5-2023

Jackknife Empirical Likelihood Tests for Equality of Generalized Lorenz Curves

Anton Butenko

Jackknife Empirical Likelihood Tests for Equality of Generalized Lorenz Curves

---

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

Anton Butenko

May 2023

Jackknife Empirical Likelihood Tests for Equality of Generalized Lorenz Curves

---

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Anton Butenko

May 2023

Approved by:

Dr. Suthakaran Ratnasingam, Committee Chair

Dr. Jeremy Aikin, Committee Member

Dr. Hani Aldirawi, Committee Member

Dr. Madeleine Jetter, Chair, Department of Mathematics

Dr. Corey Dunn, Graduate Coordinator

Abstract

A Lorenz curve is a graphical representation of the distribution of income or wealth within a population. The generalized Lorenz curve can be created by scaling the values on the vertical axis of a Lorenz curve by the average output of the distribution. In this thesis, we propose two nonparametric methods for testing the equality of two generalized Lorenz curves. Both methods are based on empirical likelihood and utilize a U-statistic. We derive the limiting distribution of the likelihood ratio, which is shown to follow a chi-squared distribution with one degree of freedom. We conduct simulations to compare the proposed methods and an existing method by examining Type I error rates and power across various sample sizes and distribution assumptions. Our results show that the proposed methods exhibit superior performance in finite samples, particularly in small sample sizes, and are robust across various scenarios. Finally, we use real-world data to illustrate the methods of testing two generalized Lorenz curves.

Acknowledgements

I would like to express my sincere gratitude to everyone who supported me during my master's thesis journey. First and foremost, I thank my thesis advisor, Dr. Ratnasingam, for his invaluable guidance, feedback, and encouragement. I also thank my committee members, Dr. Aikin and Dr. Aldirawi, for their time and support. I am grateful to the program coordinator, Dr. Dunn, for his assistance and support throughout the program. Last but not least, I thank my family for their unconditional love and understanding. They always believed in me and cheered me on, even when I was stressed and overwhelmed. I dedicate this thesis to them.

Table of Contents

Abstract **iii**

Acknowledgements **iv**

List of Figures **vii**

1 Introduction **1**

1.1 Lorenz Curves and Generalized Lorenz Curves 1

1.2 Statistical Inferences about Generalized Lorenz Curves 2

1.3 Motivation 3

2 Definitions and Preliminaries **5**

2.1 Generalized Lorenz Curves 5

2.2 Empirical Likelihood (EL) 6

2.3 Jackknife Empirical Likelihood (JEL) 7

2.4 Adjusted Empirical Likelihood (AEL) 8

3 JEL-based Tests for the Equality of Two Generalized Lorenz Curves **10**

3.1 JEL Test for the Equality of Two Generalized Lorenz Curves 10

3.2 AJEL Test for the Equality of Two Generalized Lorenz Curves 13

4 A Simulation Study **14**

4.1 Probability of Type I Error Analysis 14

4.2 Power Analysis 20

5 Real Data Applications **26**

5.1 Using Complete Data to Compare Income Distributions of Faculty at CSU Monterey Bay and CSU San Bernardino 27

5.2 Using Complete Data to Compare Income Distributions of Faculty at CSU San Bernardino and CSU San Francisco 28

5.3 Using Incomplete Data to Compare Income Distributions of All Faculty at CSU and UC 30

5.4 Using Incomplete Data to Compare Income Distributions of CSUSB Faculty between 2009 and 2020 31

6 Discussion and Conclusion **35**

Bibliography **40**

List of Figures

4.1 Type I error comparison for ADF, JEL, and AJEL methods for different distributions, sample sizes, and values of t 19
4.2 Generalized Lorenz curves for Chi-Square, Exponential, and Half-Normal distributions 20
4.3 Power comparison for ADF, JEL, and AJEL methods for different distributions, sample sizes, and values of t 25

5.1 (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSU Monterey Bay and CSU San Bernardino faculty 27
5.2 (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSU San Bernardino and CSU San Francisco faculty 29
5.3 (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSU and UC faculty 30
5.4 (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSUSB faculty in years 2009 and 2020 32
5.5 (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSUSB faculty in years 2009 and 2020 33

Chapter 1

Introduction

1.1 Lorenz Curves and Generalized Lorenz Curves

A Lorenz curve is a visual representation of an income or wealth distribution within a population. It is named after American economist Max Lorenz [Lor05] who developed it in 1905. The curve uses the percentiles of the population regarding the income or wealth as an input and the percent of the cumulative income or wealth attributed to a given percentile as the output. For example, in a population with a constant income or wealth per capita, the Lorenz curve will be a piece of the graph of the identity function, $y = x$, on the restricted domain $[0, 1]$, i.e., a 45-degree line segment starting at $(0, 0)$ and ending at $(1, 1)$. This line is commonly referred to as the line of equality and all other Lorenz curves fall under the line of equality creating a gap that visually represents the inequality in the population. For example, in a population with a uniformly distributed income (i.e., all incomes are equally likely to be between some values a and b) the Lorenz curve will be a piece of the standard parabola, $y = x^2$, on the restricted domain $[0, 1]$. The area of the region bounded by a given Lorenz curve and the line of equality divided by the area under the line of equality is called the Gini coefficient. The Gini coefficient is used to summarize the inequality as a single number which can be used to characterize the given population or to compare two different populations. The generalized Lorenz curve can be constructed from a Lorenz curve by scaling the values on the vertical axis by the average output of the distribution. While Lorenz curves are frequently used in economics to represent financial inequality, they also can be used in other fields of study to visualize the inequality of the distribution within any system. For example, the Lorenz curve has been used by several researchers to analyze physician distributions. Chang and Halfon [CH97] examined variations in the distribution of pediatricians among the states between 1982 and 1992 using Lorenz curves and Gini indices. Kobayashi and Takaki [KT92] used the Lorenz curve and the Gini coefficient to study the disparity in physician distribution in Japan.

1.2 Statistical Inferences about Generalized Lorenz Curves

Since the actual distribution is very likely to be unknown, the Lorenz curves must be constructed from incomplete data by making a statistical inference. In regard to making comparisons between two Lorenz curves, there have been multiple studies. Arora and Jain [AJ06] investigated the generalized Lorenz dominance and proposed tests for the equality of two generalized Lorenz curves over a specified interval. Li and Wei [LW18] noted that normal approximation-based methods may have poor performance, especially for the skewed income data, or the limiting distributions are nonstandard and bootstrap calibrations are needed hence more effective inferences for Lorenz curves are desirable. Empirical likelihood (EL) introduced by Owen [Owe88] is a nonparametric method that requires fewer assumptions for utilizing the likelihood ratio approach while preserving many of its appealing features such as its extension of Wilk's theorem, asymmetric confidence intervals, better coverage for small sample sizes, etc. However, it has some computational complications when using nonlinear statistics as demonstrated by Jing, Yuan, and Zhou [JYZ09] and when solutions to the corresponding constraints do not exist as demonstrated by Chen, Variyath, and Abraham [CVA08]:

- Jing et al. [JYZ09] provides an example of the EL-based approach losing its appeal when using nonlinear $U-$statistics of degree $m \geq 2$, due to the increased computational difficulty of solving a system of nonlinear equations simultaneously using Lagrange multipliers. To overcome this difficulty, Jing et al. [JYZ09] proposed the jackknife empirical likelihood (JEL) approach. In summary, JEL turns the statistic of interest into a sample mean based on jackknife pseudo-values [Que56] which are asymptotically independent under mild conditions [Shi84], and then consecutively, Owen's EL method can be applied resulting in a simpler system of equations.

- Chen et al. [CVA08] pointed out that under certain conditions it can be difficult to determine the parameter region over which the likelihood ratio function is well-defined, making it a challenge to identify the maximum likelihood ratio or to find a proper initial value. To overcome this challenge, Chen et al. [CVA08] proposed the adjusted empirical likelihood (AEL) approach to eliminate the outlined problem. In summary, the AEL extends the convex hull to include the origin by adding a pseudo-value. With this adjustment, the empirical likelihood is well-defined for all parameter values, thus finding the maximum becomes a much simpler problem.

Multiple studies have been conducted on EL for the Lorenz curve by various researchers. For instance, Belinga-Hall [BH07] and Yang et al. [YQBH12] developed plug-in empirical likelihood-based inferences to construct confidence intervals for the generalized Lorenz curve. Most recently, Ratnasingam et al. [RWAR23] developed three nonparametric EL-based methods to construct confidence intervals for the generalized Lorenz curve using adjusted empirical likelihood (AEL), transformed empirical likelihood (TEL), and transformed adjusted empirical likelihood (TAEL). Moreover, several studies have focused on comparing two Lorenz curves. For example, Arora and Jain [AJ06] investigated the generalized Lorenz dominance and proposed tests for the equality of two generalized Lorenz curves over a specified interval. Li and Wei [LW18] noted that normal approximation-based methods may have poor performance, especially for the skewed income data, or the limiting distributions are nonstandard and bootstrap calibrations are needed hence more effective inferences for Lorenz curves are desirable. All of these tests were parametric and they involve making assumptions about the underlying distribution of the data. Xu [Xu97] proposed an asymptotically distribution-free statistical test (ADF) to evaluate the equality of two generalized Lorenz curves and showed that the test statistic follows the weighted sum of χ^2 with different degrees of freedom.

1.3 Motivation

To the best of our knowledge, no previous studies have investigated testing the equality of two Lorenz curves based on JEL using a U-statistic. Thus, in this thesis, to test the equality of two generalized Lorenz curves, we propose two new nonparametric approaches using a $U-$statistic based on the jackknife empirical likelihood method and its extension to the adjusted jackknife empirical likelihood. The new methods combine two of the EL-based approaches mentioned above - the JEL as described by Jing et al. [JYZ09] and the AEL as described by Cheng et al. [CVA08].

The rest of the thesis is organized as follows. In Chapter 2, we provide a list of definitions and briefly describe the fundamental properties of the EL, JEL, and AEL methods. In Chapter 3, we propose two new approaches to test the equality of two generalized Lorenz curves. Simulation studies are conducted in Chapter 4 to evaluate the performance of the proposed methods. In Chapter 5, we apply the proposed methods to real data sets to demonstrate the procedures. We discuss our results and draw conclusions in Chapter 6. All proofs are deferred to the appendix.

Chapter 2

Definitions and Preliminaries

2.1 Generalized Lorenz Curves

Lorenz curves were first introduced by Max Lorenz [Lor05] as a graphical representation of the distribution of a variable, such as income or wealth, in a population. It is used to measure the degree of inequality in a society or group. The Lorenz curve is constructed by plotting the cumulative percentage of the population on the x-axis against the cumulative percentage of the variable (such as income or wealth) on the y-axis. The resulting curve represents the distribution of the variable in the population. A Lorenz curve that is close to the line of equality (which is a straight line that represents perfect equality) indicates that the distribution of the variable is relatively equal across the population. On the other hand, a Lorenz curve that is farther away from the line of equality indicates a higher degree of inequality in the distribution of the variable. Following [Gas71], a general definition of the Lorenz curve is given as

$$\eta(t) = \frac{1}{\mu} \int_0^{\psi_t} x dF(x), \quad t \in [0, 1] \tag{2.1}$$

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 proportion of cumulative income of the lowest t-th quantile of households. Similarly, the generalized Lorenz curve is defined by

$$\xi(t) = \int_0^{\psi_t} x dF(x), \quad t \in [0, 1] \tag{2.2}$$

where $\psi_t = F^{-1}(t) = \inf\{x : F(x) \geq t\}$ is the $t-$th quantile of F. For a fixed $t \in [0, 1]$, the generalized Lorenz ordinate $\xi(t)$ is the average income of the lowest t-th quantile of households.

