\right.\\
&\quad \left.+\frac{1}{2}\omega'(\frac{t-F(x)}{h})(\frac{F(x)-F_n(x)}{h})^2\right]dF(x) + o_p(1)\\
&= -\int_{-\infty}^{\infty} x\omega(\frac{t-F(x)}{h})\frac{\sqrt{n}(F_n(x)-F(x))}{h}dF(x)\\
&\quad + O_p(\frac{1}{\sqrt{n}h^2}).
\end{aligned}$$

Let $y = F(x)$, $Y_i = F(X_i)$ and $U_n(y) = \sqrt{n}[\frac{1}{n}\sum_{i=1}^n I(Y_i \le y) - y]$. Then, Y_i's follow uniform[0,1] distribution. Since $h \to 0$, the support of $\omega(\cdot)$ is bounded, (13) will be equal to

$$\begin{aligned}
(14)\quad I_{12} &= -\int_{-1}^{1} F^{-1}(y)\omega(\frac{t-y}{h})\frac{1}{h}U_n(y)dy + o_p(1)\\
&= \int_{\frac{t-1}{h}}^{\frac{t+1}{h}} F^{-1}(t-hz)\omega(z)U_n(t-hz)dz + o_p(1)\\
&= \xi_t U_n(t) + o_p(1).
\end{aligned}$$

Since $\sqrt{n}[F_n(x) - F(x)] \xrightarrow{d} B(x)$, which is Gaussian process, and $\sqrt{n}h^2 \to \infty$, so $I_{11} = o_p(1)$. Therefore, $I_1 = \xi_t U_n(t) + o_p(1)$. Next, let's consider I_2 of (11). Notice that

$$\begin{aligned}
(15)\quad &\lim_{h\to 0} E\left[XK(\frac{t-F(X)}{h})\right]\\
&= \lim_{h\to 0}\int_{-\infty}^{\infty} xK(\frac{t-F(x)}{h})f(x)dx\\
&= \int_{-\infty}^{\infty} x[\lim_{h\to 0}\int_{-\infty}^{\frac{t-F(x)}{h}}\omega(y)dy]f(x)dx\\
&= \int_{-\infty}^{\infty} xI[t > F(x)]dF(x) = \int_0^{\xi_t} xdF(x) = \theta(t).
\end{aligned}$$

Similarly, we have that

$$\begin{aligned}
(16)\quad &\lim_{h\to 0} E\left[X^2K^2(\frac{t-F(X)}{h})\right]\\
&= \lim_{h\to 0}\int_{-\infty}^{\infty} x^2K^2(\frac{t-F(x)}{h})f(x)dx\\
&= \int_{-\infty}^{\infty} x^2[\lim_{h\to 0}\int_{-\infty}^{\frac{t-F(x)}{h}}\omega(y)dy]^2 f(x)dx\\
&= \int_{-\infty}^{\infty} x^2I[t > F(x)]dF(x) = \int_0^{\xi_t} x^2dF(x).
\end{aligned}$$

Let $W = XK(\frac{t-F(X)}{h})$ and $W_i = X_iK(\frac{t-F(X_i)}{h})$. Then, I_2 of (11) can be rewritten as

$$\begin{aligned}
(17)\quad I_2 &= \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^n X_iK(\frac{t-F(X_i)}{h})\right.\\
&\quad \left.- E\left[XK(\frac{t-F(X)}{h})\right]\right\} + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n (W_i - EW_i) + o_p(1).
\end{aligned}$$

Let $\{U_1, U_2, \cdots, U_n\}$ be an i.i.d. sample from $U(0,1)$ (uniform distribution on [0,1]) and independent of $\{X_1, X_2, \cdots, X_n\}$. Since $Y_i = F(X_i) \overset{i.i.d.}{\sim} U(0,1)$ for any continuous distribution function F, then

$$(18)\quad \begin{aligned} I_1 &= \xi_t\sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} I(Y_i \le t) - t] + o_p(1) \\ &\stackrel{d}{=} \xi_t\sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} I(U_i \le t) - t], \end{aligned}$$

where $\stackrel{d}{=}$ means that two statistics asymptotically have the same distribution. Therefore,

$$(19)\quad I_1 + I_2 \stackrel{d}{=} \xi_t\sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} I(U_i \le t) - t] + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i - EW_i).$$

Note that $Var(W_i) = E[X^2K^2(\frac{t-F(X)}{h})] - [E(XK(\frac{t-F(X)}{h}))]^2 = \int_0^{\xi_t} x^2 dF(x) - \theta^2(t) + o(1)$. Since U_i's are independent of X_i's, U_i's are thus independent of W_i's, and the two terms in (19) are independent. By central limit theorem, the fist term in (19) follows $N(0, \xi_t^2 t(1-t))$ asymptotically, and the second term in (19) follows $N(0, \int_0^{\xi_t} x^2 dF(x) - \theta^2(t))$ asymptotically, hence

$$(20)\quad I_1 + I_2 \xrightarrow{d} N(0, \sigma^2(t)),$$

where $\sigma^2(t) = \xi_t^2 t(1-t) + \int_0^{\xi_t} x^2 dF(x) - \theta^2(t)$. The proof of Theorem 2.1 is complete. □

We need Lemma 1 and Lemma 2 to prove Theorem 3.1.

