Computational Statistics (2024) 39:3073–3090
https://doi.org/10.1007/s00180-023-01431-8

ORIGINAL PAPER

Nonparametric confidence intervals for generalized Lorenz curve using modified empirical likelihood

Suthakaran Ratnasingam[1] · Spencer Wallace[1] · Imran Amani[1] · Jade Romero[1]

Received: 5 June 2023 / Accepted: 20 October 2023 / Published online: 3 November 2023

Abstract

The Lorenz curve portrays income distribution inequality. In this article, we develop three modified empirical likelihood (EL) approaches, including adjusted empirical likelihood, transformed empirical likelihood, and transformed adjusted empirical likelihood, to construct confidence intervals for the generalized Lorenz ordinate. We demonstrate that the limiting distribution of the modified EL ratio statistics for the generalized Lorenz ordinate follows scaled Chi-Squared distributions with one degree of freedom. We compare the coverage probabilities and mean lengths of confidence intervals of the proposed methods with the traditional EL method through simulations under various scenarios. Finally, we illustrate the proposed methods using real data to construct confidence intervals.

Keywords Generalized Lorenz curve · Empirical likelihood · Modified empirical likelihood · Confidence intervals · Coverage probability

1 Introduction

The Lorenz curve developed by American economist Lorenz (Lorenz 1905) is a graphical representation used to describe income and wealth inequality. A Lorenz curve with perfect equality follows a diagonal line (45° angle) in which the income percentage is always proportional to the population percentage; however, in the real world, the Lorenz curve falls below this line. As the actual income distribution is rarely known, the distribution is typically estimated from income data. Several researchers have made contributions to Lorenz curves analysis, for example, Sen (1973), Jakobsson (1976), Goldie (1977), and Marshall and Olkin (1979). The full

Spencer Wallace, Imran Amani and Jade Romero are contributed equally to this work.

✉ Suthakaran Ratnasingam
suthakaran.ratnasingam@csusb.edu

[1] Department of Mathematics, California State University, San Bernardino, San Bernardino, CA 92407, USA

joint variance-covariance structure for the Lorenz curve ordinates was developed by Beach and Davidson (1983). John et al. (1989) proposed new results on generalized Lorenz ordinate analysis that are relevant for proving second-degree stochastic dominance. Allen (1990), Lambert (2001), and Mosler and Koshevov (2007) have made recent advances, with their findings leading to a wide range of applications, particularly in reliability theory. Ryu and Slottje (1996) proposed an exponential polynomial expansion and a Bernstein polynomial expansion as two flexible functional form approaches for approximating Lorenz curves. Hasegawa and Kozumi (2003) proposed a Bayesian non-parametric analysis approach with the Dirichlet process prior to Lorenz curve estimation with contaminated data. In particular, their method allows for heteroscedasticity in individual incomes. Further, the Lorenz curve has been used by several researchers to analyze physician distributions. For example, Chang and Halfon (1997) examined variations in the distribution of pediatricians among the states between 1982 and 1992 using Lorenz curves and Gini indices. Kobayashi and Takaki (1992) used the Lorenz curve and the Gini coefficient to study the disparity in physician distribution in Japan.

Empirical likelihood (EL) is a nonparametric method introduced by Owen (2001), an alternative to the standard parametric likelihood that inherits many alluring features such as its extension of Wilks' theorem, asymmetric confidence interval, better coverage for small sample sizes, and so on. Many researchers have studied EL for the Lorenz curve. For instance, Belinga-Hall (2007) and Yang et al. (2012) developed plug-in empirical likelihood-based inferences to construct confidence intervals for the generalized Lorenz curve. Qin et al. (2013) studied EL-based confidence interval for the Lorenz curve under the simple random sampling and the stratified random sampling designs. Shi et al. (2019) proposed new nonparametric confidence intervals using the influence function-based empirical likelihood method for the Lorenz curve and showed that the limiting distributions of the empirical log-likelihood ratio statistics for the Lorenz ordinates were standard chi-square distributions. Luo and Qin (2019) suggested a kernel smoothing estimator for the Lorenz curve and developed a smoothed jackknife empirical likelihood approach for constructing confidence intervals of Lorenz ordinates.

Despite being widely used, the EL-based approach has two major drawbacks: (1) the convex hull must have vector zero as its interior point in order to solve the profile empirical likelihood problem. According to Owen (2001), the empirical likelihood function should be set to $-\infty$ if the convex hull does not have zero as an interior point. However, Chen et al. (2008), pointed out that this makes it difficult to find the maximum of the EL function. (2) The EL technique frequently experiences under-coverage problems, see (Tsao 2013) for more details. To address these problems, numerous strategies have been proposed in the literature. Chen et al. (2008) proposed adjusted empirical likelihood method (AEL), for example, confirms the existence of a solution in the maximization problem while preserving asymptotic optimality properties. Further, Jing et al. (2017) suggested the transformed empirical likelihood to tackle the under-coverage problem for small sample sizes (TEL). Stewart and Ning (2020) proposed the transformed adjusted empirical likelihood (TAEL), a strategy that combines the AEL and TEL approaches. The AEL, TEL, and TAEL approaches are proven to be effective in many applications, for example, Li et al. (2022) investigated modified EL-based confidence intervals for quantile regression models with longitudinal data. Ratnasingam and Ning (2022) studied all three modified versions of EL procedures to construct confidence intervals of the mean residual life function with length-biased data.

In this research, we develop three modified EL-based inference procedures to construct confidence intervals for the generalized Lorenz curve. These modified EL methods aim to address the shortcomings of traditional EL, including the issue of under-coverage, while also ensuring the existence of a solution for the maximization procedure. To the best of our knowledge, this is the first study to investigate AEL, TEL, and TAEL methods for constructing confidence intervals for the generalized Lorenz ordinate.

The remainder of this paper is organized as follows. In Sect. 2, we briefly describe the fundamental properties of EL for the generalized Lorenz curve and provide the methodology of AEL, TEL, and TAEL for the generalized Lorenz curve. In Sect. 3, we conduct an extensive simulation study to compare the finite sample performances of the proposed confidence intervals for the generalized Lorenz ordinates. In Sect. 4, we use an income dataset to illustrate the proposed intervals. In Sect. 5, we discuss our results and draw conclusions. The simulations are included in the Appendix.