2.2 Empirical Likelihood (EL)

Let $X_1, X_2, \cdots, X_n$ be a random sample from a population with the cumulative distribution $F(x)$, then the Empirical cumulative distribution function (ECDF) is defined as

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

where

$$I(X_i \leq x) = \begin{cases} 1, & X_i \leq x \\ 0, & X_i > x \end{cases}$$

The nonparametric likelihood of distribution function F is defined as

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_{i^-}))$$

where $F(X_i)$ denotes the probability $P(X_i \leq x)$ and $F(X_{i^-})$ denotes the probability $P(X_i < x)$ so that $F(X_i) - F(X_{i^-}) = P(X_i = x)$. It can be easily shown [Owe88] that ECDF is the nonparametric maximum likelihood estimate of $F(x)$. For a distribution $F(x)$, the Empirical likelihood ratio is defined as

$$R(F) = \frac{L(F)}{L(F_n)}$$

Similar to how parametric likelihood ratios are used to construct confidence intervals and hypothesis tests, the Empirical likelihood ratios can be used as a basis for statistical inferences but with fewer assumptions. Thus, if we are interested in some parameter θ defined by the estimating function $g(X;\theta)$, then the profile Empirical likelihood function for θ is

$$L(\theta) = \sup\left\{\prod_{i=1}^{n} p_i \middle| \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \sum_{i=1}^{n} p_i g(X_i;\theta) = 0\right\}$$

where p_i is the probability of observing X_i. Similarly, we define the profile empirical likelihood ratio function as

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \sup\left\{\prod_{i=1}^{n} np_i \middle| \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \sum_{i=1}^{n} p_i g(X_i;\theta) = 0\right\} \tag{2.3}$$

Therefore, the profile empirical log-likelihood ratio function, evaluated at θ is defined as

$$W(\theta) = \sup \left\{ \sum_{i=1}^{n} \log(np_i) \middle| \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \sum_{i=1}^{n} p_i g(X_i;\theta) = 0 \right\}$$

Calculating $R(\theta)$ and $W(\theta)$ is equivalent to solving an optimization problem with constraints which can be done using the Lagrange multipliers method. It can be easily shown [Owe88] that $W(\theta)$ reaches its maximum when

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda g(X_i;\theta)}$$

where λ is the Lagrange multiplier that solves the equation

$$\sum_{i=1}^{n} \frac{g(X_i;\theta)}{1 + \lambda' g(X_i;\theta)} = 0$$

When λ is obtained, $W(\theta)$ can be computed

$$W(\theta) = 2 \sum_{i=1}^{n} \log\left[1 + \lambda g(X_i;\theta)\right]$$

It can be easily shown [Owe88] that when $g(X,\mu) = X - \mu$, a $(1-\alpha)100\%$ confidence region for μ can be constructed using

$$C = \left\{ \mu \middle| W(\mu) \leq \chi_q^2(\alpha) \right\}$$

where C is a convex set and q is the dimension of the set of θ values.

2.3 Jackknife Empirical Likelihood (JEL)

Jing et al. [JYZ09] provides an example of the EL-based approach losing its appeal when using nonlinear $U-$statistics of degree $m \geq 2$, due to the increased computational difficulty of solving a system of nonlinear equations simultaneously using Lagrange multipliers. To overcome this difficulty, Jing et al. [JYZ09] proposed the Jackknife Empirical Likelihood (JEL) approach.

Let $Z_1, \ldots, Z_n$ be a random sample of n independent but possibly not identically distributed observations and let $T_n = T(Z_1, \ldots, Z_n)$ be consistent estimator of the parameter θ. We define the jackknife pseudo-values as

$$\hat{V}_i = nT_n - (n-1)T_{n-1}^{(-i)}$$

where $T_{n-1}^{(-i)} := T(Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n)$ is the statistic T_{n-1} computed on the sample of $n-1$ observations formed from the original data set by deleting the i-th observation. We define the jackknife estimator of θ as the average of the pseudo-values

$$\hat{T}_{n,jack} := \frac{1}{n}\sum_{i=1}^{n} \hat{V}_i$$

Jing et al. [JYZ09] showed that for one or two-sample U-statistic, the θ's estimators T_n and $\hat{T}_{n,jack}$ coincide, that is

$$T_n = \frac{1}{n}\sum_{i=1}^{n} \hat{V}_i$$

Since Owen's empirical likelihood is particularly easy to apply for the sample mean, Jing, Tsao, and Zhou [JTZ17] proposed the Jackknife Empirical Likelihood (JEL) approach in which the EL is applied to the jackknife pseudo-values $\hat{V}_i$'s.

2.4 Adjusted Empirical Likelihood (AEL)

Chen et al. [CVA08] pointed out that under certain conditions it can be difficult to determine the parameter region over which the likelihood ratio function is well-defined, making it a challenge to identify the maximum likelihood ratio or to find a proper initial value. To overcome this challenge, Chen et al. [CVA08] proposed the Adjusted Empirical Likelihood (AEL) approach to eliminate the outlined problem.
For any given θ, denote $g_i = g_i(\theta) = g(X_i;\theta)$ and $\bar{g}_n = \bar{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i$. For some positive constant a_n, define

$$g_{n+1} := g_{n+1}(\theta) = -\frac{a_n}{n}\sum_{i=1}^{n} g_i = -a_n\bar{g}_n$$

The adjusted profile empirical log-likelihood ration function is now

$$W^*(\theta) = \sup\left\{\sum_{i=1}^{n+1}\log((n+1)p_i)\middle|\sum_{i=1}^{n+1}p_i = 1, p_i \geq 0, \sum_{i=1}^{n+1}p_i g(X_i;\theta) = 0\right\}$$

Since the convex hull of $\{g_i, i = 1, 2, \ldots, n, n+1\}$ for any given θ contains 0, $W^*(\theta)$ is well defined without exceptions. The value of a_n should be chosen to fit the problem of the user's particular application and the general recommendation is to have $a_n = \max(1, \log(n)/2)$ coupled with trimmed version of $\bar{g}_n$ when appropriate.

Chapter 3

JEL-based Tests for the Equality of Two Generalized Lorenz Curves

In this chapter, we develop two new testing procedures using jackknife EL methods.

3.1 JEL Test for the Equality of Two Generalized Lorenz Curves

Let $X_1, X_2, \cdots, X_{n_1}$ and $Y_1, Y_2, \cdots, Y_{n_2}$ be two random samples from two independent populations. The generalized Lorenz curve for these two samples are

$$\eta_1(t) = \int_0^{\psi_t} x dF(x), \quad t \in [0,1] \tag{3.1}$$

and

$$\eta_2(t) = \int_0^{\psi_t} y dF(y), \quad t \in [0,1] \tag{3.2}$$

where $\psi_t = F^{-1}(t) = \inf\{x : F(x) \geq t\}$ is the $t-$th quantile of F. We are interested in testing the following hypotheses.

$$H_0 : \eta_1(t) = \eta_2(t) \quad \text{vs} \quad H_1 : \eta_1(t) \neq \eta_2(t) \tag{3.3}$$

From the definition of the generalized Lorenz curve, it can be clearly seen that

$$E[X\, I(X \leq \psi_t)] - \eta_1(t) = 0.$$

and

$$E[Y\, I(Y \leq \psi_t)] - \eta_2(t) = 0.$$

As a result, the generalized Lorenz ordinates $\eta_1(t)$ and $\eta_2(t)$ are the means of the random variable X and Y truncated at ψ_t respectively. Let's consider the kernel function,

$$h(X, Y) = X\, I(X \leq \psi_t) I(Y \leq \psi_t) - Y\, I(X \leq \psi_t) I(Y \leq \psi_t) \tag{3.4}$$

We can easily show that $\theta(t) \equiv E\big[h(X_i, Y_j)\big] = \big(\eta_1(t) - \eta_2(t)\big) P\big(X \leq \psi_t\big) P\big(Y \leq \psi_t\big)$. Thus, we are interested in testing

$$H_0 : \theta(t) = 0 \quad \text{vs} \quad H_1 : \theta(t) \neq 0. \tag{3.5}$$

Now consider, the two-sample U-statistics of degree (1,1) with the kernel h is given by,

$$\begin{aligned} U_{n_1,n_2} &= \frac{1}{n_1}\frac{1}{n_2} \sum_{1\leq i\leq n_1} \sum_{1\leq j\leq n_2} h(X_i, Y_j) \\ &= \frac{1}{n_1}\frac{1}{n_2} \sum_{1\leq i\leq n_1} \sum_{1\leq j\leq n_2} X_i\, I(X_i \leq \psi_t) I(Y_j \leq \psi_t) - Y_j\, I(X_i \leq \psi_t) I(Y_j \leq \psi_t) \end{aligned} \tag{3.6}$$

Let $n = n_1 + n_2$. We can write the U-statistics

$$U_{n_1,n_2}(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}) = U_n(Z_1, Z_2, \dots, Z_n) \tag{3.7}$$

where

$$Z_k = \begin{cases} X_k & k = 1, 2, \dots, n_1 \\ Y_{k-n_1} & k = n_1 + 1, \dots, n \end{cases}$$

We define the corresponding jackknife pseudo-values by

$$\widehat{V}_k = nU_n - (n-1)U_{n-1}^{-k}, \quad k = 1, 2, \cdots n, \tag{3.8}$$

where $U_{n-1}^{-k} = U_n\big(Z_1, Z_2, \dots, Z_{k-1}, Z_{k+1}, \dots, Z_n\big)$. Further, the jackknife estimator of θ is $n^{-1}\sum_{k=1}^{n} \widehat{V}_k$. In particular, under mild conditions, the $\widehat{V}_k$'s are asymptotically independent. For more details, readers are referred to Shi's paper [Shi84]. Thus, we can use the EL approach to the $\widehat{V}_k$'s. It should be noted that $\widehat{V}_k(t)$ is the function of t and can be calculated at a fixed value t_0 such that $t_0 \in [0, 1]$. For the simplicity of notations we use $\widehat{V}_k$ instead of $\widehat{V}_k(t)$. The JEL for $\theta(t)$ is defined as follows:

$$L(\theta(t)) = \sup_{\mathbf{p}} \left\{ \prod_{k=1}^{n} p_k : \sum_{k=1}^{n} p_k = 1, \sum_{k=1}^{n} p_k\big(\widehat{V}_k - \mathbf{E}\widehat{V}_k\big) = 0 \right\} \tag{3.9}$$

where $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is a probability vector satisfying $\sum_{k=1}^{n} p_k = 1$ and $p \geq 0$ for all k, and $\mathbf{E}\widehat{V}_k$ can be determined using the equation (14) in [JTZ17]. Note that $\prod_{k=1}^{n} p_k$, subject to $\sum_{k=1}^{n} p_k = 1$, attains its maximum n^{-n} at $p_k = n^{-1}$. Thus, the JEL ratio for $\theta(t)$ is given as

$$\mathcal{R}(\theta(t)) = \sup\left\{ \prod_{k=1}^{n} np_k : \sum_{k=1}^{n} p_k = 1, \sum_{k=1}^{n} p_k\big(\widehat{V}_k - \mathbf{E}\widehat{V}_k\big) = 0 \right\} \tag{3.10}$$

Further, under null hypothesis $H_0 : \theta(t) = 0$, the JEL ratio becomes,

$$\mathcal{R}(0) = \sup\left\{ \prod_{k=1}^{n} np_k : \sum_{k=1}^{n} p_k = 1, \sum_{k=1}^{n} p_k \widehat{V}_k = 0 \right\}. \tag{3.11}$$

Using the Lagrange multiplier method, we have

$$p_k = \frac{1}{n}\left\{1 + \lambda \widehat{V}_k\right\}^{-1}, \quad k = 1, \ldots, n.$$

where λ is the solution to

$$\frac{1}{n}\sum_{k=1}^{n} \frac{\widehat{V}_k}{1 + \lambda \widehat{V}_k} = 0.$$

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

$$\ell(\theta(t)) = -2\log \mathcal{R}(\theta(t)) = 2\sum_{k=1}^{n} \log\left\{1 + \lambda \widehat{V}_k\right\} \tag{3.12}$$

Let $h_{1,0}(x) = \mathbf{E}h(x, Y_1)$, $\sigma_{1,0}^2 = Var\big(h_{1,0}(X_1)\big)$, $h_{0,1}(y) = \mathbf{E}h(X_1, y)$, and $\sigma_{0,1}^2 = Var\big(h_{0,1}(Y_1)\big)$. We have the following theorem for the JEL.