Lemma 1. Under the conditions in Theorem 2.1, we have

$$(21)\quad \sqrt{n}\left[\frac{1}{n}\sum_{k=1}^{n}\hat{V}_k(t) - \theta(t)\right] \xrightarrow{d} N(0, \sigma^2(t)),$$

where $\sigma^2(t)$ is defined in Theorem 2.1.

Proof. Note that $\frac{1}{n}\sum_{k=1}^{n}\hat{V}_k(t)$ can be decomposed into

$$(22)\quad \frac{1}{n}\sum_{k=1}^{n}\hat{V}_k(t) = \frac{n-1}{n}\sum_{k=1}^{n}(\hat{T}_n(t) - \hat{T}_{n-1,k}(t)) + \hat{T}_n(t),$$

while

(23)

$$\begin{aligned} &\hat{T}_n(t) - \hat{T}_{n-1,k}(t) \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i K(\frac{t-F_n(X_i)}{h}) \\ &\quad - \frac{1}{(n-1)}\sum_{j\ne k} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i K(\frac{t-F_n(X_i)}{h}) - \frac{1}{n}\sum_{j\ne k}^{n} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \\ &\quad + \frac{1}{n}\sum_{j\ne k}^{n} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \\ &\quad - \frac{1}{(n-1)}\sum_{j\ne k} X_j K(\frac{t-F_{n,k}(X_j)}{h}), \end{aligned}$$

and

(24)

$$\begin{aligned} &\sum_{k=1}^{n}(\hat{T}_n(t) - \hat{T}_{n-1,k}(t)) \\ &= \frac{1}{n}\{\sum_{k=1}^{n}\sum_{i=1}^{n} X_i K(\frac{t-F_n(X_i)}{h}) \\ &\quad - \sum_{k=1}^{n}\sum_{j\ne k} X_j K(\frac{t-F_{n,k}(X_j)}{h})\} \\ &\quad + \sum_{k=1}^{n}\{\frac{1}{n}\sum_{j\ne k} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \\ &\quad - \frac{1}{n-1}\sum_{j\ne k} X_j K(\frac{t-F_{n,k}(X_j)}{h})\} \\ &= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} X_i[K(\frac{t-F_n(X_i)}{h}) - K(\frac{t-F_{n,k}(X_i)}{h})] \\ &\quad + \frac{1}{n}\sum_{k=1}^{n} X_k K(\frac{t-F_{n,k}(X_k)}{h}) \\ &\quad + \frac{1}{n}\sum_{k=1}^{n}\sum_{j\ne k} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \\ &\quad - \frac{1}{n-1}\sum_{k=1}^{n}\sum_{j\ne k} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \\ &= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} X_i[K(\frac{t-F_n(X_i)}{h}) - K(\frac{t-F_{n,k}(X_i)}{h})] \\ &\quad + \left[\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{n} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \right. \\ &\quad - \frac{1}{n-1}\sum_{k=1}^{n}\sum_{j=1}^{n} X_j K(\frac{t-F_{n,k}(X_j)}{h}) \\ &\quad \left. + \frac{1}{n-1}\sum_{k=1}^{n} X_k K(\frac{t-F_{n,k}(X_k)}{h})\right] \\ &\equiv H_1 + H_2. \end{aligned}$$

Using Taylor expansion, H_1 of (24) can be written as

(25)

$$\begin{aligned} H_1 &= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} X_i[K(\frac{t-F_n(X_i)}{h}) - K(\frac{t-F_{n,k}(X_i)}{h})] \\ &= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} X_i\{\omega(\frac{t-F_n(X_i)}{h})\frac{F_{n,k}(X_i) - F_n(X_i)}{h} \end{aligned}$$

$$- \frac{1}{2}\omega^{'}(\frac{t-\xi_{n,k,i}}{h})(\frac{F_{n,k}(X_i)-F_n(X_i)}{h})^2\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i\{\omega(\frac{t-F_n(X_i)}{h})\sum_{k=1}^{n}\frac{F_{n,k}(X_i)-F_n(X_i)}{h}$$

$$- \frac{1}{2}\sum_{k=1}^{n}\omega^{'}(\frac{t-\xi_{n,k,i}}{h})(\frac{F_{n,k}(X_i)-F_n(X_i)}{h})^2\}$$

$$= -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n} X_i\omega^{'}(\frac{t-\xi_{n,k,i}}{h})(\frac{F_{n,k}(X_i)-F_n(X_i)}{h})^2,$$

where $\xi_{n,k,i}$ is a random variable between $F_n(X_i)$ and $F_{n,k}(X_i)$.