2 Empirical likelihood based methods

2.1 Empirical likelihood

Let X be a random variable with cumulative distribution function (CDF) denoted by $F(x)$ with finite support. For instance, $F(\cdot)$ denotes the CDF of income or wealth distribution. Following (Gastwirth 1971), a general definition of the Lorenz curve is provided below.

$$\eta(t) = \frac{1}{\mu}\int_0^{\psi_t} xdF(x), \quad t \in [0,1] \tag{1}$$

where μ denotes the mean of F, and $\psi_t = F^{-1}(t) = \inf\{x : F(x) \geq t\}$ is the t-th quantile of F. For a fixed $t \in [0,1]$, the Lorenz ordinate $\eta(t)$ is the ratio of the mean income of the lowest t-th fraction of households and the mean income of total households. The generalized Lorenz curve is defined as follows.

$$\theta(t) = \int_0^{\psi_t} xdF(x), \quad t \in [0,1] \tag{2}$$

Because the income distribution $F(x)$ is rarely known in practice, the Lorenz curve is typically estimated from income data. Hence, the empirical estimator for $\eta(t)$ is defined as

$$\hat{\eta}(t) = \frac{1}{\hat{\mu}}\int_0^{\hat{\psi}_t} xd\hat{F}_n(x), \quad t \in [0,1] \tag{3}$$

where $\widehat{F}_n(x)$ is the empirical distribution function of the X_i's, $\widehat{\mu}$ is the sample mean, $\widehat{\psi}_t$ is the t-th sample quantile of the X_i's. From the definition of the generalized Lorenz curve, we observe that

$$E[XI(X \leq \psi_t)] - \theta(t) = 0.$$

Therefore, the empirical likelihood of $\theta(t)$ can be expressed as

$$L^*(\theta(t)) = \sup_p \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i W_i(t) = 0 \right\}, \tag{4}$$

where $\mathbf{p} = (p_1, p_2, \dots, p_n)$ is a probability vector satisfying $\sum_{i=1}^n p_i = 1$ and $p \geq 0$ for all i, and $W_i(t) = X_i I(X_i \leq \psi_t) - \theta(t), i = 1, 2, \dots, n$. It can be seen that $W_i(t)$ in (4) depends on the unknown t-quantile ψ_t. As a result, the generalized Lorenz ordinate $\theta(t)$ is the mean of the random variable X truncated at ψ_t. Using sample data, the empirical likelihood for $\theta(t)$ as follows:

$$L(\theta(t)) = \sup_p \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \widehat{W}_i(t) = 0 \right\}, \tag{5}$$

where $\widehat{W}_i(t) = X_i I(X_i \leq \widehat{\psi}_t) - \theta(t), i = 1, 2, \dots, n$. When the vector $\mathbf{p} = (p_1, p_2, \dots, p_n)$ is contained within the convex hull defined by $\{X_1 I(X_1 \leq \widehat{\psi}_t), X_2 I(X_2 \leq \widehat{\psi}_t), \dots, X_n I(X_n \leq \widehat{\psi}_t)\}$, Eq. (5) attains its unique maximum value. By applying the Lagrange multiplier method, we can determine, $L(\theta(t))$ as follows.

$$p_i = \frac{1}{n} \left\{ 1 + \lambda(t) \widehat{W}_i(t) \right\}^{-1}, \quad i = 1, \dots, n.$$

where λ is the solution to

$$\frac{1}{n} \sum_{i=1}^n \frac{\widehat{W}_i(t)}{1 + \lambda(t) \widehat{W}_i(t)} = 0.$$

Note that $\prod_{i=1}^n p_i$, subject to $\sum_{i=1}^n p_i = 1$, attains its maximum n^{-n} at $p_i = n^{-1}$. Thus, the EL ratio for $\theta(t)$ is given as

$$\mathcal{R}(\theta(t)) = \prod_{i=1}^n n p_i = \prod_{i=1}^n \left\{ 1 + \lambda(t) \widehat{W}_i(t) \right\}^{-1}. \tag{6}$$

Hence, the profile empirical log-likelihood ratio for $\theta(t)$ is

$$\ell(\theta(t)) = -2 \log \mathcal{R}(\theta(t)) = 2 \sum_{i=1}^n \log \left\{ 1 + \lambda(t) \widehat{W}_i(t) \right\}. \tag{7}$$

Theorem 2.1 *If* $E(X^2) < \infty$ *and* $\theta(t_0) = E[XI(X \leq \psi_{t_0})]$ *for any given* $t = t_0 \in (0,1)$ *then the limiting distribution of* $\mathrm{l}(\theta(t_0))$ *is a scaled chi-square distribution with one degree of freedom,*

$$\left(\frac{\sigma_p^2}{\sigma_v^2}\right)\ell(\theta(t_0)) \longrightarrow \chi_1^2, \quad \text{as } n \longrightarrow \infty. \tag{8}$$

where $\sigma_p^2 = Var\big(X\,I(X \leq \psi_t)\big)$ *and* $\sigma_v^2 = Var\big((X - \psi_t)I(X \leq \psi_t)\big)$.

Although the scale constant $\left(\frac{\sigma_p^2}{\sigma_v^2}\right)$ is unknown, it can be consistently estimated by using the following formula.

$$\begin{aligned} \sigma_p^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i\,I(X_i \leq \hat{\psi}_t) - \frac{1}{n}\sum_{i=1}^{n} X_i\,I(X_i \leq \hat{\psi}_t)\right)^2, \quad \text{and} \\ \sigma_v^2 &= \frac{1}{n}\sum_{i=1}^{n}\left((X_i - \hat{\psi}_t)I(X_i \leq \hat{\psi}_t) - \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\psi}_t)I(X_i \leq \hat{\psi}_t)\right)^2. \end{aligned} \tag{9}$$

Thus, an asymptotic $(1-\alpha)100\%$ confidence interval for generalized Lorenz ordinate, $\theta(t)$ at a fixed time t is given as follows

$$C(t) = \left\{\theta(t) : \left(\frac{\sigma_p^2}{\sigma_v^2}\right)\ell(\theta(t)) \leq \chi_{1,\alpha}^2\right\}, \tag{10}$$

where $\chi_{1,\alpha}^2$ is the upper α-quantile of the distribution of χ_1^2. For more details, we refer to Yang et al. (2012). As previously stated, the original EL method experiences low coverage probability, particularly for small sample sizes, for example, Chen et al. (2008) and Jing et al. (2017). Next, we describe the technical details of three modified EL-based methods for constructing confidence intervals. These methods are called AEL, TEL, and TAEL, and they are used for constructing confidence intervals for the generalized Lorenz ordinate $\theta(t)$, at a fixed time t for $t \in (0,1)$.