Theorem 3.1. *Assume that*

1. $E(X^2) < \infty$, *and* $E(Y^2) < \infty$
2. $\sigma_{1,0}^2 > 0$, *and* $\sigma_{0,1}^2 > 0$
3. $n_1/n_2 \longrightarrow r$, *where* $0 < r < \infty$

For any given $t = t_0 \in (0, 1)$, *the limiting distribution of* $\ell(\theta(t_0))$ *is a chi-square distribution with one degree of freedom,*

$$\ell(\theta(t_0)) \longrightarrow \chi_1^2, \quad as \ \min(n_1, n_2) \longrightarrow \infty. \tag{3.13}$$

Proof. Proof of Theorem 3.1 is given in Appendix. □

3.2 AJEL Test for the Equality of Two Generalized Lorenz Curves

Further, Chen et al. [CVA08] proposed the AEL method by adding a pseudo-observation to the data set. This method bypasses the convex hull constraint and ensures a solution at any parameter point. By adopting this idea, we extend the proposed JEL method by employing the adjusted jackknife empirical likelihood (AJEL) to examine the equality of two Lorenz curves. The AJEL for $\theta(t)$ is defined as follows:

$$L^{\text{Adj}}(\theta(t)) = \sup_{\mathbf{p}} \left\{ \prod_{k=1}^{n+1} p_k^{\text{Adj}} : \sum_{k=1}^{n+1} p_k^{\text{Adj}} = 1, \sum_{k=1}^{n+1} p_k^{\text{Adj}} g_k^{\text{Adj}}(t) = 0 \right\} \tag{3.14}$$

where $g_k^{\text{Adj}}(t) = \widehat{V}_k - \mathbf{E}\widehat{V}_k$, $k = 1, \ldots, n$, and $g_{n+1}^{\text{Adj}} = -a_n \bar{g}_n(t) = -\frac{a_n}{n} \sum_{i=1}^{n} g_i^{\text{Adj}}(t)$. As recommended by Chen et al. [CVA08], $a_n = \max\{1, \log(n)/2\}$. Using the Lagrange multiplier method, we can determine $L^{\text{Adj}}(\theta(t))$ as follows.

$$p_k^{\text{Adj}} = \frac{1}{n+1} \left\{ 1 + \lambda^{\text{Adj}}(t) g_k^{\text{Adj}}(t) \right\}^{-1}, \quad k = 1, \ldots, n+1.$$

where λ^{Adj} is the solution to

$$\frac{1}{n+1} \sum_{k=1}^{n+1} \frac{g_k^{\text{Adj}}(t)}{1 + \lambda^{\text{Adj}}(t) g_k^{\text{Adj}}(t)} = 0.$$

Note that $\prod_{k=1}^{n+1} p_k^{\text{Adj}}$, subject to $\sum_{k=1}^{n+1} p_k^{\text{Adj}} = 1$, attains its maximum $(n+1)^{-n-1}$ at $p_k = (n+1)^{-1}$. Thus, the AJEL ratio for $\theta(t)$ is given as

$$\mathcal{R}^{\text{Adj}}(\theta(t)) = \prod_{k=1}^{n+1} (n+1) p_k^{\text{Adj}} = \prod_{k=1}^{n+1} \left\{ 1 + \lambda^{\text{Adj}}(t) g_k^{\text{Adj}}(t) \right\}^{-1} \tag{3.15}$$

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

$$\ell^{\text{Adj}}(\theta(t)) = -2 \log \mathcal{R}^{\text{Adj}}(\theta(t)) = 2 \sum_{k=1}^{n+1} \log \left\{ 1 + \lambda^{\text{Adj}}(t) g_k^{\text{Adj}}(t) \right\} \tag{3.16}$$

Theorem 3.2. *Under the same conditions of Theorem 3.1 and for any given $t = t_0 \in (0,1)$, the limiting distribution of $\ell^{\text{Adj}}(\theta(t_0))$ is a chi-square distribution with one degree of freedom,*

$$\ell^{\text{Adj}}(\theta(t_0)) \longrightarrow \chi_1^2 \quad \text{as } \min(n_1, n_2) \longrightarrow \infty. \tag{3.17}$$

Proof. Proof of Theorem 3.2 is given in Appendix. □

Chapter 4

A Simulation Study

In this chapter, we conduct a simulation study to evaluate the performance of the proposed testing methods, JEL and AJEL, and compare the results with the ADF. In our simulation analysis, we chose Chi-Square, Exponential, and Half-Normal distributions as the overall distribution function $F(x)$ because the majority of income distributions are positively skewed. Under the null hypothesis, we examine the distributions of χ^2_4, $Exp(4)$, and $HN(1)$ using different sample sizes (n_1, n_2) such as $(20, 30), (40, 50), (75, 75)$, and $(100, 100)$ for each of the methods.

4.1 Probability of Type I Error Analysis

We first assess the Type I error probabilities of the ADF, JEL and AJEL methods with a nominal level of $\alpha = 0.05$. Tables 4.1-4.3 provide a summary of the findings, including the probabilities of Type I errors (TE) and their corresponding standard errors (SE), whereas Figure 4.1 depicts the outcomes graphically. The JEL method appears to perform slightly better or similar to the AJEL method. For instance, when using the χ^2_4 distribution with sample sizes of (20, 30) at $t = 0.1$, the Type I error probability for the ADF method is 0.007 with a standard error of 0.0027, the JEL method is 0.027 with a standard error of 0.0051, and the AJEL method has a probability of 0.075 and a standard error of 0.0083. When using the χ^2_4 test with sample sizes of 20 and 30, the AJEL method produces a Type I error rate that is slightly higher than the expected level. The ADF method comes next, with the JEL method following. However, when testing for the $Exp(1)$ distribution, the ADF method results in a Type I error rate that is much lower than the expected level, and the test becomes more conservative for $t > 0.2$. When using the $HN(1)$ distribution, the JEL method performs the best among the three methods, while the ADF method performs the worst for all sample sizes. The Type I error probabilities are slightly above the nominal level for small sample sizes, but improve for larger sample sizes and remain within an acceptable range.

Table 4.1: Type I error (TE) and standard error (SE) comparison of ADF, JEL, and AJEL tests with nominal level $\alpha = 0.05$ when $X, Y \sim \chi_4^2$

		ADF		JEL		AJEL	
(n_1, n_2)	t	TE	SE	TE	SE	TE	SE
(20, 30)	0.1	0.007	0.0027	0.027	0.0051	0.075	0.0083
	0.2	0.011	0.0033	0.019	0.0043	0.058	0.0074
	0.3	0.027	0.0051	0.018	0.0042	0.062	0.0076
	0.4	0.037	0.0060	0.017	0.0041	0.065	0.0078
	0.5	0.069	0.0080	0.017	0.0041	0.065	0.0078
	0.6	0.062	0.0076	0.059	0.0075	0.086	0.0089
	0.7	0.069	0.0080	0.052	0.0070	0.070	0.0081
	0.8	0.071	0.0081	0.046	0.0066	0.065	0.0078
	0.9	0.070	0.0081	0.038	0.0060	0.052	0.0075
(40,50)	0.1	0.005	0.0023	0.016	0.0040	0.055	0.0072
	0.2	0.009	0.0030	0.019	0.0042	0.039	0.0061
	0.3	0.014	0.0037	0.010	0.0031	0.044	0.0065
	0.4	0.027	0.0051	0.013	0.0036	0.054	0.0071
	0.5	0.047	0.0067	0.022	0.0046	0.061	0.0076
	0.6	0.044	0.0065	0.049	0.0068	0.072	0.0082
	0.7	0.051	0.0070	0.050	0.0069	0.065	0.0078
	0.8	0.059	0.0075	0.042	0.0063	0.065	0.0078
	0.9	0.060	0.0075	0.043	0.0063	0.061	0.0076
(75,75)	0.1	0.001	0.0012	0.010	0.0029	0.012	0.0034
	0.2	0.002	0.0014	0.011	0.0033	0.013	0.0039
	0.3	0.015	0.0038	0.019	0.0042	0.016	0.0040
	0.4	0.016	0.0040	0.012	0.0034	0.018	0.0048
	0.5	0.022	0.0046	0.023	0.0047	0.024	0.0051
	0.6	0.024	0.0048	0.037	0.0060	0.046	0.0073
	0.7	0.025	0.0049	0.031	0.0056	0.031	0.0056
	0.8	0.028	0.0052	0.030	0.0054	0.035	0.0071
	0.9	0.036	0.0059	0.032	0.0056	0.031	0.0056
(100, 100)	0.1	0.002	0.0016	0.011	0.0030	0.038	0.0060
	0.2	0.006	0.0024	0.012	0.0034	0.040	0.0062
	0.3	0.014	0.0037	0.020	0.0044	0.048	0.0068
	0.4	0.024	0.0048	0.015	0.0038	0.046	0.0066
	0.5	0.033	0.0056	0.010	0.0031	0.044	0.0065
	0.6	0.036	0.0059	0.031	0.0055	0.046	0.0066
	0.7	0.038	0.0060	0.035	0.0058	0.046	0.0066
	0.8	0.047	0.0067	0.035	0.0058	0.046	0.0066
	0.9	0.045	0.0066	0.033	0.0057	0.045	0.0065

Table 4.2: Type I error and standard error comparison of ADF, JEL, and AJEL tests with nominal level $\alpha = 0.05$ when $X, Y \sim Exp(4)$

		ADF		JEL		AJEL	
(n_1, n_2)	t	TE	SE	TE	SE	TE	SE
(20,30)	0.1	0.037	0.0060	0.081	0.0086	0.091	0.0091
	0.2	0.022	0.0046	0.047	0.0067	0.060	0.0075
	0.3	0.010	0.0031	0.074	0.0083	0.075	0.0083
	0.4	0.006	0.0024	0.061	0.0076	0.074	0.0083
	0.5	0.006	0.0024	0.060	0.0075	0.061	0.0076
	0.6	0.001	0.0010	0.096	0.0093	0.101	0.0095
	0.7	0.001	0.0010	0.088	0.0090	0.087	0.0089
	0.8	0.001	0.0010	0.065	0.0078	0.075	0.0083
	0.9	0.000	0.0000	0.062	0.0076	0.064	0.0078
(40,50)	0.1	0.024	0.0049	0.060	0.0075	0.076	0.0084
	0.2	0.010	0.0031	0.043	0.0064	0.052	0.0070
	0.3	0.000	0.0000	0.048	0.0068	0.049	0.0068
	0.4	0.000	0.0000	0.045	0.0066	0.052	0.0070
	0.5	0.000	0.0000	0.057	0.0073	0.056	0.0073
	0.6	0.000	0.0000	0.068	0.0080	0.070	0.0081
	0.7	0.000	0.0000	0.065	0.0078	0.064	0.0077
	0.8	0.000	0.0000	0.067	0.0079	0.072	0.0082
	0.9	0.000	0.0000	0.059	0.0074	0.061	0.0075
(75,75)	0.1	0.018	0.0043	0.062	0.0076	0.071	0.0081
	0.2	0.005	0.0022	0.050	0.0069	0.055	0.0072
	0.3	0.000	0.0000	0.049	0.0058	0.050	0.0059
	0.4	0.000	0.0000	0.051	0.0070	0.056	0.0073
	0.5	0.000	0.0000	0.056	0.0074	0.060	0.0076
	0.6	0.000	0.0000	0.069	0.0080	0.069	0.0080
	0.7	0.000	0.0000	0.052	0.0072	0.052	0.0072
	0.8	0.000	0.0000	0.054	0.0071	0.056	0.0073
	0.9	0.000	0.0000	0.052	0.0072	0.055	0.0074
(100, 100)	0.1	0.012	0.0035	0.062	0.0076	0.071	0.0081
	0.2	0.006	0.0024	0.049	0.0068	0.055	0.0072
	0.3	0.001	0.0010	0.063	0.0077	0.061	0.0076
	0.4	0.001	0.0010	0.054	0.0071	0.059	0.0075
	0.5	0.000	0.0000	0.054	0.0071	0.051	0.0070
	0.6	0.000	0.0000	0.051	0.0070	0.051	0.0070
	0.7	0.000	0.0000	0.057	0.0073	0.057	0.0073
	0.8	0.000	0.0000	0.057	0.0073	0.057	0.0073
	0.9	0.000	0.0000	0.052	0.0072	0.052	0.0072