Since

(26)
$$F_n(X) - F_{n,k}(X) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le X) - \frac{1}{n-1}\sum_{j\neq k} I(X_i \le X) = \frac{1}{n-1}\{I(X_k \le X) - F_n(X)\} = O_p(n^{-1}),$$

under conditions of Theorem 2.1, we have that

(27)
$$|H_1| \le \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}|X_i\omega^{'}(\frac{t-\xi_{n,k,i}}{h})|\left(\frac{F_{n,k}(X_i)-F_n(X_i)}{h}\right)^2$$

$$= \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}|X_i\omega^{'}(\frac{t-\xi_{n,k,i}}{h})|O_p(\frac{1}{n^2h^2})$$

$$= O_p(\frac{1}{nh^2}).$$

Meanwhile, H_2 from (24) can be written as

(28)
$$\sum_{k=1}^{n}\{(\frac{1}{n}-\frac{1}{n-1})\sum_{j=1}^{n} X_jK(\frac{t-F_{n,k}(X_j)}{h}) + \frac{X_k}{(n-1)}K(\frac{t-F_{n,k}(X_k)}{h})\}$$

$$= \sum_{k=1}^{n}\{\frac{-1}{n(n-1)}\sum_{j=1}^{n} X_jK(\frac{t-F_{n,k}(X_j)}{h}) + \frac{1}{n-1}X_kK(\frac{t-F_{n,k}(X_k)}{h})\}$$

$$= \frac{-1}{n-1}\sum_{k=1}^{n}\{\frac{1}{n}\sum_{j=1}^{n} X_jK(\frac{t-F_{n,k}(X_j)}{h}) - X_kK(\frac{t-F_{n,k}(X_k)}{h})\}$$

$$= \frac{-1}{n-1}\sum_{k=1}^{n}\{\frac{1}{n}\sum_{j=1}^{n} X_jK(\frac{t-F_{n,k}(X_j)}{h}) - \frac{1}{n}\sum_{j=1}^{n} X_jK(\frac{t-F_n(X_j)}{h}) + \frac{1}{n}\sum_{j=1}^{n} X_jK(\frac{t-F_n(X_j)}{h})\}$$

$$- \frac{-1}{n-1}\sum_{k=1}^{n}\{X_kK(\frac{t-F_{n,k}(X_k)}{h}) - X_kK(\frac{t-F_n(X_k)}{h}) + X_kK(\frac{t-F_n(X_k)}{h})\}$$

$$= \frac{-1}{n(n-1)}\sum_{k=1}^{n}\sum_{j=1}^{n}[X_jK(\frac{t-F_{n,k}(X_j)}{h}) - X_jK(\frac{t-F_n(X_j)}{h})]$$

$$+ \frac{-1}{n(n-1)}\sum_{k=1}^{n}\sum_{j=1}^{n} X_jK(\frac{t-F_n(X_j)}{h})$$

$$+ \frac{1}{n-1}\sum_{k=1}^{n}[X_kK(\frac{t-F_{n,k}(X_k)}{h}) - X_kK(\frac{t-F_n(X_k)}{h})] + \frac{1}{n-1}\sum_{k=1}^{n} X_kK(\frac{t-F_n(X_k)}{h})$$

$$= \frac{1}{n-1}O_p(\frac{1}{nh^2}) - \frac{1}{n(n-1)}\sum_{k=1}^{n}\sum_{j=1}^{n} X_jK(\frac{t-F_n(X_j)}{h})$$

$$+ O_p(\frac{1}{n}) + \frac{1}{n-1}\sum_{k=1}^{n} X_kK(\frac{t-F_n(X_k)}{h})$$

$$= -\frac{1}{n-1}\sum_{j=1}^{n} X_jK(\frac{t-F_n(X_j)}{h})$$

$$+ \frac{1}{n-1}\sum_{k=1}^{n} X_kK(\frac{t-F_n(X_k)}{h}) + O_p(\frac{1}{n}) + O_p(\frac{1}{n^2h^2})$$

$$= O_p(\frac{1}{n}).$$

From (22)–(28), it follows that

$$\frac{1}{n}\sum_{k=1}^{n}\hat{V}_k(t) = \frac{n-1}{n}\sum_{k=1}^{n}(\hat{T}_n(t) - \hat{T}_{n-1,k}(t)) + \hat{T}_n(t) = \hat{T}_n(t) + O_p(\frac{1}{nh^2}).$$