2.2 Adjusted Empirical Likelihood for Generalized Lorenz Ordinate

Chen et al. (2008) proposed the adjusted empirical likelihood (AEL) in order to address the challenge of the non-existence of a solution in the Eq. (7). We adopted the idea of the AEL method for generalized Lorenz ordinate. We define $\overline{W}_n = (1/n)\sum_{i=1}^{n} W_i(t)$. The pseudo value $W_{n+1}(t) = -a_n\overline{W}_n$, where $a_n = \max\left\{1, \frac{1}{2}\log n\right\}$. Using the $(n+1)$ observations, we define the adjusted empirical likelihood as

$$L^*(\theta(t)) = \sup_p\left\{\prod_{i=1}^{n+1} p_i \middle| p_i \geq 0, \sum_{i=1}^{n+1} p_i = 1, \sum_{i=1}^{n+1} p_i\hat{W}_i(t) = 0\right\} \tag{11}$$

Thus, the adjusted empirical log-likelihood ratio is given by

$$\ell^*(\theta(t)) = 2\sum_{i=1}^{n+1} \log\left(1 + \lambda(t)\widehat{W}_i(t)\right). \tag{12}$$

Theorem 2.2 *Assume that $E(X^2) < \infty$. For all $t_0 \in (0,1)$, let $\ell^*(\theta(t_0))$ be the adjusted log-empirical likelihood ratio function defined by (12) and $a_n = o_p(n^{2/3})$. We have*

$$\left(\frac{\sigma_p^2}{\sigma_v^2}\right)\ell^*(\theta(t_0)) \longrightarrow \chi_1^2, \quad \text{as } n \longrightarrow \infty. \tag{13}$$

in distribution.

Thus, an asymptotic $(1-\alpha)100\%$ confidence interval for $\theta(t)$ at a fixed time t is given as follows

$$C(t) = \left\{\theta(t) : \left(\frac{\sigma_p^2}{\sigma_v^2}\right)\ell^*(\theta(t)) \le \chi_{1,\alpha}^2\right\}, \tag{14}$$

where $\chi_{1,\alpha}^2$ is the upper α-quantile of the distribution of χ_1^2.

Proof Under the conditions of Theorem 2.1, Yang et al. (2012) showed that

$$\text{A1. } \frac{1}{n}\sum_{i=1}^{n} \widehat{W}_i^2(t) \xrightarrow{p} \sigma_p^2(t)$$

$$\text{A2. } \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widehat{W}_i(t) \xrightarrow{\mathcal{L}} N\left(0, \sigma_v^2(t)\right)$$

Following similar arguments as in the proof of Theorem 3.1 in Ratnasingam and Ning (2022) we have

$$\begin{aligned} \ell^*(\theta(t)) &= 2\sum_{i=1}^{n+1} \log\left(1 + \lambda(t)\widehat{W}_i(t)\right) \\ &= 2\sum_{i=1}^{n+1} \left\{\lambda(t)\widehat{W}_i(t) - \frac{\left(\lambda(t)\widehat{W}_i(t)\right)^2}{2}\right\} + o_p(1). \end{aligned} \tag{15}$$

Using (15), A1 and A2, we have

$$\left(\frac{\sigma_p^2}{\sigma_v^2}\right)\ell^*(\theta(t)) = \left(\frac{\sigma_p^2}{\sigma_v^2}\right)\sum_{i=1}^{n+1}\left(\lambda(t)\widehat{W}_i(t)\right)^2 + o_p(1)$$
$$= \frac{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n+1}\lambda(t)\widehat{W}_i(t)\right)^2}{\sigma_v^2}\;\frac{\sigma_p^2}{\frac{1}{n}\sum_{i=1}^{n+1}\widehat{W}_i(t)^2} + o_p(1)$$
$$\xrightarrow{d} \chi_1^2. \tag{16}$$

This completes the proof. □

2.3 Transformed empirical likelihood for generalized Lorenz ordinate

Jing et al. (2017) proposed the transformed empirical likelihood (TEL) as a simple transformation of the original EL to tackle the under-coverage problem. They claimed that TEL is superior in small sample sizes and multidimensional situations. The transformed empirical log-likelihood ratio can be defined as

$$\ell_t\big(\theta(t),\gamma\big) = \ell(\theta(t)) \times \max\left\{1 - \frac{\ell(\theta(t))}{n}, 1-\gamma\right\}, \tag{17}$$

where $\ell(\theta(t))$ is given in (7) and $\gamma \in [0, 1]$. It should be noted that $\gamma = 1/2$ ensures the maximum expansion without violating the conditions (C2) stated in Jing et al. (2017). Hence, the transformed empirical log-likelihood ratio is defined as

$$\ell_t(\theta(t)) = \ell_t\left(\theta(t),\gamma = \frac{1}{2}\right) = \ell(\theta(t)) \times \max\left\{1 - \frac{\ell(\theta(t))}{n},\ \frac{1}{2}\right\}. \tag{18}$$

Thus, the transformed empirical log-likelihood ratio is given as

$$\ell_t(\theta(t)) = \begin{cases} \ell(\theta(t))\left(1 - \frac{\ell(\theta(t))}{n}\right) & \text{if } \ell(\theta(t)) \le \frac{n}{2}, \\ \frac{\ell(\theta(t))}{2} & \text{if } \ell(\theta(t)) > \frac{n}{2}. \end{cases} \tag{19}$$

Further, Jing et al. (2017) showed that the TEL ratio meets four conditions that ensure the likelihood ratio's asymptotic properties.

Theorem 2.3 *Assume that $E(X^2) < \infty$, for all $t_0 \in (0, 1)$, let $\ell_t(\theta(t_0))$ be the transformed log-empirical likelihood ratio function defined by (18). We have*

$$\left(\frac{\sigma_p^2}{\sigma_v^2}\right)\ell_t(\theta(t_0)) \longrightarrow \chi_1^2, \quad \text{as } n \longrightarrow \infty. \tag{20}$$

in distribution.

Thus, an asymptotic $(1-\alpha)100\%$ confidence interval for $\theta(t)$ at a fixed time t is given as follows

$$C(t) = \left\{ \theta(t) : \left(\frac{\sigma_p^2}{\sigma_v^2} \right) \ell_t(\theta(t)) \leq \chi^2_{1,\alpha} \right\}, \tag{21}$$

where $\chi^2_{1,\alpha}$ is the upper α-quantile of the distribution of χ^2_1.