Table 4.3: Type I error and standard error comparison of ADF, JEL, and AJEL tests with nominal level $\alpha = 0.05$ when $X, Y \sim HN(1)$

		ADF		JEL		AJEL	
(n_1, n_2)	t	TE	SE	TE	SE	TE	SE
(20,30)	0.1	0.068	0.0080	0.074	0.0083	0.094	0.0092
	0.2	0.076	0.0084	0.063	0.0077	0.065	0.0078
	0.3	0.084	0.0088	0.050	0.0069	0.047	0.0067
	0.4	0.087	0.0089	0.051	0.0070	0.052	0.0070
	0.5	0.129	0.0106	0.050	0.0069	0.056	0.0073
	0.6	0.115	0.0101	0.077	0.0084	0.090	0.0090
	0.7	0.099	0.0094	0.063	0.0077	0.069	0.0080
	0.8	0.108	0.0098	0.062	0.0076	0.078	0.0085
	0.9	0.103	0.0096	0.041	0.0063	0.071	0.0081
(40,50)	0.1	0.062	0.0076	0.056	0.0073	0.066	0.0079
	0.2	0.066	0.0079	0.037	0.0060	0.042	0.0063
	0.3	0.068	0.0080	0.049	0.0068	0.054	0.0071
	0.4	0.081	0.0086	0.065	0.0078	0.058	0.0074
	0.5	0.101	0.0095	0.049	0.0068	0.059	0.0075
	0.6	0.094	0.0092	0.080	0.0086	0.087	0.0089
	0.7	0.096	0.0093	0.066	0.0079	0.064	0.0077
	0.8	0.108	0.0098	0.053	0.0071	0.070	0.0081
	0.9	0.098	0.0094	0.034	0.0057	0.055	0.0072
(75,75)	0.1	0.038	0.0061	0.051	0.0070	0.053	0.0071
	0.2	0.045	0.0066	0.055	0.0072	0.054	0.0071
	0.3	0.059	0.0075	0.051	0.0058	0.058	0.0063
	0.4	0.058	0.0074	0.053	0.0071	0.053	0.0071
	0.5	0.077	0.0084	0.044	0.0048	0.048	0.0052
	0.6	0.079	0.0085	0.057	0.0073	0.065	0.0078
	0.7	0.081	0.0086	0.045	0.0050	0.048	0.0052
	0.8	0.091	0.0091	0.055	0.0072	0.061	0.0076
	0.9	0.098	0.0094	0.036	0.0059	0.062	0.0076
(100, 100)	0.1	0.049	0.0069	0.046	0.0066	0.047	0.0067
	0.2	0.055	0.0072	0.071	0.0081	0.064	0.0077
	0.3	0.060	0.0075	0.054	0.0071	0.056	0.0073
	0.4	0.067	0.0079	0.060	0.0075	0.059	0.0075
	0.5	0.072	0.0082	0.048	0.0068	0.053	0.0071
	0.6	0.078	0.0085	0.065	0.0078	0.071	0.0081
	0.7	0.082	0.0087	0.066	0.0079	0.066	0.0079
	0.8	0.087	0.0089	0.063	0.0077	0.072	0.0082
	0.9	0.086	0.0089	0.039	0.0061	0.057	0.0073

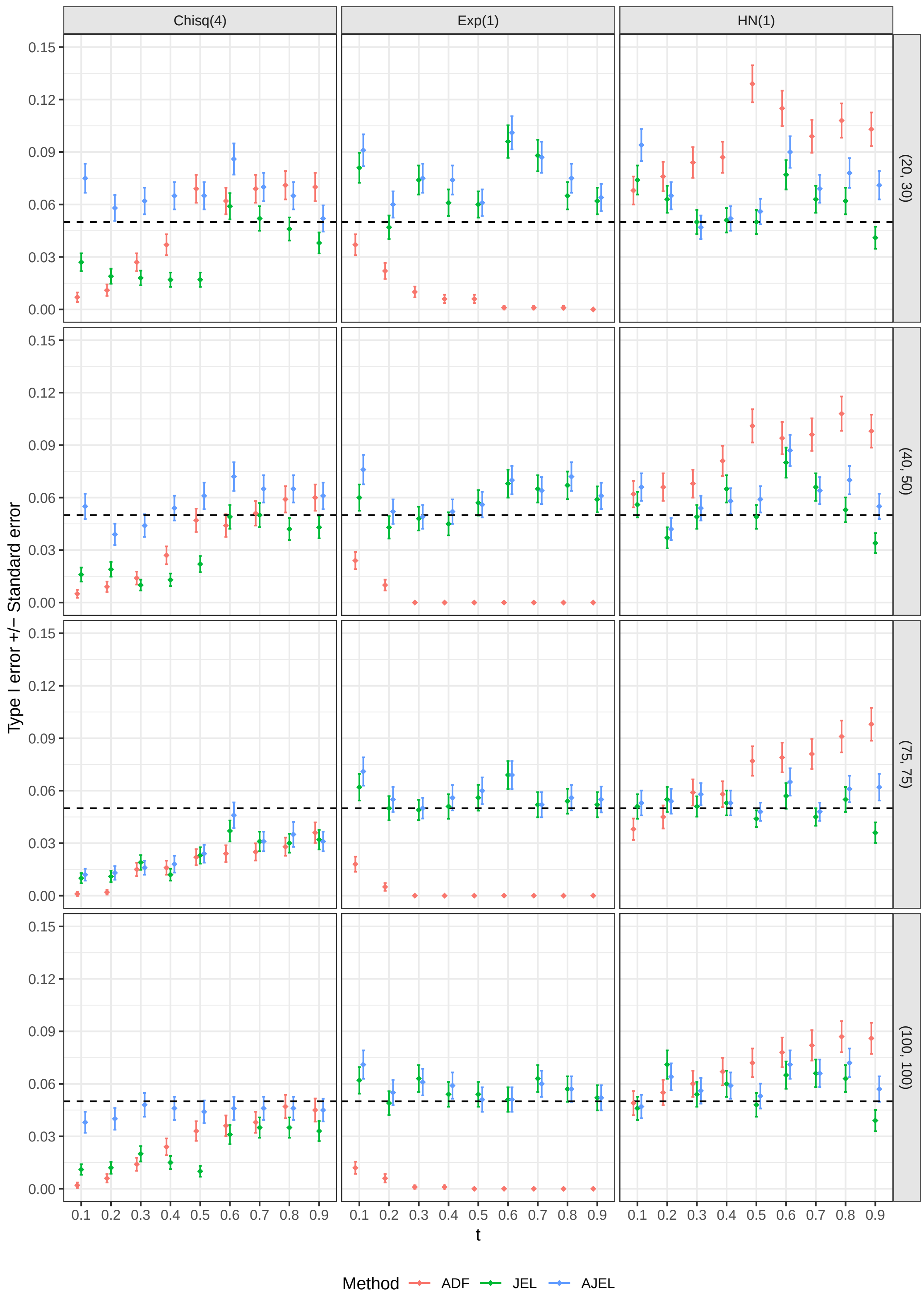

Figure 4.1: Type I error comparison for ADF, JEL, and AJEL methods for different distributions, sample sizes, and values of t

4.2 Power Analysis

Next, we conduct a power analysis for the ADF, JEL, and AJEL methods. Figure 4.2 displays the generalized Lorenz curves for Chi-Square, Exponential, and Half-Normal distribution under two sets of parameters. The difference between the generalized Lorenz curves of $\chi^2(4)$ and $\chi^2(5.5)$ increases as t changes from 0 to 0.5, then the difference decreases as t changes from 0.5 to 1. The difference between the generalized Lorenz curves for $Exp(2)$ and $Exp(4)$ increases significantly as t changes from 0 to 1, while the difference between the generalized Lorenz curves for $HN(1)$ and $HN(1.5)$ also increases but not as significant as in the case of the Exponential distributions.

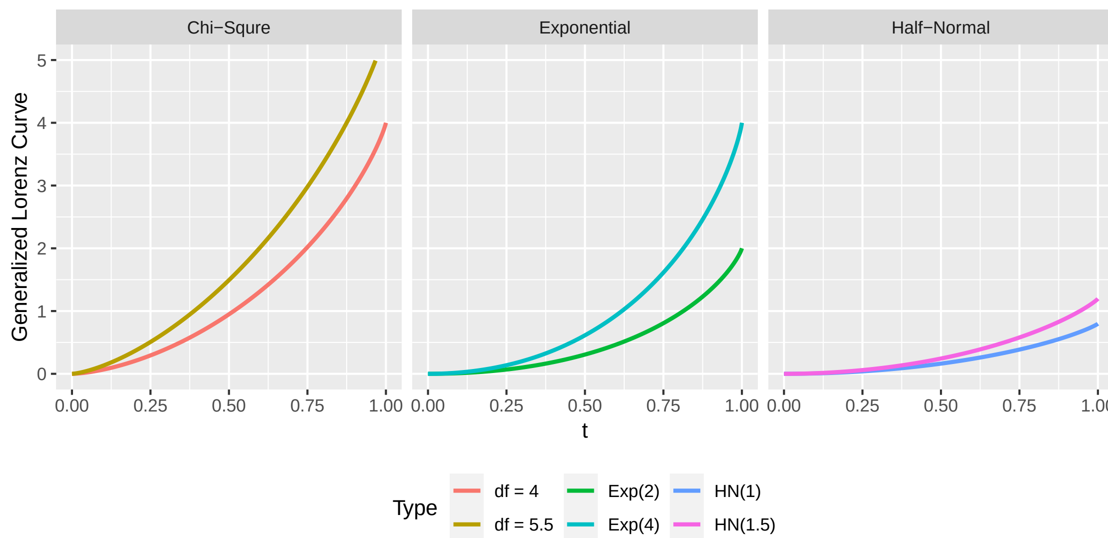

Figure 4.2: Generalized Lorenz curves for Chi-Square, Exponential, and Half-Normal distributions

Tables 4.4-4.6 present a summary of the outcomes, depicting the power and standard errors, whereas Figure 4.3 illustrates the results graphically. As expected, the power for Chi-Square distributions tends to increase as the value of t ranges from 0.1 to 0.5, followed by a slight drop as t ranges from 0.5 to 0.9. The AJEL method exhibits better power among the three methods, while the ADF method is the weakest. Moreover, for Exponential distributions, the power of the JEL and AJEL methods tends to increase as t ranges from 0 to 0.9, while the power of the ADF method tends to decrease as the value of t changes from 0 to 0.9. When considering the Half-Normal distributions, all three methods show a similar pattern, with the ADF method being the most effective when $t \leq 0.5$, and the JEL and AJEL methods being superior when $t > 0.5$. The increase in power is more pronounced for Exponential distributions than for Half-Normal distributions in the case of the JEL and AJEL methods. In all three cases, the AJEL method outperforms the JEL method slightly.