Therefore,

$$\sqrt{n}\{\frac{1}{n}\sum_{k=1}^{n}\hat{V}_k(t) - \theta(t)\} = \sqrt{n}[\hat{T}_n(t) - \theta(t)] + O_p(\frac{1}{\sqrt{n}h^2}) \xrightarrow{d} N(0, \sigma^2(t)).$$

Thus, Lemma 1 holds. □

Lemma 2. Under the conditions in Theorem 2.1, we have

$$\frac{1}{n}\sum_{k=1}^{n}\{\hat{V}_k(t) - \theta(t)\}^2 \xrightarrow{p} \sigma^2(t).$$

Proof. We have the following decompositions:

$$\frac{1}{n}\sum_{k=1}^{n}\{\hat{V}_k(t)-\theta(t)\}^2 = \frac{1}{n}\sum_{k=1}^{n}\hat{V}_k^2(t)-2\theta(t)\frac{1}{n}\sum_{k=1}^{n}\hat{V}_k(t)+\frac{1}{n}\sum_{k=1}^{n}\theta^2(t), \tag{29}$$

where

$$\begin{aligned}\hat{V}_k(t) &= n\hat{T}_n(t)-(n-1)\hat{T}_{n-1,k}(t)\\ &= \sum_{i=1}^{n}X_iK(\frac{t-F_n(X_i)}{h})-\sum_{j\neq k}X_jK(\frac{t-F_{n,k}(X_j)}{h})\\ &= \sum_{i=1}^{n}X_i[K(\frac{t-F_n(X_i)}{h})-K(\frac{t-F_{n,k}(X_i)}{h})]\\ &\quad + X_kK(\frac{t-F_{n,k}(X_k)}{h})\end{aligned} \tag{30}$$

and

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{n}\hat{V}_k^2(t)\\ &= \frac{1}{n}\sum_{k=1}^{n}\{\sum_{i=1}^{n}X_i[K(\frac{t-F_n(X_i)}{h})-K(\frac{t-F_{n,k}(X_i)}{h})]\}^2\\ &\quad+\frac{1}{n}\sum_{k=1}^{n}X_k^2K^2(\frac{t-F_{n,k}(X_k)}{h})\\ &\quad+\frac{2}{n}\sum_{k=1}^{n}X_kK(\frac{t-F_{n,k}(X_k)}{h})\sum_{i=1}^{n}X_i[K(\frac{t-F_n(X_i)}{h})\\ &\quad-K(\frac{t-F_{n,k}(X_i)}{h})]\\ &\equiv J_1+J_2+J_3.\end{aligned} \tag{31}$$

Note that $\sqrt{n}(F_n(x)-F(x))\xrightarrow{d}B(x)$, which is a Gaussian process. Also, by Lemma 1, $F_{n,k}(x)-F_n(x)=O_p(\frac{1}{n})$. Therefore, based on Taylor series, J_1 of (31) can be written as:

$$\begin{aligned}J_1 &= \frac{1}{n}\sum_{k=1}^{n}\{\sum_{i=1}^{n}X_i[K(\frac{t-F_n(X_i)}{h})-K(\frac{t-F_{n,k}(X_i)}{h})]\}^2\\ &= \frac{1}{n}\sum_{k=1}^{n}\{\sum_{i=1}^{n}X_i\omega(\frac{t-F_n(X_i)}{h})\frac{F_{n,k}(X_i)-F_n(X_i)}{h}\\ &\quad-\frac{1}{2}\sum_{i=1}^{n}X_i\omega^{'}(\frac{t-\xi_{n,k,i}}{h})(\frac{F_{n,k}(X_i)-F_n(X_i)}{h})^2\}^2\\ &= \frac{1}{n}\sum_{k=1}^{n}\{\sum_{i=1}^{n}X_i\omega(\frac{t-F_n(X_i)}{h})\frac{F_{n,k}(X_i)-F_n(X_i)}{h}\}^2\\ &\quad-\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}X_i\omega(\frac{t-F_n(X_i)}{h})\frac{F_{n,k}(X_i)-F_n(X_i)}{h}\\ &\quad\times\sum_{i=1}^{n}X_i\omega^{'}(\frac{t-\xi_{n,k,i}}{h})(\frac{F_{n,k}(X_i)-F_n(X_i)}{h})^2\\ &\quad+\frac{1}{4n}\sum_{k=1}^{n}\{\sum_{i=1}^{n}X_i\omega^{'}(\frac{t-\xi_{n,k,i}}{h})(\frac{F_{n,k}(X_i)-F_n(X_i)}{h})^2\}^2\\ &= \frac{1}{n}\sum_{k=1}^{n}\{\sum_{i=1}^{n}X_i\omega(\frac{t-F_n(X_i)}{h})\frac{F_{n,k}(X_i)-F_n(X_i)}{h}\}^2\\ &\quad+O_p(\frac{1}{nh})+O_p(\frac{1}{(nh)^2})\\ &= \frac{1}{n}\sum_{k=1}^{n}\{n\int_{-\infty}^{\infty}x\omega(\frac{t-F_n(x)}{h})\frac{F_{n,k}(x)-F_n(x)}{h}\\ &\quad dF_n(x)\}^2+o_p(1)\\ &= \frac{1}{n}\sum_{k=1}^{n}\{\frac{n}{\sqrt{n}}\int_{-\infty}^{\infty}x\omega(\frac{t-F_n(x)}{h})\frac{F_{n,k}(x)-F_n(x)}{h}\\ &\quad d[\sqrt{n}(F_n(x)-F(x))]+\int_{-\infty}^{\infty}nx\omega(\frac{t-F_n(x)}{h})\\ &\quad\frac{F_{n,k}(x)-F_n(x)}{h}dF(x)\}^2+o_p(1)\\ &= \frac{1}{n}\sum_{k=1}^{n}\{\int_{-\infty}^{\infty}nx\omega(\frac{t-F_n(x)}{h})\frac{F_{n,k}(x)-F_n(x)}{h}dF(x)\}^2\\ &\quad+o_p(1),\end{aligned} \tag{32}$$

where $\xi_{n,k,i}$ is a random variable between $F_{n,k}(X_i)$ and $F_n(X_i)$.

Let $y_1=F_n(x_1), F_{n,k}(x_1)=\frac{1}{n-1}\sum_{i\neq k}^{n}I(X_i\leq x_1)$. Exactly similar to Gong *et al.* (2010), J_1 can be written as

$$\begin{aligned}J_1 &= \frac{n^2}{n}\sum_{k=1}^{n}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2\\ &\quad(\frac{F_{n,k}(x_1)-F_n(x_1)}{h})(\frac{F_{n,k}(x_2)-F_n(x_2)}{h})\\ &\quad\omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_2)}{h})dF_n(x_1)dF_n(x_2)\}\\ &\quad+o_p(1)\\ &= \frac{n}{h^2}\sum_{k=1}^{n}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2\\ &\quad[\frac{1}{n-1}\sum_{i\neq k}^{n}I(X_i\leq x_1)-\frac{1}{n}\sum_{i=1}^{n}I(X_i\leq x_1)]\\ &\quad[\frac{1}{n-1}\sum_{i\neq k}^{n}I(X_i\leq x_2)-\frac{1}{n}\sum_{i=1}^{n}I(X_i\leq x_2)]\\ &\quad\omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_2)}{h})dF_n(x_1)dF_n(x_2)\}\\ &\quad+o_p(1)\end{aligned} \tag{33}$$