Proof The proof is omitted as it is similar to the proof of Theorem 3.2 in Ratnasingam and Ning (2022). □

2.4 Transformed adjusted empirical likelihood for generalized Lorenz ordinate

Stewart and Ning (2020) developed a hybrid method based upon AEL and TEL methods called transformed adjusted empirical likelihood (TAEL). The TAEL method combines the benefits of both AEL and TEL methods. Let $\gamma \in [0,1]$. We define

$$\ell_t^*\left(\ell^*(\theta(t)), \gamma\right) = \ell^*(\theta(t)) \times \max\left\{ 1 - \frac{\ell^*(\theta(t))}{n},\ 1-\gamma \right\}, \tag{22}$$

where $\ell^*(\cdot)$ defined in (12). Thus, for $\gamma = 1/2$, the transformed empirical log-likelihood ratio $\ell_t^*(\theta(t))$ is defined as,

$$\ell_t^*\left(\ell^*(\theta(t)), \frac{1}{2}\right) = \ell^*(\theta(t)) \times \max\left\{ 1 - \frac{\ell^*(\theta(t))}{n},\ \frac{1}{2} \right\}. \tag{23}$$

This can be further viewed as

$$\ell_t^*(\theta(t)) = \begin{cases} l^*(\theta(t))\left(1 - \dfrac{\ell^*(\theta(t))}{n}\right) & \text{if } \ell^*(\theta(t)) \leq \dfrac{n}{2}, \\ \dfrac{\ell^*(\theta(t))}{2} & \text{if } \ell^*(\theta(t)) > \dfrac{n}{2}. \end{cases} \tag{24}$$

Theorem 2.4 *Assume that $E(X^2) < \infty$, for all $t_0 \in (0,1)$, let $\ell_t^*(\theta(t_0))$ be the transformed log-empirical likelihood ratio function defined by (24). We have*

$$\left(\frac{\sigma_p^2}{\sigma_v^2} \right) \ell_t^*(\theta(t_0)) \longrightarrow \chi_1^2, \quad \text{as } n \longrightarrow \infty. \tag{25}$$

in distribution.

Thus, an asymptotic $(1-\alpha)100\%$ confidence interval for $\theta(t)$ at a fixed time t is given as follows

$$C(t) = \left\{ \theta(t) : \left(\frac{\sigma_p^2}{\sigma_v^2} \right) \ell_t^*(\theta(t)) \leq \chi_{1,\alpha}^2 \right\}, \tag{26}$$

where $\chi_{1,\alpha}^2$ is the upper α-quantile of the distribution of χ_1^2.

Proof The proof of Theorem 2.4 is similar to Theorem 2.3. In this case, the EL ratio defined in (7) is replaced by the adjusted empirical log-likelihood ratio defined in (12). Thus, details are omitted here. □

3 Simulation study

In this section, we conduct a simulation study to compare the performance of the proposed AEL, TEL, and TAEL-based confidence regions for the generalized Lorenz curve with EL-based confidence regions under various sample sizes in terms of coverage probabilities (CP) and mean lengths (ML) of the confidence intervals. The CP represents the proportion of times that the confidence regions contain the true value of the parameter among N simulation runs.

Since most income distributions are positively skewed, the Weibull, Chi-square, and Skew-Normal distributions appear to provide a good fit for the income data. Thus, in our simulation study, we consider that the overall distribution function $F(x)$ is:

1. Weibull distribution with shape parameter $a = 1$, scale parameter $b = 2$. The pdf of the Weibull distribution is given by

$$f_X(x) = \left(\frac{a}{b}\right)\left(\frac{x}{b}\right)^{a-1} e^{-\left(x/b\right)^a}, \quad x > 0$$

2. χ^2 distribution with $n = 3$ degrees of freedom. The pdf of the Chi-square distribution is given by

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2} e^{-x/2}, \quad x > 0$$

3. Skew-Normal distribution with location parameter $\mu = 1$, scale parameter $\sigma = 3$, and shape parameter $\lambda = 5$. The pdf of the skew-normal distribution is given by:

$$f_X(x) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution. Further, in short-hand notation, we denote the skew-normal distribution by $X \sim SN(\mu, \sigma, \lambda)$.

We choose the sample size, $n = 25, 50, 100, 150, 300$ and 500 representing a range from small to large, and values of $t_0 = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

Further, we set the nominal significance level $\alpha = 5\%$. The results are based on 10,000 iterations. To assess the performance of the proposed methods, we consider two commonly used criteria for evaluating the goodness of a confidence interval procedure. These criteria are:

1. Coverage probability: Preferably, close to 95%.
2. Mean lengths: Smaller is preferable.

First, we compute bias and mean squared error (MSE) of the estimates for the generalized Lorenz ordinates for various distributions, including Weibull (1, 2), χ^2_3, and $SN(1, 3, 5)$. The results are summarized in Table 1 and graphed in Fig. 1. It is evident that the bias of the estimate is consistently close to zero across all scenarios. Moreover, as the sample size increases, both bias and MSE generally decrease. Furthermore, it's notable that irrespective of the sample size, both bias and MSE increase as the value of t increases.

Next, we compute the coverage probability and mean lengths of the confidence regions for the generalized Lorenz ordinates $\theta(t)$. The coverage probabilities are graphed in Fig. 2. In all cases, the CP tends to increase as the sample size increases. Among all four methods, the TAEL method consistently provides the highest CP. In particular, the TAEL approach occasionally results in over-coverage issues. For

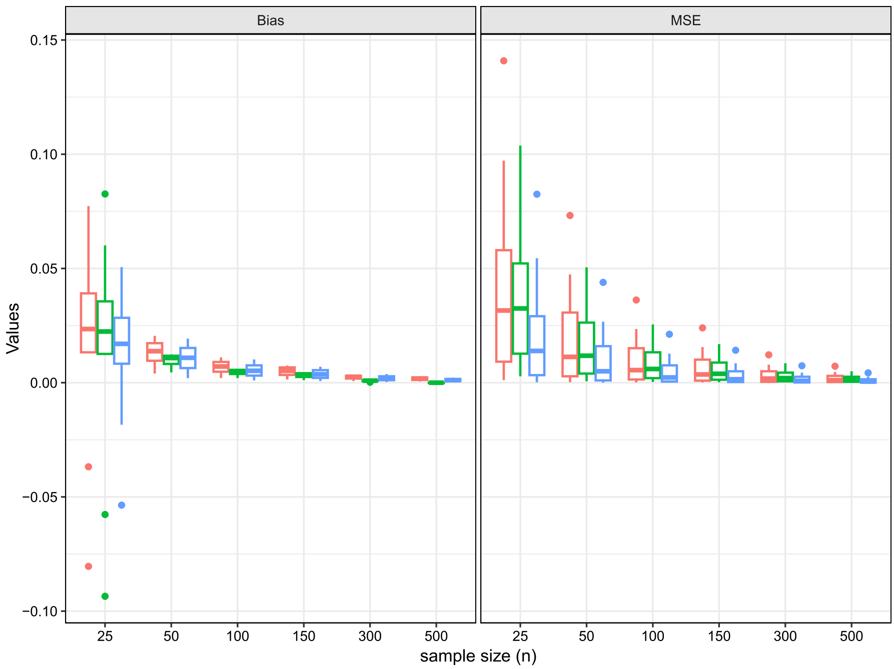

Fig. 1 Bias and MSE of the generalized Lorenz ordinate of Weibull(1, 2), χ^2_3, and $SN(1, 3, 5)$ for various sample sizes