Table 4.4: Power comparison of ADF, JEL, and AJEL tests with nominal level $\alpha = 0.05$ when $X \sim \chi^2_4$ and $Y \sim \chi^2_{5.5}$

		ADF		JEL		AJEL	
(n_1, n_2)	t	Power	SE	Power	SE	Power	SE
(20,30)	0.1	0.128	0.0106	0.357	0.0152	0.372	0.0153
	0.2	0.190	0.0124	0.400	0.0155	0.415	0.0156
	0.3	0.330	0.0149	0.423	0.0156	0.441	0.0157
	0.4	0.360	0.0152	0.454	0.0157	0.471	0.0158
	0.5	0.454	0.0157	0.443	0.0157	0.461	0.0158
	0.6	0.475	0.0158	0.834	0.0118	0.840	0.0116
	0.7	0.439	0.0157	0.771	0.0133	0.781	0.0131
	0.8	0.450	0.0157	0.717	0.0142	0.725	0.0141
	0.9	0.488	0.0158	0.594	0.0155	0.610	0.0154
(40,50)	0.1	0.182	0.0122	0.483	0.0158	0.497	0.0158
	0.2	0.369	0.0153	0.595	0.0155	0.607	0.0154
	0.3	0.508	0.0158	0.660	0.0150	0.673	0.0148
	0.4	0.578	0.0156	0.664	0.0149	0.676	0.0148
	0.5	0.657	0.0150	0.644	0.0151	0.654	0.0150
	0.6	0.646	0.0151	0.935	0.0078	0.936	0.0077
	0.7	0.655	0.0150	0.914	0.0089	0.915	0.0088
	0.8	0.670	0.0149	0.878	0.0103	0.883	0.0102
	0.9	0.694	0.0146	0.783	0.0130	0.793	0.0128
(75,75)	0.1	0.438	0.0157	0.688	0.0147	0.695	0.0146
	0.2	0.595	0.0155	0.811	0.0124	0.820	0.0036
	0.3	0.757	0.0136	0.986	0.0037	0.987	0.0036
	0.4	0.790	0.0129	0.869	0.0107	0.872	0.0106
	0.5	0.841	0.0116	0.983	0.0041	0.983	0.0041
	0.6	0.852	0.0112	0.979	0.0045	0.980	0.0044
	0.7	0.856	0.0111	0.791	0.0129	0.797	0.0127
	0.8	0.872	0.0106	0.954	0.0066	0.957	0.0064
	0.9	0.875	0.0105	0.927	0.0082	0.928	0.0082
(100,100)	0.1	0.625	0.0153	0.809	0.0124	0.813	0.0123
	0.2	0.786	0.0130	0.899	0.0095	0.901	0.0094
	0.3	0.871	0.0106	0.931	0.0080	0.933	0.0079
	0.4	0.923	0.0084	0.932	0.0080	0.933	0.0079
	0.5	0.932	0.0080	0.936	0.0077	0.94	0.0075
	0.6	0.933	0.0079	0.993	0.0026	0.993	0.0026
	0.7	0.942	0.0074	0.991	0.0030	0.991	0.0030
	0.8	0.951	0.0068	0.984	0.0040	0.986	0.0037
	0.9	0.953	0.0067	0.972	0.0052	0.973	0.0051

Table 4.5: Power comparison of ADF, JEL, and AJEL tests with nominal level $\alpha = 0.05$ when $X \sim Exp(4)$ and $Y \sim Exp(2)$

		ADF		JEL		AJEL	
(n_1, n_2)	t	Power	SE	Power	SE	Power	SE
(20, 30)	0.1	0.695	0.0146	0.363	0.0152	0.382	0.0154
	0.2	0.678	0.0148	0.463	0.0158	0.478	0.0158
	0.3	0.520	0.0158	0.551	0.0157	0.566	0.0157
	0.4	0.417	0.0156	0.613	0.0154	0.626	0.0153
	0.5	0.418	0.0156	0.664	0.0149	0.682	0.0147
	0.6	0.367	0.0152	0.893	0.0098	0.901	0.0094
	0.7	0.295	0.0144	0.884	0.0101	0.894	0.0097
	0.8	0.284	0.0143	0.868	0.0107	0.875	0.0105
	0.9	0.238	0.0135	0.834	0.0118	0.846	0.0114
(40, 50)	0.1	0.835	0.0117	0.471	0.0158	0.485	0.0158
	0.2	0.812	0.0124	0.606	0.0155	0.615	0.0154
	0.3	0.582	0.0156	0.724	0.0141	0.730	0.0140
	0.4	0.492	0.0158	0.792	0.0128	0.803	0.0126
	0.5	0.470	0.0158	0.837	0.0117	0.851	0.0113
	0.6	0.368	0.0153	0.969	0.0055	0.971	0.0053
	0.7	0.324	0.0148	0.971	0.0053	0.972	0.0052
	0.8	0.317	0.0147	0.961	0.0061	0.966	0.0057
	0.9	0.290	0.0143	0.958	0.0063	0.960	0.0062
(75, 75)	0.1	0.901	0.0094	0.616	0.0154	0.623	0.0153
	0.2	0.886	0.0101	0.787	0.0129	0.791	0.0129
	0.3	0.719	0.0142	0.975	0.0049	0.976	0.0048
	0.4	0.586	0.0156	0.930	0.0081	0.930	0.0081
	0.5	0.546	0.0157	0.993	0.0026	0.993	0.0026
	0.6	0.440	0.0157	0.995	0.0022	0.995	0.0022
	0.7	0.408	0.0155	0.987	0.0036	0.988	0.0034
	0.8	0.371	0.0153	0.995	0.0022	0.995	0.0022
	0.9	0.348	0.0151	0.994	0.0024	0.995	0.0022
(100, 100)	0.1	0.946	0.0071	0.661	0.0150	0.667	0.0149
	0.2	0.930	0.0081	0.851	0.0113	0.856	0.0111
	0.3	0.717	0.0142	0.927	0.0082	0.928	0.0082
	0.4	0.629	0.0153	0.965	0.0058	0.968	0.0056
	0.5	0.551	0.0157	0.984	0.0040	0.985	0.0038
	0.6	0.483	0.0158	1.000	0.0000	1.000	0.0000
	0.7	0.429	0.0157	1.000	0.0000	1.000	0.0010
	0.8	0.419	0.0156	0.999	0.0010	0.999	0.0010
	0.9	0.355	0.0151	0.999	0.0010	0.999	0.0010

Table 4.6: Power comparison of ADF, JEL, and AJEL tests with nominal level $\alpha = 0.05$ when $X \sim HN(1)$ and $Y \sim HN(1.5)$

		ADF		JEL		AJEL	
(n_1, n_2)	t	Power	SE	Power	SE	Power	SE
(20, 30)	0.1	0.415	0.0156	0.260	0.0139	0.274	0.0141
	0.2	0.438	0.0157	0.292	0.0144	0.303	0.0145
	0.3	0.479	0.0158	0.309	0.0146	0.327	0.0148
	0.4	0.485	0.0158	0.373	0.0153	0.390	0.0154
	0.5	0.541	0.0158	0.377	0.0153	0.395	0.0155
	0.6	0.559	0.0157	0.721	0.0142	0.737	0.0139
	0.7	0.520	0.0158	0.694	0.0146	0.716	0.0143
	0.8	0.537	0.0158	0.670	0.0149	0.688	0.0147
	0.9	0.548	0.0157	0.597	0.0155	0.616	0.0154
(40, 50)	0.1	0.654	0.0150	0.312	0.0147	0.330	0.0149
	0.2	0.661	0.0150	0.393	0.0154	0.400	0.0155
	0.3	0.676	0.0148	0.455	0.0157	0.468	0.0158
	0.4	0.690	0.0146	0.544	0.0158	0.561	0.0157
	0.5	0.725	0.0141	0.577	0.0156	0.589	0.0156
	0.6	0.713	0.0143	0.838	0.0117	0.845	0.0114
	0.7	0.723	0.0142	0.840	0.0116	0.847	0.0114
	0.8	0.733	0.0140	0.841	0.0116	0.847	0.0114
	0.9	0.746	0.0138	0.809	0.0124	0.816	0.0123
(75, 75)	0.1	0.839	0.0116	0.381	0.0154	0.390	0.0154
	0.2	0.843	0.0115	0.508	0.0158	0.515	0.0158
	0.3	0.853	0.0112	0.896	0.0097	0.897	0.0096
	0.4	0.866	0.0108	0.689	0.0146	0.700	0.0145
	0.5	0.877	0.0104	0.912	0.0090	0.913	0.0089
	0.6	0.876	0.0104	0.933	0.0079	0.936	0.0077
	0.7	0.882	0.0102	0.846	0.0114	0.851	0.0113
	0.8	0.891	0.0099	0.932	0.0080	0.934	0.0079
	0.9	0.892	0.0098	0.929	0.0081	0.931	0.0080
(100, 100)	0.1	0.938	0.0076	0.390	0.0155	0.400	0.0155
	0.2	0.940	0.0075	0.554	0.0157	0.561	0.0157
	0.3	0.941	0.0075	0.691	0.0146	0.695	0.0146
	0.4	0.944	0.0073	0.787	0.0129	0.792	0.0128
	0.5	0.951	0.0068	0.837	0.0117	0.839	0.0116
	0.6	0.953	0.0067	0.958	0.0063	0.958	0.0063
	0.7	0.952	0.0068	0.969	0.0055	0.970	0.0054
	0.8	0.957	0.0064	0.973	0.0051	0.973	0.0051
	0.9	0.958	0.0063	0.972	0.0052	0.975	0.0049

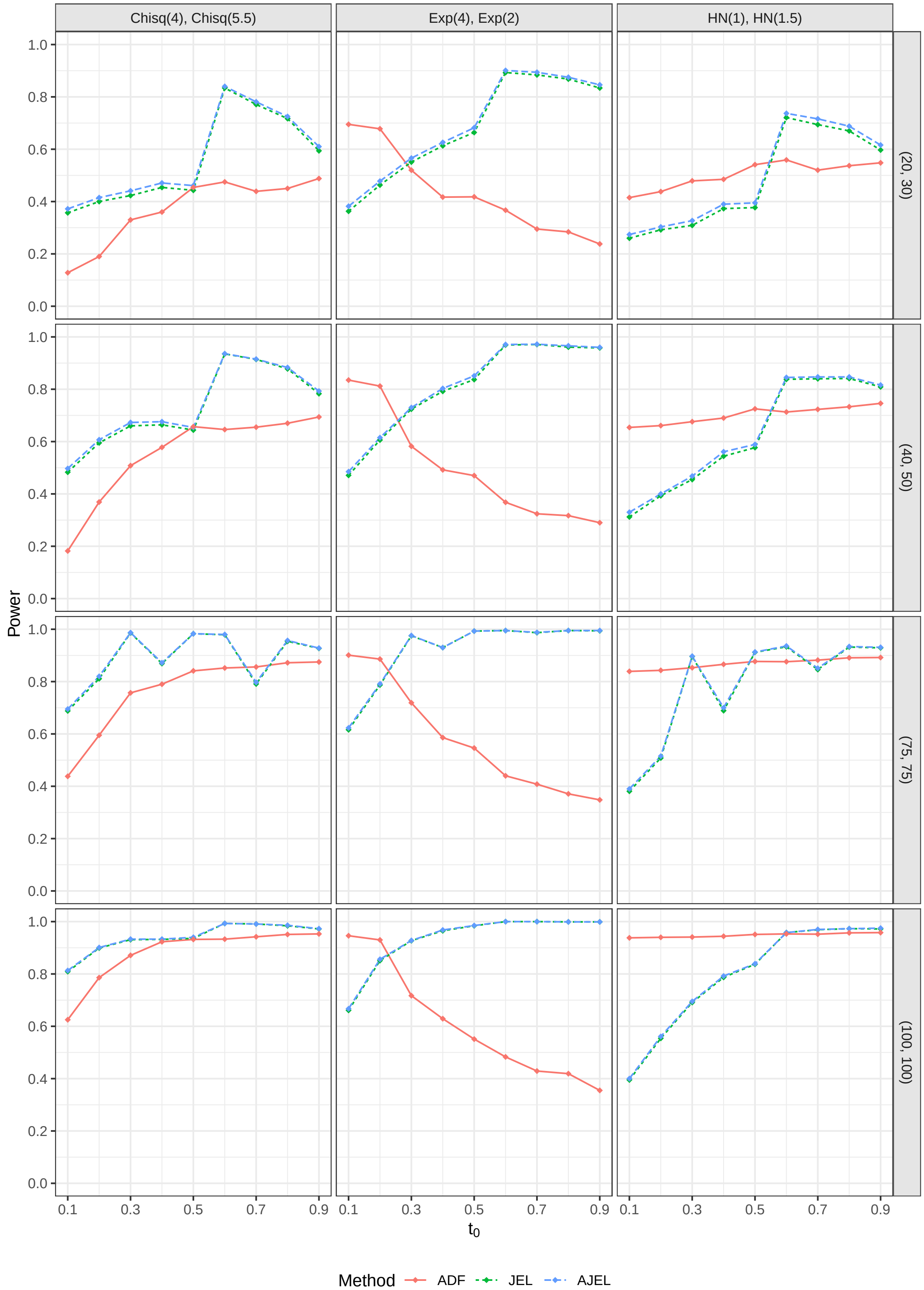

Figure 4.3: Power comparison for ADF, JEL, and AJEL methods for different distributions, sample sizes, and values of t

Chapter 5

Real Data Applications

In this chapter, we use the proposed methods to evaluate the equality of the generalized Lorenz curves for various subgroups of employees of California State University (CSU) and University of California (UC) systems in 2021.[1] The 2021 data comprises of 105,414 records of salaries for CSU and 299,448 records of salaries for UC. The data is anonymous but grouped by employer name and types of positions.[2] For the purpose of this analysis, we considered testing the hypothesis defined in equation (3.5) at a significance level of 5% for the following four scenarios.[3] In each scenario, we apply the proposed testing procedures[4] for $t = 0.0$, 0.2, 0.4, 0.6, 0.8, 1.0, and obtain the corresponding test statistics and p-values.