$$
\begin{aligned}
&= \frac{n}{h^2}\sum_{k=1}^{n}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 \quad [\frac{1}{n-1}\sum_{i=1}^{n} I(X_i \le x_1) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x_1) - \frac{1}{n-1} I(X_k \le x_1)] \\
&\quad [\frac{1}{n-1}\sum_{i=1}^{n} I(X_i \le x_2) - \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x_2) \\
&\quad - \frac{1}{n-1} I(X_k \le x_2)] \\
&\quad \omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_2)}{h}) dF_n(x_1)dF_n(x_2)\} \\
&\quad + o_p(1) \\
&= \frac{n}{h^2}\sum_{k=1}^{n}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 \quad [\frac{1}{n(n-1)}\sum_{i=1}^{n} I(X_i \le x_1) \\
&\quad - \frac{1}{n-1} I(X_k \le x_1)] \\
&\quad [\frac{1}{n(n-1)}\sum_{i=1}^{n} I(X_i \le x_2) \\
&\quad - \frac{1}{n-1} I(X_k \le x_2)]\omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_2)}{h}) \\
&\quad dF_n(x_1)dF_n(x_2)\} + o_p(1) \\
&= \frac{n}{h^2}\sum_{k=1}^{n}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 \frac{1}{n-1}[\frac{1}{n}\sum_{i=1}^{n} I(X_i \le x_1) \\
&\quad - I(X_k \le x_1)] \\
&\quad \frac{1}{n-1}[\frac{1}{n}\sum_{i=1}^{n} I(X_i \le x_2) - I(X_k \le x_2)] \\
&\quad \omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_2)}{h}) dF_n(x_1)dF_n(x_2)\} \\
&\quad + o_p(1) \\
&= \frac{n}{(n-1)^2h^2}\sum_{k=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2[F_n(x_1) \\
&\quad - I(X_k \le x_1)][F_n(x_2) - I(X_k \le x_2)] \\
&\quad \omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_x)}{h}) dF_n(x_1)dF_n(x_2) \\
&\quad + o_p(1) \\
&= \frac{n}{(n-1)^2h^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 \\
&\quad \sum_{k=1}^{n}[F_n(x_1)F_n(x_2) + I(X_k \le x_1)I(X_k \le x_2) \\
&\quad - I(X_k \le x_1)F_n(x_2) - I(X_k \le x_2)F_n(x_1)] \\
&\quad \omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_2)}{h}) dF_n(x_1)dF_n(x_2) \\
&\quad + o_p(1) \\
&= \frac{1}{h^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2[F_n(x_1\wedge x_2) - F_n(x_1)F_n(x_2)] \\
&\quad \omega(\frac{t-F_n(x_1)}{h})\omega(\frac{t-F_n(x_2)}{h}) dF_n(x_1)dF_n(x_2) \\
&\quad + o_p(1) \\
&= \frac{1}{h^2}\int_{-1}^{1}\int_{-1}^{1} F_n^{-1}(y_1)F_n^{-1}(y_2)\{F_n[F_n^{-1}(y_1)\wedge F_n^{-1}(y_2)] \\
&\quad - y_1y_2\}\omega(\frac{t-y_1}{h})\omega(\frac{t-y_2}{h}) dy_1dy_2 + o_p(1) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^{-1}(t)F^{-1}(t)\{t\wedge t - t^2\} \\
&\quad \omega(u_1)\omega(u_2)du_1du_2 + o_p(1) \\
&= \xi_t^2 t(1-t) + o_p(1).
\end{aligned}
$$

Based on (16), $E[X^2K^2(\frac{t-F(X)}{h})] = \int_0^{\xi_t} x^2dF(x) + o(1)$. Therefore,

$$
\begin{aligned}
(34)\ J_2 &= \frac{1}{n}\sum_{k=1}^{n} X_k^2K^2(\frac{t-F_{n,k}(X_k)}{h}) \\
&= \frac{1}{n}\sum_{k=1}^{n} X_k^2K^2(\frac{t-F_{n,k}(X_k)}{h}) \\
&\quad - \frac{1}{n}\sum_{k=1}^{n} X_k^2K^2(\frac{t-F(X_k)}{h}) \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} X_k^2K^2(\frac{t-F(X_k)}{h}) \\
&= \frac{1}{n}\sum_{k=1}^{n} X_k^2[K^2(\frac{t-F_{n,k}(X_k)}{h}) - K^2(\frac{t-F(X_k)}{h})] \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} X_k^2K^2(\frac{t-F(X_k)}{h}) = \frac{1}{n}\sum_{k=1}^{n} X_k^2 \\
&\quad \{2K(\frac{t-F(X_k)}{h})\omega(\frac{t-F(X_k)}{h})\frac{F_{n,k}(X_k)-F(X_k)}{h} \\
&\quad + \frac{1}{2}[2\omega^2(\frac{t-F(X_k)}{h})(\frac{F_{n,k}(X_k)-F(X_k)}{h})^2 \\
&\quad + 2K(\frac{t-F(X_k)}{h})\omega'(\frac{t-\xi_{n,k,i}}{h}) \\
&\quad (\frac{F_{n,k}(X_k)-F(X_k)}{h})^2]\} \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} X_k^2K^2(\frac{t-F(X_k)}{h}) \\
&= \frac{1}{n}\sum_{k=1}^{n} X_k^2K^2(\frac{t-F(X_k)}{h}) + o_p(1) \\
&= E\left[X^2K^2(\frac{t-F(X)}{h})\right] + o_p(1) \\
&= \int_0^{\xi_t} x^2dF(x) + o_p(1).
\end{aligned}
$$

By Chebyshev's Inequality, we have that

$$P(|X_kK(\frac{t-F(X_k)}{h})| \geq M) \leq \frac{1}{M^2}E[X^2K^2(\frac{t-F(X)}{h})] = O(\frac{1}{M^2}) \to 0,\ as\ M \to \infty.$$

So, $X_kK(\frac{t-F_{n,k}(X_k)}{h}) = O_p(1)$ uniformly for $k = 1, 2, ..., n$. Using (27), (31) and the similar proof to (27), we can get that