Table 1 Bias and mean squared error (MSE) for the generalized Lorenz curve with various probability distributions

n	t	Weibull (1, 2)		χ^2_3		$SN(1, 3, 5)$	
		Bias	MSE	Bias	MSE	Bias	MSE
25	0.1	0.0095	0.0002	0.0214	0.0011	0.0356	0.0028
	0.2	0.0083	0.0007	0.0133	0.0023	0.0126	0.0039
	0.3	0.0284	0.0033	0.0489	0.0092	0.0601	0.0127
	0.4	0.0170	0.0052	0.0235	0.0129	0.0184	0.0147
	0.5	0.0506	0.0139	0.0773	0.0316	0.0826	0.0325
	0.6	0.0253	0.0191	0.0319	0.0398	0.0224	0.0365
	0.7	−0.0184	0.0291	−0.0368	0.0580	−0.0577	0.0522
	0.8	0.0328	0.0545	0.0391	0.0972	0.0236	0.0744
	0.9	−0.0536	0.0825	−0.0804	0.1409	−0.0935	0.1038
50	0.1	0.0020	0.0000	0.0040	0.0002	0.0045	0.0006
	0.2	0.0042	0.0003	0.0070	0.0010	0.0065	0.0019
	0.3	0.0064	0.0010	0.0096	0.0028	0.0082	0.0040
	0.4	0.0087	0.0024	0.0118	0.0060	0.0096	0.0072
	0.5	0.0109	0.0050	0.0138	0.0113	0.0108	0.0118
	0.6	0.0130	0.0092	0.0156	0.0192	0.0118	0.0181
	0.7	0.0152	0.0160	0.0173	0.0307	0.0123	0.0263
	0.8	0.0173	0.0267	0.0190	0.0474	0.0124	0.0369
	0.9	0.0193	0.0439	0.0205	0.0732	0.0114	0.0505
100	0.1	0.0010	0.0000	0.0020	0.0001	0.0020	0.0003
	0.2	0.0020	0.0001	0.0035	0.0005	0.0030	0.0009
	0.3	0.0031	0.0005	0.0048	0.0014	0.0038	0.0020
	0.4	0.0042	0.0011	0.0059	0.0029	0.0046	0.0036
	0.5	0.0052	0.0023	0.0071	0.0055	0.0051	0.0060
	0.6	0.0064	0.0043	0.0082	0.0093	0.0054	0.0091
	0.7	0.0076	0.0076	0.0091	0.0151	0.0057	0.0133
	0.8	0.0090	0.0127	0.0101	0.0235	0.0058	0.0186
	0.9	0.0102	0.0212	0.0111	0.0362	0.0055	0.0255
150	0.1	0.0007	0.0000	0.0014	0.0001	0.0011	0.0002
	0.2	0.0014	0.0001	0.0025	0.0003	0.0019	0.0006
	0.3	0.0021	0.0003	0.0034	0.0009	0.0025	0.0013
	0.4	0.0029	0.0007	0.0043	0.0020	0.0031	0.0024
	0.5	0.0037	0.0015	0.0052	0.0036	0.0036	0.0039
	0.6	0.0045	0.0029	0.0061	0.0062	0.0039	0.0060
	0.7	0.0055	0.0050	0.0067	0.0101	0.0041	0.0088
	0.8	0.0063	0.0085	0.0072	0.0156	0.0043	0.0123
	0.9	0.0070	0.0142	0.0075	0.0240	0.0042	0.0169
300	0.1	0.0003	0.0000	0.0007	0.0000	0.0002	0.0001
	0.2	0.0007	0.0000	0.0013	0.0002	0.0005	0.0003
	0.3	0.0011	0.0001	0.0018	0.0004	0.0007	0.0006
	0.4	0.0015	0.0004	0.0021	0.0010	0.0009	0.0012
	0.5	0.0019	0.0008	0.0026	0.0018	0.0010	0.0020
	0.6	0.0023	0.0015	0.0030	0.0031	0.0010	0.0030

Table 1 (continued)

n	t	Weibull (1, 2)		χ^2_3		$SN(1, 3, 5)$	
		Bias	MSE	Bias	MSE	Bias	MSE
	0.7	0.0028	0.0026	0.0032	0.0050	0.0011	0.0044
	0.8	0.0034	0.0044	0.0032	0.0079	0.0010	0.0062
	0.9	0.0038	0.0074	0.0030	0.0122	0.0008	0.0085
500	0.1	0.0002	0.0000	0.0005	0.0000	−0.0002	0.0001
	0.2	0.0004	0.0000	0.0009	0.0001	−0.0001	0.0002
	0.3	0.0007	0.0001	0.0012	0.0003	0.0000	0.0004
	0.4	0.0009	0.0002	0.0015	0.0006	0.0000	0.0007
	0.5	0.0011	0.0005	0.0018	0.0011	0.0000	0.0012
	0.6	0.0014	0.0009	0.0021	0.0019	0.0000	0.0018
	0.7	0.0016	0.0015	0.0023	0.0030	0.0000	0.0026
	0.8	0.0019	0.0026	0.0024	0.0047	−0.0001	0.0037
	0.9	0.0022	0.0043	0.0024	0.0072	−0.0002	0.0050

$t \geq 0.5$, the TEL method outperforms EL, and AEL performs either slightly better or on par with the EL method. However, for $t < 0.5$, the EL approach performs significantly better than AEL and slightly better than the TEL method.

When considering the mean lengths of confidence intervals, the TAEL method yields a slightly longer mean length, but it remains within an acceptable range. Among all four methods, the AEL results in the shortest confidence intervals. For $t < 0.5$, the confidence intervals based on the TEL and TAEL approaches have approximately the same mean lengths. In addition, regardless of the method, as t increases, the mean length also increases. However, as the sample size increases, the mean length of the confidence interval decreases. The mean lengths of confidence intervals are illustrated in Fig. 3.