[1]The most recent data was obtained and is available on the State Controller's Office website at https://publicpay.ca.gov/Reports/Explore.aspx.

[2]To identify CSU instructional faculty, we filtered the data set based on the following key words: "Instructional Faculty", "Teaching Associate", "Visiting Faculty", "Lecturer", "Academic-Related", "Department Chair". To identify UC instructional faculty, we filtered the data set based on the following key words: "Assoc Prof", "Assoc Adj", "Asst Adj", "Asst Prof", "Prof In", "VIS Prof", "Adj Instr", "Lect", "Grad", "Adj". All other employees that didn't possess the listed key words in description of their position were identified as non-teaching staff.

[3]Log-transformed data was used to avoid unnecessary computational burden.

[4]Note that for ADF procedure, $t = 0.0$ can not be used, thus we used $t = 0.2$, 0.4, 0.6, 0.8, 1.0

5.1 Using Complete Data to Compare Income Distributions of Faculty at CSU Monterey Bay and CSU San Bernardino

In this scenario, we examine the salaries of all instructional faculty from CSU Monterey Bay and CSU San Bernardino. Filtering data based on the employer name and the type of position resulted in obtaining the total of 546 records for CSU Monterey Bay faculty salaries and 1,265 records for CSU San Bernardino faculty. These two institutions were chosen because their Lorenz and generalized Lorenz curves appear to be very similar as shown in Figure 5.1.

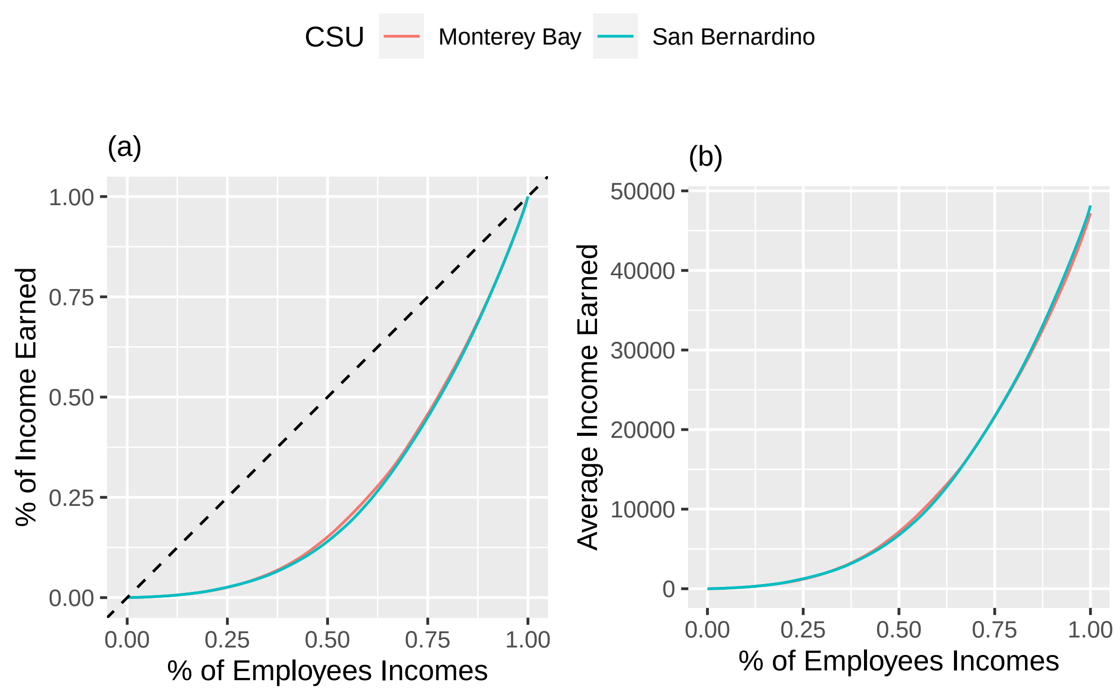

Figure 5.1: (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSU Monterey Bay and CSU San Bernardino faculty

Table 5.1: Test statistic and p-value

		t					
Method	Value	0.0	0.2	0.4	0.6	0.8	1.0
ADF	Test statistic	-	0.0119	0.0119	0.0119	0.0128	0.0135
	p-value	-	0.9133	0.9133	0.9133	0.9100	0.9076
JEL	Test statistic	0.2116	667.8884	417.0143	0.0380	0.0070	0.0138
	p-value	0.6455	0.0000	0.0000	0.8454	0.9332	0.9065
AJEL	Test statistic	0.2126	669.2835	417.9359	0.0382	0.0071	0.0139
	p-value	0.6448	0.0000	0.0000	0.8450	0.9331	0.9063

Table 5.1 shows the calculated test statistics and p-values for JEL and AJEL methods at $t = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ and for ADF method at $t = 0.2, 0.4, 0.6, 0.8, 1.0$. While ADF method suggest lack of evidence to conclude that the two generalized Lorenz curves are different for all values of t, the other two methods lead to different conclusions depending on values of t. On the one hand, both methods lead to the rejection of the null hypothesis $\theta = 0$ for $t = 0.2$ and $t = 0.4$ which means that at 5% significance level we have sufficient evidence to conclude that the considered generalized Lorenz curves are significantly different for 20th and 40th percentiles. On the other hand, both methods produce p-values that are higher than 0.05 for $t = 0.0$, $t = 0.6$, $t = 0.8$, and $t = 1.0$ which means that we do not have sufficient evidence to conclude that the considered generalized Lorenz curves are significantly different for 0th, 60th, 80th, and 100th percentiles. Such mixed results can be explained by the fact that the considered generalized Lorentz curves intersect multiple times.

5.2 Using Complete Data to Compare Income Distributions of Faculty at CSU San Bernardino and CSU San Francisco

In this scenario, we examine the salaries of all instructional faculty from CSU San Bernardino and CSU San Francisco. Filtering data based on the employer name and the type of position resulted in obtaining the total of 1,265 records for CSU San Bernardino faculty salaries and 2,294 records for CSU San Francisco faculty. These two institutions were chosen because their Lorenz and generalized Lorenz curves appear to be very distinct as shown in Figure 5.2.

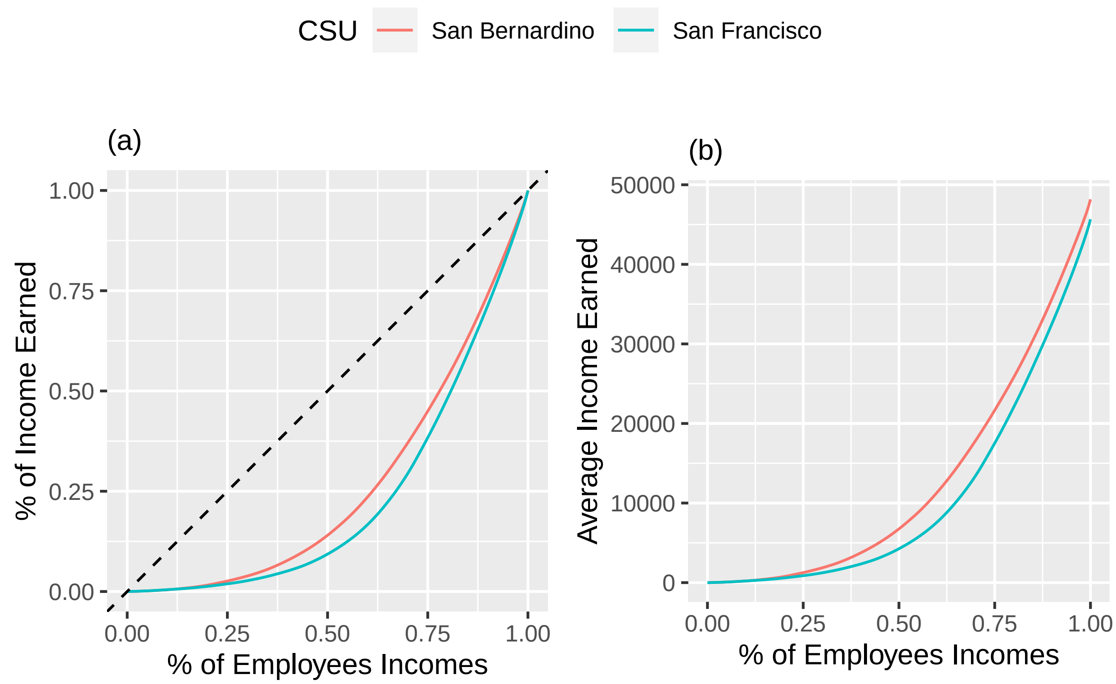

Figure 5.2: (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSU San Bernardino and CSU San Francisco faculty

Table 5.2: Test statistic and p-value

		t					
Method	Value	0.0	0.2	0.4	0.6	0.8	1.0
ADF	Test statistic	-	17.3485	20.1831	23.0600	24.9908	28.8444
	p-value	-	0.0000	0.0000	0.0000	0.0000	0.0000
JEL	Test statistic	1.5821	217.8804	942.6197	250.0000	30.2795	20.0987
	p-value	0.2085	0.0000	0.0000	0.0000	0.0000	0.0000
AJEL	Test statistic	1.5860	218.2452	943.7490	250.0000	30.3477	20.1438
	p-value	0.2079	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.2 shows the calculated test statistics and p-values for JEL and AJEL methods at $t = 0.0$, 0.2, 0.4, 0.6, 0.8, 1.0 and for ADF method at $t = 0.2$, 0.4, 0.6, 0.8, 1.0. On the one hand, all three methods lead to the rejection of the null hypothesis $\theta = 0$ for $t = 0.2$, $t = 0.4$, $t = 0.6$, $t = 0.8$, and $t = 1.0$ which means that at 5% significance level we have sufficient evidence to conclude that the considered generalized Lorenz curves are significantly different for 20th, 40th, 60th, 80th, adn 100th percentiles. On the other

hand, while ADF is inapplicable at $t = 0$, the other two methods at $t = 0$ produce p-values that are higher than 0.05, which means that we do not have sufficient evidence to conclude that the considered generalized Lorenz curves are significantly different for 0th percentiles. These results can be explained by the fact that the minimum salaries are nearly equal for both institutions but as t increases the difference increases as well.