(35)

$$\begin{aligned} J_3 &= \frac{2}{n}\sum_{k=1}^{n} X_kK(\frac{t-F_{n,k}(X_k)}{h})\sum_{i=1}^{n} X_i[K(\frac{t-F_n(X_i)}{h}) \\ &\quad - K(\frac{t-F_{n,k}(X_i)}{h})] \\ &\leq \frac{2}{n}\sum_{k=1}^{n} X_kK(\frac{t-F_{n,k}(X_k)}{h})\sum_{i=1}^{n} |X_i[K(\frac{t-F_n(X_i)}{h}) \\ &\quad - K(\frac{t-F_{n,k}(X_i)}{h})]| \\ &= O_p(\frac{1}{n})\sum_{k=1}^{n}\sum_{i=1}^{n} |X_i[K(\frac{t-F_n(X_i)}{h}) \\ &\quad - K(\frac{t-F_{n,k}(X_i)}{h})]| = O_p(\frac{1}{nh^2}). \end{aligned}$$

From (33), (34) and (35), it follows that

$$\frac{1}{n}\sum_{k=1}^{n} \hat{V}_k^2(t) \xrightarrow{p} \xi_t^2 t(1-t) + \int_0^{\xi_t} x^2 dF(x).$$

Hence,

$$\frac{1}{n}\sum_{k=1}^{n}\{\hat{V}_k(t) - \theta(t)\}^2 \xrightarrow{p} \xi_t^2 t(1-t) + \int_0^{\xi_t} x^2 dF(x) - \theta^2(t) = \sigma^2(t),$$

and Lemma 2 is proved. □

Proof of Theorem 3.1. It follows immediately from Lemma 1 and Lemma 2. □

Proof of Theorem 3.2. Define $g(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{V}_i(t)-\theta}{1+\lambda(\hat{V}_i(t)-\theta)}$. It is easy to check that

(36)

$$\begin{aligned} 0 = |g(\lambda)| &= \frac{1}{n}|\sum_{i=1}^{n}(\hat{V}_i(t)-\theta) - \lambda\sum_{i=1}^{n}\frac{(\hat{V}_i(t)-\theta)^2}{1+\lambda(\hat{V}_i(t)-\theta)}| \\ &\geq |\frac{\lambda}{n}\sum_{i=1}^{n}\frac{(\hat{V}_i(t)-\theta)^2}{1+\lambda(\hat{V}_i(t)-\theta)}| - |\frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta)| \\ &\geq \frac{|\lambda|S_n}{1+|\lambda|Z_n} - |\frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta)|, \end{aligned}$$

where $S_n = \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t) - \theta)^2$ and $Z_n = max_{1\leq i\leq n}|\hat{V}_i(t)-\theta|$.

From Lemma 1 and Lemma 2, we have $|\lambda| = O_p\{n^{-\frac{1}{2}}\}$. Put $\gamma_i = \lambda(\hat{V}_i(t)-\theta)$, then we have $max_{1\leq i\leq n}|\gamma_i| = o_p(1)$, and

(37)

$$\begin{aligned} 0 = g(\lambda) &= \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta)(1-\gamma_i+\frac{\gamma_i^2}{1+\gamma_i}) \\ &= \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta) - S_n\lambda + \frac{\lambda^2}{n}\sum_{i=1}^{n}\frac{(\hat{V}_i(t)-\theta)^3}{1+\gamma_i} \\ &= \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta) - S_n\lambda + o_p(n^{-1/2}) \end{aligned}$$

which implies that $\lambda = S_n^{-1}\frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta) + \beta_n$, where $\beta_n = o_p(n^{-1/2})$. So,

(38)

$$\begin{aligned} l_n(\theta(t)) &= -2\ log\ L_n(\theta(t)) \\ &= 2\sum_{i=1}^{n} log\ \{1+\lambda(\hat{V}_i(t)-\theta)\} \\ &= 2\sum_{i=1}^{n}\gamma_i - \sum_{i=1}^{n}\gamma_i^2 + o_p(1) \\ &= 2n\lambda\frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta) - nS_n\lambda^2 + o_p(1) \\ &= \frac{n\{\frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta)\}^2}{S_n} - nS_n\beta_n^2 + o_p(1) \\ &= \frac{n\{\frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(t)-\theta)\}^2}{S_n} + o_p(1) \\ &\xrightarrow{d} \chi^2(1). \end{aligned}$$

The proof of Theorem 3.2 is completed. □

Received 8 October 2014

REFERENCES

Beach, C. M. and Davidson, R. (1983). Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economics and Statistics*. **50**, 723–735.