4 Application to real data

In this section, we demonstrate the effectiveness of the proposed AEL, TEL, and TAEL methods for generalized Lorenz ordinate by constructing confidence intervals for Median Household Income in 2020. The data set is available https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/, which contains 3194 observations of the median household income in 2020, and they are grouped by state or county name. We mainly focus on examining the median income distribution of households in Arizona (AZ), California (CA), Nevada (NV), Oregon (OR), and the US as a whole. The Lorenz curves and the generalized Lorenz curves for the four states and the US are graphed in Fig. 4. The black dashed line represents the equality line. It is evident that Arizona has the Lorenz curve that is closest to the equality line. Further, the USA has the most unequal income distribution, followed by California and Nevada. We also compute a 95% confidence interval for the generalized Lorenz ordinate

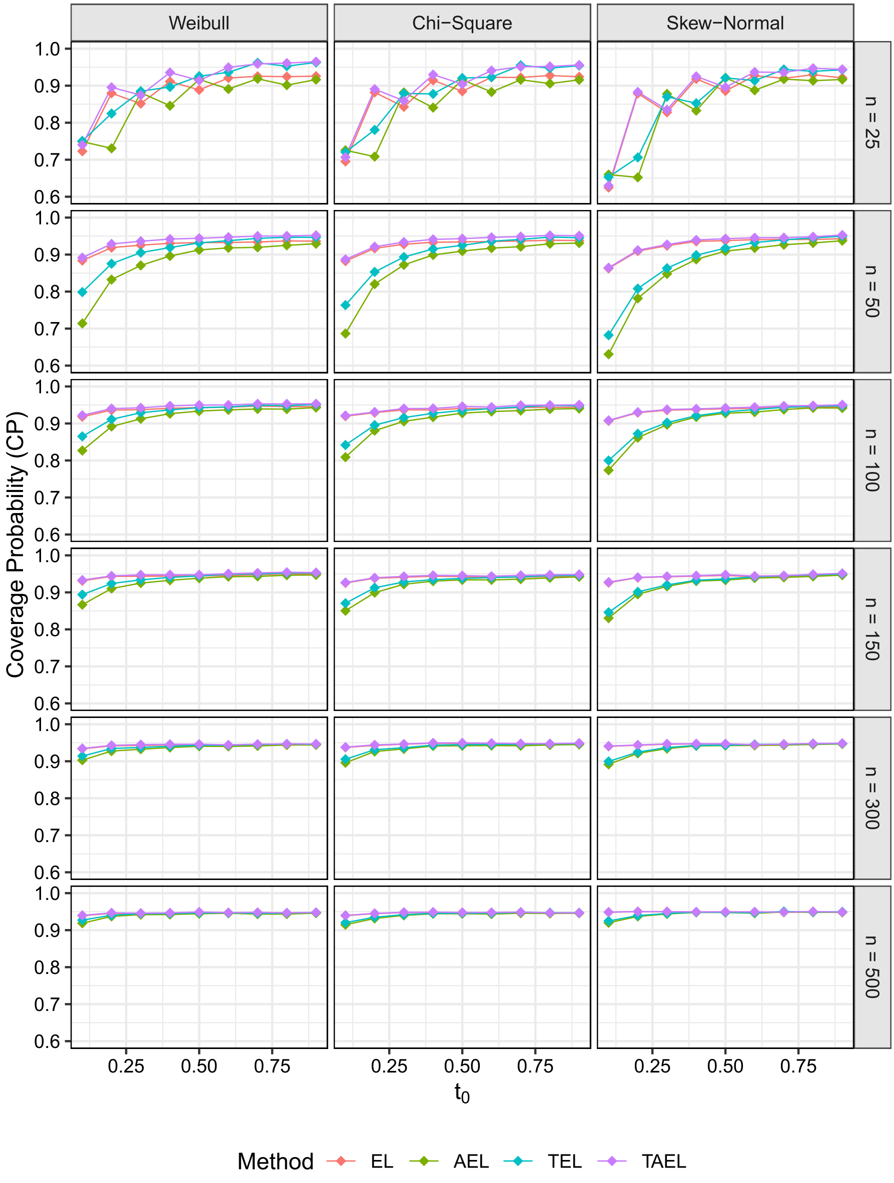

Fig. 2 Coverage probabilities of the EL, AEL, TEL, and TAEL methods with a range of t_0 values and various sample sizes for the generalized Lorenz ordinate of Weibull(1, 2), χ^2_3, and $SN(1, 3, 5)$

using EL, AEL, TEL, and TAEL methods with various t values considering $t = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. We considered all 3194 observations for the purposes of this analysis. The results are summarized in Table 2 and they are plotted in Fig. 5. We notice that EL and AEL perform similarly while