5.3 Using Incomplete Data to Compare Income Distributions of All Faculty at CSU and UC

In this scenario, we examine the instructional faculty salaries across all CSU and UC institutions. Filtering data based on the type of position resulted in obtaining the total of 34,927 records for CSU faculty salaries and 16,798 records for UC faculty salaries. The Lorenz and generalized Lorenz curves are shown in Figure 5.3. Since both data sets

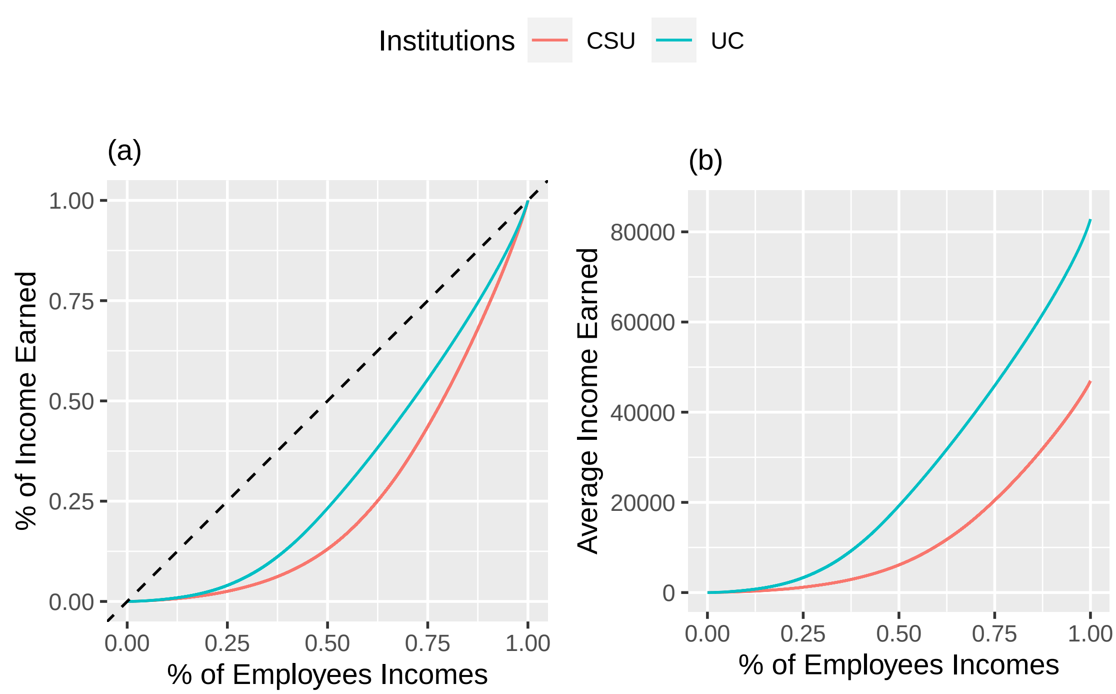

Figure 5.3: (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSU and UC faculty

are quite large, we decided to utilize the proposed procedures using the samples. Given that empirical likelihood approach is effective for relatively small samples, we chose the sample sizes $n_1 = n_2 = 100$. Since the difference between the generalized Lorenz curves is somewhat distinct, we expect both procedures to be able to capture the difference even with such small sample sizes.

Table 5.3: Test statistic and p-value

		t					
Method	Value	0.0	0.2	0.4	0.6	0.8	1.0
ADF	Test statistic	-	14.9673	17.3378	18.9842	19.8941	24.9073
	p-value	-	0.0001	0.0000	0.0000	0.0000	0.0000
JEL	Test statistic	0.6586	48.6775	20.5069	66.4861	38.4036	25.2639
	p-value	0.4171	0.0000	0.0000	0.0000	0.0000	0.0000
AJEL	Test statistic	0.6833	49.7646	20.9485	67.5089	39.0850	25.7541
	p-value	0.4084	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.3 shows the calculated test statistics and p-values for JEL and AJEL methods at $t = 0.0$, 0.2, 0.4, 0.6, 0.8, 1.0 and for ADF method at $t = 0.2$, 0.4, 0.6, 0.8, 1.0. On the one hand, all three methods lead to the rejection of the null hypothesis $\theta = 0$ for $t = 0.2$, $t = 0.4$, $t = 0.6$, $t = 0.8$, and $t = 1.0$ which means that at 5% significance level we have sufficient evidence to conclude that the considered generalized Lorenz curves are significantly different for 20th, 40th, 60th, 80th, and 100th percentiles. On the other hand, both methods produce p-values that are higher than 0.05 for $t = 0.0$ which means that we do not have sufficient evidence to conclude that the considered generalized Lorenz curves are significantly different for 0th percentiles. Similar to the previous scenario, these results can be explained by the fact that the minimum salaries are nearly equal for both institutions but as t increases the difference increases as well. However, what is worth noting about these results is that we were able to capture these differences using relatively small samples.

5.4 Using Incomplete Data to Compare Income Distributions of CSUSB Faculty between 2009 and 2020

In this scenario, we examine the salaries of all instructional faculty at CSUSB in the years 2009 and 2020. Filtering data based on the employer name and the type of position resulted in obtaining a total of 1,111 records for 2009 and 1,331 records for 2020.

The Lorenz and generalized Lorenz curves are graphed in Figure 5.4.

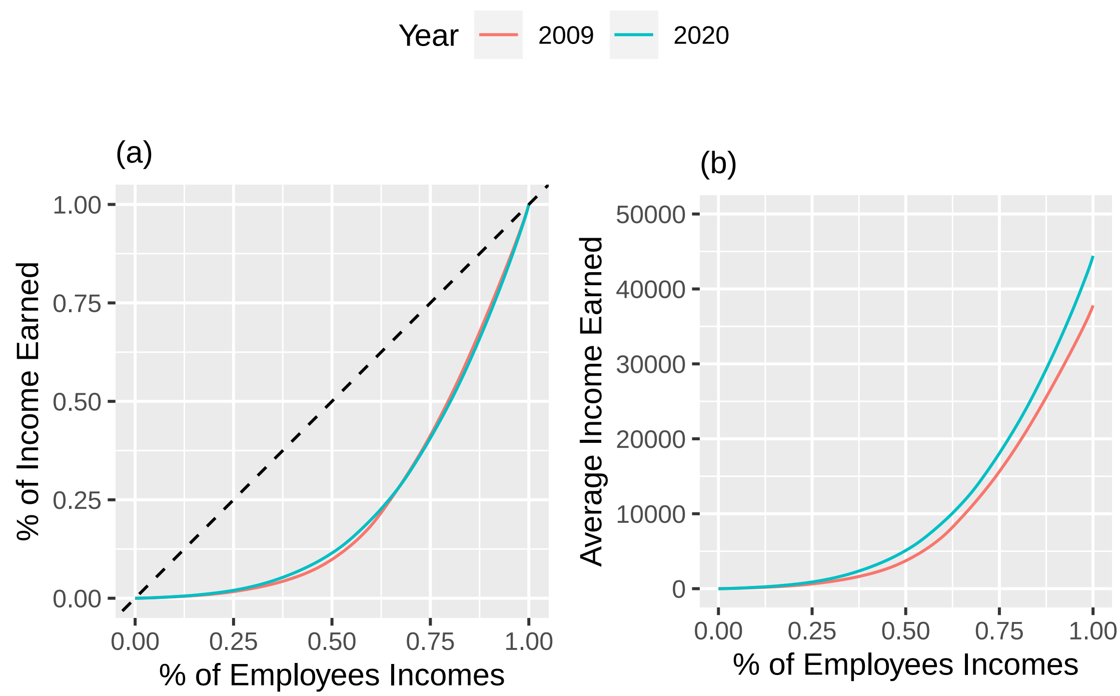

Figure 5.4: (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSUSB faculty in years 2009 and 2020

After adjusting for inflation and expressing the wages in 2020 US dollar[5], the adjusted Lorenz and generalized Lorenz curves are graphed in Figure 5.5.

[5]The adjustment coefficient 1.23 was obtained using the CPI inflation calculator from https://www.bls.gov/data/inflation_calculator.htm between January 2009 and January 2020.

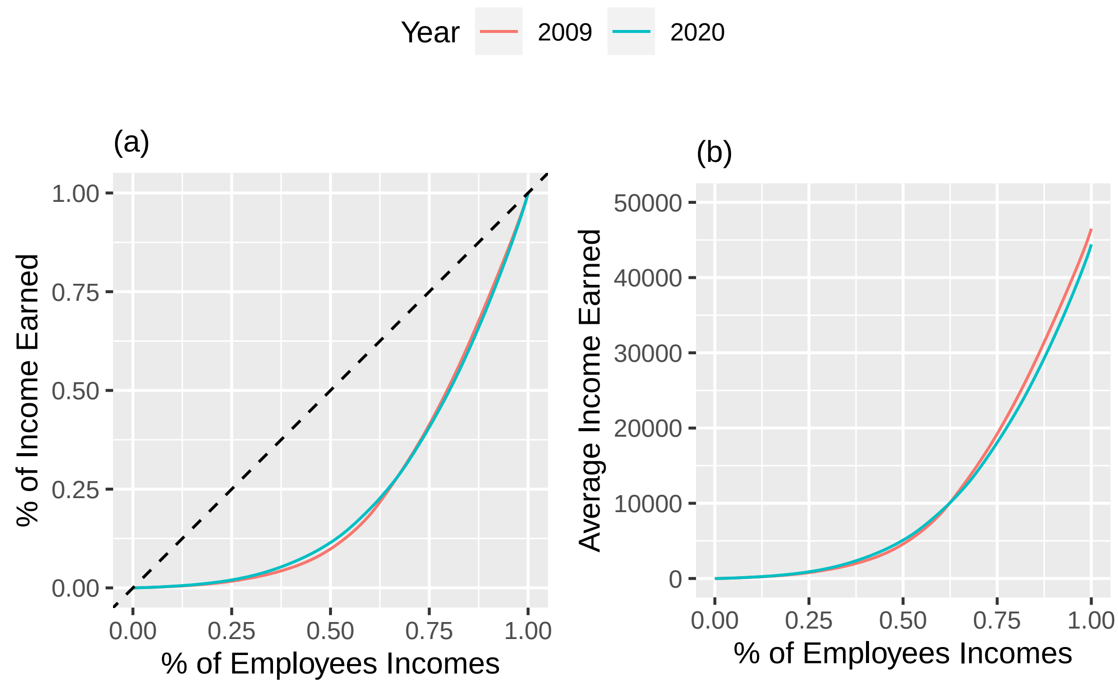

Figure 5.5: (a) Lorenz curves and (b) generalized Lorenz curves for salaries of CSUSB faculty in years 2009 and 2020

Once again, since both data sets are quite large, we utilized the proposed procedures using the samples of sizes $n_1 = n_2 = 100$.

Table 5.4: Test statistic and p-value

				t			
Method	Value	0.0	0.2	0.4	0.6	0.8	1.0
ADF	Test statistic	-	1.5990	2.2190	2.2614	2.5820	2.5776
	p-value	-	0.2060	0.1363	0.1326	0.1081	0.1084
JEL	Test statistic	0.1719	52.4042	29.2928	0.1161	0.0624	0.1486
	p-value	0.6785	0.0000	0.0000	0.7333	0.8028	0.6999
AJEL	Test statistic	0.1725	52.5651	29.3841	0.1165	0.0626	0.1490
	p-value	0.6779	0.0000	0.0000	0.7329	0.8025	0.6994

Table 5.4 shows the calculated test statistics and p-values for JEL and AJEL methods at $t = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ and for ADF method at $t = 0.2, 0.4, 0.6, 0.8, 1.0$. On the one hand, the ADF method lead to failure of rejection of the null hypothesis $\theta = 0$ for all values of t, which means that at 5% significance level, we do

not have sufficient evidence to conclude that the considered generalized Lorenz curves are significantly different for 20th, 40th, 60th, 80th, and 100th percentiles. On the other hand, the JEL and AJEL methods result in large p-values for $t = 0.0, 0.6, 0.8, 1.0$ and small p-values at $t = 0.2$, $t = 0.4$, which means that we have sufficient evidence to conclude that the generalized Lorenz curves in this scenario are significantly diffirent at $t = 0.2$ and $t = 0.4$. Such mixed results can be explained by the fact that the considered generalized Lorentz curves intersect multiple times. The two proposed methods were able to capture these differences using relatively small samples.

Chapter 6

Discussion and Conclusion

In this thesis, we introduced two non-parametric JEL-based methods using a U-statistic to test the equality of two generalized Lorenz curves. The limiting distribution of the likelihood ratios are shown to follow a chi-squared distribution with one degree of freedom. Simulations for different distribution types and various sample sizes illustrate that as the sample size increases both methods improve in terms of controlling the Type I error probability and the power. Simulations showed that AJEL resulted in higher test powers in comparison to JEL across all distributions, sample sizes, and values of t with a few exceptions. However, AJEL also resulted in higher Type I errors in comparison to JEL, but still within an acceptable range. Applications to real data sets for four different scenarios demonstrated that both testing procedures produce reasonable results for complete and incomplete data.