Chan, N. H., Li, D., and Peng, L. (2012). Toward a unified interval estimation of autoregressions. *Econometric Theory*. **28**, 705–717. MR2927926

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimator of smooth distribution functions. *Statist. Neerlandica*. **37**, 73–83. MR0711744

Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. *Ann. Statist.* **13**, 428–433. MR0773180

Feng, H. and Peng, L. (2012). Jackknife empirical likelihood tests for error distributions in regression models. *J. Multivariate Anal.* **112**, 63–75. MR2957286

Gong, Y., Peng, L., and Qi, Y. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis*. **101**, 1520–1531. MR2609511

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statist. Review*. **58**, 109–127.

Jing, B., Yuan, J., and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*. **104**, 1224–1232.

Kakwani, N. (1984). Welfare ranking of income distributions. *Advances in Econometrics*. **3**, 191–213.

Kleiber, C. and Kramer, W. (2003). Efficiency, equity and generalized Lorenz dominance. *Estadistica*. **55** (Special Issue on Income Distribuiton, Inequality and Poverty, ed. C. Dagum), 173–186.

Li, Z., Gong, Y., and Peng, L. (2011). Empirical likelihood intervals for conditional value-at-risk in heteroscedastic regression models. *Scand. J. Stat.* **38**, 781–787. MR2859750

Lloyd, C. J. and Yong, Z. (1999). Kernel estimators for the ROC curve are better than empirical. *Statistics and Probability Letters*. **44**, 221–228. MR1711637

Lorenz, M. C. (1905). Methods of measuring the concentration of wealth. *Journal of the American Statistical Association*. **9**, 209–219.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for single functional. *Biometrika*. **75**, 237–249. MR0946049

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*. **18**, 90–120. MR1041387

Owen, A. (2001). *Empirical likelihood*. Chapman and Hall/CRC: New York.

Peng, L., Qi, Y., Wang, R., and Yang, J. (2012). Jackknife empirical likelihood methods for risk measures and related quantities. *Insurance: Mathematics and Economics*. **51**, 142–150. MR2928751

Qin, G. S. and Zhou, X. H. (2006). Empirical likelihood inference for the area under the ROC curve. *Biometrics*. **62**, 613–622. MR2236844

Rao, J. N. K. and Wu, C. (2010). Pseudo-empirical likelihood inference for multiple frame surveys. *Journal of the American Statistical Association*. **105**, 1494–1503. MR2796566

Shorrocks, A. F. (1983). Ranking income distributions. *Economica*. **50**, 3–17.

Thistle, P. D. (1989a). Duality between generalized Lorenz curves and distribution functions. *Economic Studies Quarterly*. **40**, 183–187.

Thistle, P. D. (1989b). Ranking distribution with generalized Lorenz curves. *Southern Economic Journal*. **56**, 1–12.

Tukey, J. W. (1958). Bias and confidence in not-quite large sample (Abstract). *Annals of Mathematical Statisitics*. **29**, 614.

Wang, R., Peng, L., and Qi, Y. (2013). Jackknife empirical likelihood test for the equality of two high dimensional means. *Statistica Sinica*. **23**, 667–690. MR3086651

Wei, Z., Wen, S., and Zhu, L. (2009). Empirical likelihood-based evaluations of value at risk models. *Science in China Series A: Mathematics*. **52**, 1995–2006. MR2545004

Wei, Z. and Zhu, L. (2010). Evaluation of value at risk: an empirical likelihood approach. *Statistica Sinica*. **20**, 455–468. MR2640704

Wood, A. T. A., Do, K. A., and Broom, N. M. (1996). Sequential linearization of empirical likelihood constraints with application to u-statistics. *Journal of Computational and Graphical Statistics*. **5**, 365–385. MR1422116

Xu, K. (1997). Asymptotically distribution-free statistical test for generalized Lorenz curves: an alternative approach. *Journal of Income Distribution*. **7**, 45–62.

Yang, B. Y., Qin, G. S., and Belinga-Hall, N. E. (2012). Nonparametric inferences for the generalized Lorenz curve (in Chinese). *Scientia Sinica Mathematica (Science in China Series A)*. **42**, 235–250.

Zheng, B. (2002). Testing Lorenz curves with non-simple random samples. *Econometrica*. **70**, 1235–1243. MR1910418

Zhou, X. H., Qin, G. S., Lin, H. Z., and Li, G. (2006). Inferences in censored cost regression models with empirical likelihood. *Statistica Sinica*. **16**, 1213–1232. MR2327487

Shan Luo
Department of Mathematics and Statistics
Georgia State University
Atlanta, GA 30303
USA
E-mail address: lsluols@hotmail.com

Gengsheng Qin
Department of Mathematics and Statistics
Georgia State University
Atlanta, GA 30303
USA
E-mail address: gqin@gsu.edu