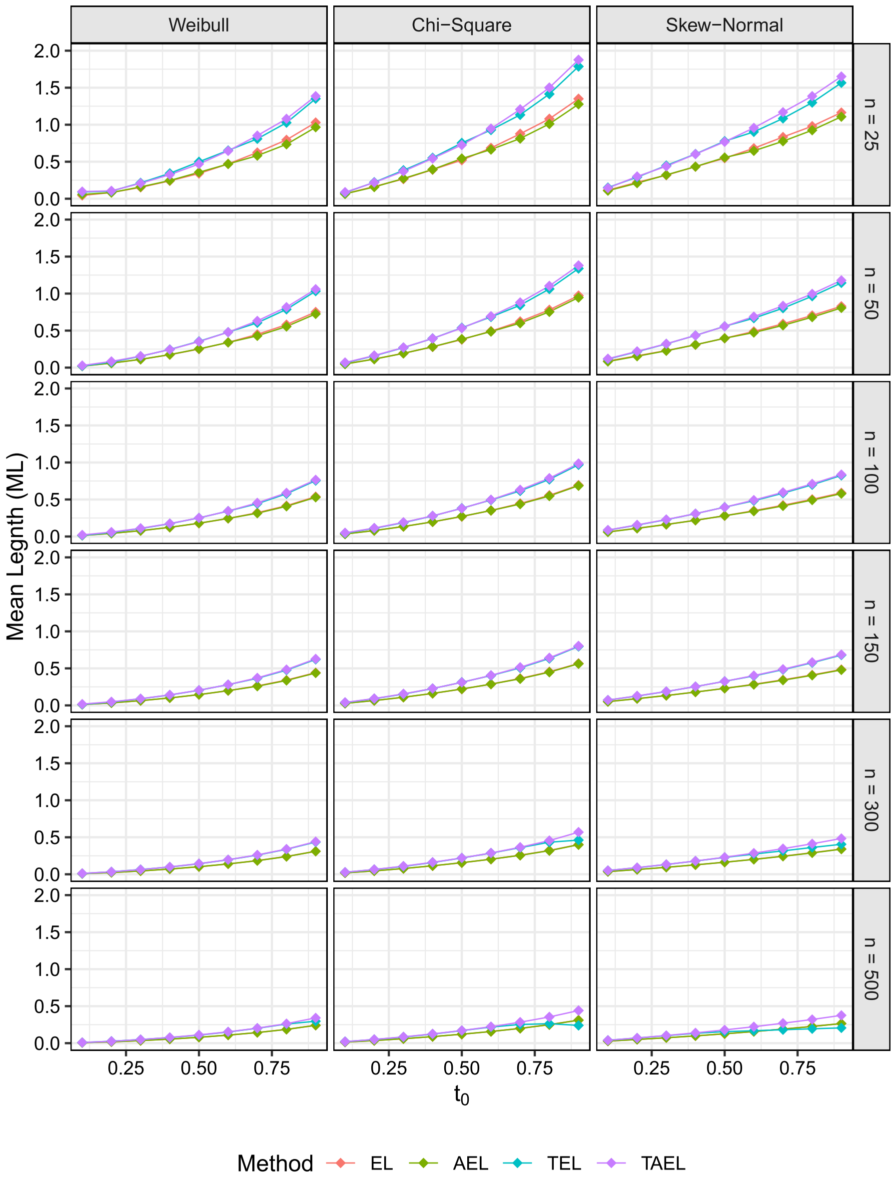

Fig. 3 Mean lengths (ML) of the confidence intervals based on the EL, AEL, TEL, and TAEL methods with a range of t_0 values and various sample sizes for the generalized Lorenz ordinate of Weibull(1, 2), χ^2_3, and $SN(1, 3, 5)$

TEL and TAEL roughly produce the same confidence intervals. Additionally, the AEL approach consistently yields a shorter confidence length than the other three methods.

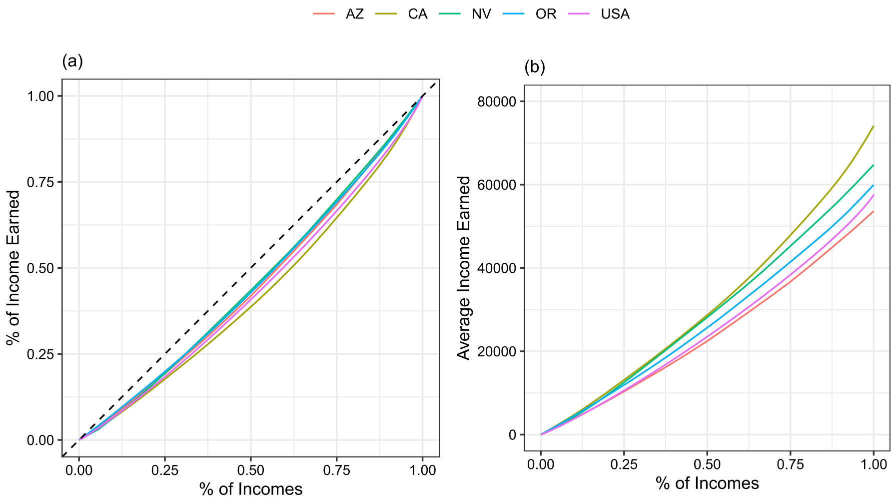

Fig. 4 **a** Lorenz Curves and **b** generalized Lorenz curves for Arizona (AZ), California (CA), Nevada (NV), Oregon (OR), and USA

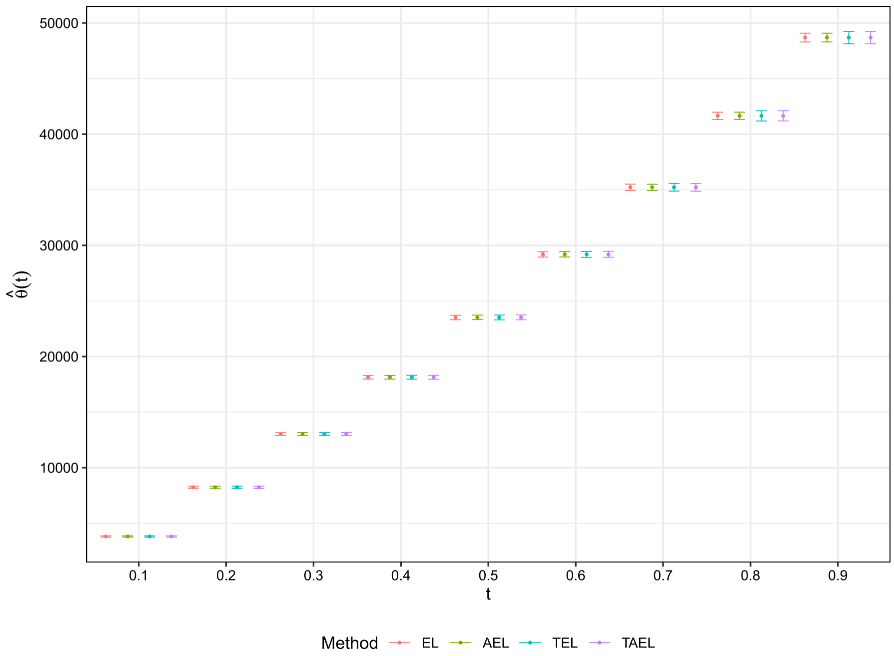

Fig. 5 A 95% confidence interval for the generalized Lorenz ordinate based on EL, AEL, TEL, and TAEL methods with various *t* values

Table 2 A 95% confidence intervals for the generalized Lorenz ordinates with various values of t for Median Household Income in 2020 in America

t	$\hat{\theta}(t)$	Method	Lower	Upper	Length
0.1	3815.195	EL	3759.1760	3871.7118	112.5535
		AEL	3757.9974	3870.5021	112.5048
		TEL	3756.8501	3871.6701	114.8200
		TAEL	3758.0284	3872.8800	114.8516
0.2	8240.058	EL	8145.7837	8334.8855	189.1324
		AEL	8143.2260	8332.2851	189.0591
		TEL	8140.5307	8335.0126	194.4819
		TAEL	8143.0881	8337.6133	194.5251
0.3	13,027.020	EL	12896.9940	13157.5745	260.6236
		AEL	12,908.4210	13,168.9741	260.5531
		TEL	12,903.4394	13,173.9962	270.5568
		TAEL	12,892.0084	13,162.6007	270.5923
0.4	18,138.190	EL	17,973.2271	18,303.5235	330.3522
		AEL	17,967.5732	18,297.8236	330.2504
		TEL	17,959.0887	18,306.3482	347.2595
		TAEL	17,964.7429	18,312.0477	347.3048
0.5	23,514.140	EL	23,312.6016	23,715.7678	403.2358
		AEL	23,322.5887	23,725.6854	403.0967
		TEL	23,307.9250	23,740.3629	432.4379
		TAEL	23,297.9302	23,730.4531	432.5230
0.6	29,186.760	EL	28,947.5843	29,425.4732	477.9736
		AEL	28,956.5766	29,434.2810	477.7044
		TEL	28,930.0349	29,460.7166	530.6816
		TAEL	28,921.0092	29,451.9423	530.9331
0.7	35,224.900	EL	34,941.7069	35,506.5735	564.9692
		AEL	34,929.3294	35,494.0013	564.6719
		TEL	34,863.9490	35,558.6065	694.6575
		TAEL	34,876.1887	35,571.3135	695.1248
0.8	41,650.550	EL	41,323.3627	41,974.1675	650.9197
		AEL	41,328.4634	41,979.0655	650.6021
		TEL	41,191.9480	42,112.0198	920.0719
		TAEL	41,186.8051	42,107.1635	920.3584
0.9	48,694.690	EL	48,301.5894	49,079.4584	777.9221
		AEL	48,304.5085	49,082.1298	777.6213
		TEL	48,139.2496	49,239.0692	1099.8196
		TAEL	48,136.2785	49,236.4482	1100.1697

5 Discussions

In this article, we proposed powerful nonparametric EL-based methods for constructing confidence intervals for generalized Lorenz ordinate. These methods include the adjusted empirical likelihood (AEL), the transformed empirical likelihood (TEL), and the transformed adjusted empirical likelihood (TAEL). We derive the limiting distributions of the generalized Lorenz ordinate based on the AEL, TEL, and TAEL methods. Simulations show that the proposed TEL and TAEL methods improve the coverage probability compared to the EL method. According to the simulation study, we highly recommend the TAEL method for $t < 0.5$ and small sample sizes. When $t \geq 0.5$, both the EL and AEL approaches yield comparable results for medium and large samples, making AEL an additional option. While the TEL method is suitable for large samples ($n \geq 300$), the TAEL method is appropriate for all sample sizes. In real-world applications, we recommend the TAEL approach as it consistently offers superior coverage compared to the other three methods. It's worth noting although the confidence intervals based on the TAEL approach are longer than others, they remain within an acceptable range. Our real-world data application demonstrates that the proposed methods are competitive with the EL method while also addressing its limitations.

Acknowledgements We would like to thank two anonymous referees for their comments, which have contributed to this improved version of the work. We also would like to express our appreciation to the Office of Student Research (OSR) at California State University, San Bernardino, for creating a supportive environment for conducting this research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Allen A (1990) Probability, statistics, and queuing theory with computer science applications, 2nd edn. Academic Press

Beach CM, Davidson R (1983) Distribution-free statistical inference with Lorenz curves and income shares. Rev Econ Stud 50(4):723–735

Belinga-Hall N (2007) Empirical likelihood confidence intervals for generalized Lorenz curve. Thesis, Georgia State University

Chang R, Halfon N (1997) Graphical distribution of pediatricians in the united states: an analysis of the fifty states and Washington, DC. Pediatrics 100:172–179

Chen J, Variyath A, Abraham B (2008) Adjusted empirical likelihood and its properties. J Comput Graph Stat 17(2):426–443

Gastwirth JL (1971) A general definition of Lorenz curve. Econometrica 39:1037–1039

Goldie CM (1977) Convergence theorems for empirical Lorenz curves and their inverses. Adv Appl Probab 9:765–791

Hasegawa H, Kozumi H (2003) Estimation of Lorenz curves: a Bayesian nonparametric approach. J Econometr 115(2):277–291

Jakobsson U (1976) On the measurement of the degree of progression. J Public Econ 5:161–168

Jing B-Y, Tsao M, Zhou W (2017) Transforming the empirical likelihood towards better accuracy. Can J Stat 45(3):340–352

John AB, Chakraborti S, Paul DT (1989) Asymptotically distribution-free statistical inference for generalized Lorenz curves. Rev Econ Stat 71(4):725–727

Kobayashi Y, Takaki H (1992) Geographic distribution of physicians in japan. Lancet 340:1391–1393

Lambert PJ (2001) The distribution and redistribution of income: a mathematical analysis, 2nd edn. Manchester University Press, Manchester

Li M, Ratnasingam S, Ning W (2022) Empirical-likelihood-based confidence intervals for quantile regression models with longitudinal data. J Stat Comput Simul 92(12):2536–2553

Lorenz MC (1905) Method of measuring the concentration of wealth. J Am Stat 9:209–219

Luo S, Qin G (2019) Jackknife empirical likelihood-based inferences for Lorenz curve with kernel smoothing. Commun Stat-Theory Methods 48(3):559–582

Marshall W, Olkin I (1979) Inequalities: theory of majorization and its applications. Academic Press, New York

Mosler K, Koshevov G (2007) Multivariate Lorenz dominance based on Zonoids. ASTA: Adv Stat Anal 91:57–76

Owen AB (2001) Empirical likelihood. Chapman & Hall, New York

Qin G, Yang B, Belinga-Hall N (2013) Empirical likelihood-based inferences for the Lorenz curve. Ann Inst Stat Math 65:1–21

Ratnasingam S, Ning W (2022) Confidence intervals of mean residual life function in length-biased sampling based on modified empirical likelihood. J Biopharm Stat. https://doi.org/10.1080/10543406.2022.2089157

Ryu HK, Slottje DJ (1996) Two flexible functional form approaches for approximating the Lorenz curve. J Econometr 72:251–274

Sen A (1973) On economic inequality. Norton, New York

Shi Y, Liu B, Qin G (2019) Influence function-based empirical likelihood and generalized confidence intervals for the Lorenz curve. Stat Methods Appl 29:427–446

Stewart P, Ning W (2020) Modified empirical likelihood-based confidence intervals for data containing many zero observations. Comput Stat 35(4):2019–2042

Tsao M (2013) Extending the empirical likelihood by domain expansion. Can J Stat 41(2):257–274

Yang BY, Qin GS, Belinga-Hill NE (2012) Non-parametric inferences for the generalized Lorenz curve. Sci Sin Math 42(3):235–250