In future research, we can extend the results by utilizing other modified approaches such as transformed empirical likelihood and transformed adjusted empirical likelihood. Further, we are also interested in utilizing a kernel-smoothing estimator to extend the proposed methods via a smoothed jackknife empirical likelihood approach.

Appendix

Proofs of Theorems

Proof. **Theorem 3.1**

Let $n_1 \leq n_2$. As shown in [Arv69], the jackknife procedure for the two sample U-statistics, U_{n_1,n_2}, we have

$$\begin{aligned} V_{i,0} &= n_1 U_{n_1,n_2} - (n_1 - 1)U^{-i,0}_{n_1-1,n_2}, \quad i = 1, \cdots, n_1 \\ V_{0,j} &= n_2 U_{n_1,n_2} - (n_2 - 1)U^{0,-j}_{n_1-1,n_2}, \quad j = 1, \cdots, n_2 \end{aligned} \tag{6.1}$$

Further, they proposed a consistent estimator of $Var(U_{n_1,n_2})$ given as,

$$\widehat{Var}_{\text{Jack}}(U_{n_1,n_2}) = \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} \Big(V_{i,0} - \bar{V}_{\cdot,0} \Big)^2 + \frac{1}{n_2(n_2-1)} \sum_{j=1}^{n_2} \Big(V_{0,j} - \bar{V}_{0,\cdot} \Big)^2$$

where $\bar{V}_{\cdot,0}$ and $\bar{V}_{0,\cdot}$ are the means of $V_{i,0}$ and $V_{0,j}$ respectively.

Lemma 6.0.1. *(See [Arv69])*

1. *Assume that* $E|h(X,Y)| < \infty$, *then* $U_{n_1,n_2} \xrightarrow{a.s.} \theta$ *as* $n \longrightarrow \infty$.
2. *Assume that* $Eh^2(X,Y) < \infty, \sigma^2_{1,0} > 0$ *and* $\sigma^2_{0,1} > 0$, *let* $S^2_{n_1,n_2} = \frac{1}{n_1}\sigma^2_{1,0} + \frac{1}{n_2}\sigma^2_{0,1}$, *then*

$$\frac{U_{n_1,n_2} - \theta}{S_{n_1,n_2}} \xrightarrow{d} N(0,1) \quad \textit{and} \quad \widehat{Var}_{Jack}(U_{n_1,n_2}) - S^2_{n_1,n_2} = o_p(n_1^{-1}) \quad \textit{as} \quad n_1 \longrightarrow \infty.$$

In order to apply JEL, using (6.1), we can determine

$$V_{i,0} = \frac{1}{n_2} I(X_i \leq \psi_t) \sum_{r=1}^{n_2} \big(X_i - Y_r\big)\, I(Y_r \leq \psi_t), \quad i = 1, \cdots, n_1$$

and

$$V_{0,j} = \frac{1}{n_1} I(Y_j \le \psi_t) \sum_{s=1}^{n_1} \big(X_s - Y_j\big) I(X_s \le \psi_t), \quad j = 1, \cdots, n_2$$

Let $n = n_1 + n_2$. Consider

$$U_n = \frac{1}{n_1 n_2} \sum_{1 \le i \le n_1 < j \le n} \big(X_i - Y_{j-n_1}\big) I(X_i \le t) I(Y_{j-n_1} \le \psi_t)$$

and

$$\begin{aligned} U_n^{-i} &= U\big(Z_1, Z_2, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n\big) \\ &= \binom{n-1}{2}^{-1} \frac{1}{n_1 n_2} \sum_{\substack{1 \le r < s \le n \\ r,s \ne i}} \big(X_r - Y_{s-n_1}\big) I(X_r \le t) I(Y_{s-n_1} \le \psi_t) \\ &= \begin{cases} \frac{n}{n-2}\Big[U_n - \frac{1}{n_1 n_2} \sum_{i<j} \big(X_i - Y_{j-n_1}\big) I(X_i \le t) I(Y_{j-n_1} \le \psi_t)\Big] & , \ 1 \le i \le n_1 \\ \\ \frac{n}{n-2}\Big[U_n - \frac{1}{n_1 n_2} \sum_{j<i} \big(X_j - Y_{i-n_1}\big) I(X_j \le t) I(Y_{i-n_1} \le \psi_t)\Big] & , \ n_1 < i \le n \end{cases} \end{aligned}$$

It can be seen that

$$\frac{1}{n_1 n_2} \sum_{i<j} \big(X_i - Y_{j-n_1}\big) I(X_i \le t) I(Y_{j-n_1} \le \psi_t) = \frac{1}{n_1} V_{i,0} \ , \quad 1 \le i \le n_1$$

and

$$\frac{1}{n_1 n_2} \sum_{j<i} \big(X_j - Y_{i-n_1}\big) I(X_j \le t) I(Y_{i-n_1} \le \psi_t) = \frac{1}{n_2} V_{0,i} \ , \quad n_1 < i \le n$$

Now, consider JEL given in (3.8), for $1 \le k \le n$, we have

$$\begin{aligned} \widehat{V}_k &= nU_n - (n-1)U_{n-1}^{-k} \\ &= \frac{n(n-1)}{n-2}\left[\left(\frac{V_{k,0}}{n_1}\right) I_{(1 \le k \le n_1)} + \left(\frac{V_{0,k-n_1}}{n_2}\right) I_{(n_1+1 \le k \le n)}\right] - \frac{n}{n-2} U_{n_1,n_2} \end{aligned}$$

Thus,

$$E\widehat{V}_k = \frac{n\theta}{n-2}\left[\left(\frac{n_2 - 1}{n_1}\right) I_{(1 \le k \le n_1)} + \left(\frac{n_1 - 1}{n_2}\right) I_{(n_1+1 \le k \le n)}\right]$$

Under H_0, $E\widehat{V}_k = 0$. Next, following the similar arguments given in [JYZ09], for fixed $t = t_0 \in [0,1]$, it can be proven that $\ell(\theta(t_0)) \longrightarrow \chi_1^2$, as $n_1 \longrightarrow \infty$. Thus, details are omitted here.

□

Proof. **Theorem 3.2**

The proof of this theorem is similar to Theorem 1 given in [CVA08]. Let $\lambda^{\text{Adj}}(t)$ be the solution to

$$\sum_{k=1}^{n+1} \frac{g_k^{\text{Adj}}(t)}{1+\lambda^{\text{Adj}}(t)g_k^{\text{Adj}}(t)} = 0. \tag{6.2}$$

The first step is to show that $\lambda^{\text{Adj}}(t) = O_p(n^{-1/2})$. By using Lemma 3 of [Owe90] and the fact that $E(\hat{V}_1^2(t)) < \infty$, we can establish that $g^* = \max_{1\le k\le n} ||\hat{V}_k|| = o_p(n^{1/2})$ and $\bar{g}_n(t) = O_p(n^{-1/2})$. Let $\rho = ||\lambda^{\text{Adj}}(t)||$, $a_n = o_p(n)$ and $\hat{\lambda}^{\text{Adj}}(t) = \lambda^{\text{Adj}}(t)/\rho$. Multiplying $\hat{\lambda}^{\text{Adj}}(t)/n$ to both sides gives

$$\begin{aligned} 0 &= \frac{\hat{\lambda}^{\text{Adj}}(t)}{n} \sum_{k=1}^{n+1} \frac{g_k^{\text{Adj}}(t)}{1+\lambda^{\text{Adj}}(t)g_k^{\text{Adj}}(t)} \\ &= \frac{\hat{\lambda}^{\text{Adj}}(t)}{n} \sum_{k=1}^{n+1} g_k^{\text{Adj}}(t) - \frac{\rho}{n}\sum_{k=1}^{n+1} \frac{(\hat{\lambda}^{\text{Adj}}(t)g_k^{\text{Adj}}(t))^2}{1+\rho\hat{\lambda}^{\text{Adj}}(t)g_k^{\text{Adj}}(t)} \\ &\le \hat{\lambda}^{\text{Adj}}(t)\bar{g}_n(t)(1-a_n/n) - \frac{\rho}{n(1+\rho g^*(t))}\sum_{k=1}^{n}(\hat{\lambda}^{\text{Adj}}(t)g_k^{\text{Adj}}(t))^2 \\ &= \hat{\lambda}^{\text{Adj}}(t)\bar{g}_n(t) - \frac{\rho}{n(1+\rho g^*(t))}\sum_{k=1}^{n}\left(\hat{\lambda}^{\text{Adj}}(t)g_k^{\text{Adj}}(t)\right)^2 + O_p(n^{-3/2}a_n). \end{aligned} \tag{6.3}$$

The inequality stated above is valid due to the non-negativity of the $(n+1)$th term in the second summation. According to [CVA08], for any given $\epsilon > 0$, we have

$$\frac{1}{n}\sum_{k=1}^{n}\left(\lambda^{\text{Adj}}(t)g_k^{\text{Adj}}(t)\right)^2 \ge 1-\epsilon. \tag{6.4}$$

Therefore, as long as $a_n = o_p(n)$, equation (6.3) implies that

$$\frac{\rho}{(1+\rho g^*(t))} \le \hat{\lambda}^{\text{Adj}}(t)\frac{\bar{g}_n(t)(t)}{(1-\epsilon)} = O_p(n^{-1/2}). \tag{6.5}$$

Thus, we get $\rho = O_p(n^{-1/2})$ and hence $\lambda^{\text{Adj}}(t) = O_p(n^{-1/2})$. Now, consider

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{k=1}^{n+1} \frac{g_k^{\text{Adj}}(t)}{1+\lambda^{\text{Adj}}(t)g_k^{\text{Adj}}(t)} \\ &= \bar{g}_n(t)(t) - \lambda^{\text{Adj}}(t)\hat{V}_n(t) + o_p(n^{-1/2}), \end{aligned} \tag{6.6}$$

where $\hat{V}_n = (1/n)\sum_{k=1}^{n} g_k^{\text{Adj}}(t)^2$. Hence, when $n \longrightarrow \infty$, $\lambda^{\text{Adj}}(t) = \hat{V}_n^{-1}\bar{g}_n(t) + o_p(n^{-1/2})$. Now, we expand $l^*(\theta(t))$ as follows

$$
\begin{aligned}
l^*(\theta(t)) &= \sum_{k=1}^{n+1} \log\left(1 + \lambda^{\mathrm{Adj}}(t) g_k^{\mathrm{Adj}}(t)\right) \\
&= \sum_{k=1}^{n+1} \left\{ \lambda^{\mathrm{Adj}}(t) g_k^{\mathrm{Adj}}(t) - \frac{\left(\lambda^{\mathrm{Adj}}(t) g_k^{\mathrm{Adj}}(t)\right)^2}{2} \right\} + o_p(1).
\end{aligned}
\tag{6.7}
$$

Substituting the expansion of λ^{Adj}, we get that

$$
\begin{aligned}
-2l^*(\theta(t_0)) &= n\hat{V}_n^{-1}\bar{g}_n(t)^2 + o_p(1) \\
&\xrightarrow{d} \chi_1^2.
\end{aligned}
\tag{6.8}
$$

This completes the proof. □

Bibliography

[AJ06] S. Arora and K. Jain. Testing for generalized lorenz dominance. *Statistical Methods and Applications*, 15:75–88, 2006.

[Arv69] J. N. Arvesen. Jackknifing u-statistics. *Ann. Math. Stat*, 40:2076–2100, 1969.

[BH07] N Belinga-Hall. Empirical likelihood confidence intervals for generalized lorenz curve. *Thesis, Georgia State University*, 2007.

[CH97] R. Chang and N. Halfon. Graphical distribution of pediatricians in the united states: An analysis of the fifty states and washington, dc. *Pediatrics*, 100:172–179, 1997.

[CVA08] J. Chen, A.M. Variyath, and B. Abraham. Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17(2):426–443, 2008.

[Gas71] J. L. Gastwirth. A general definition of lorenz curve. *Econometrica*, 39:1037–1039, 1971.

[JTZ17] B-Y. Jing, M. Tsao, and W. Zhou. Transforming the empirical likelihood towards better accuracy. *The Canadian Journal of Statistics*, 45(3):340–352, 2017.

[JYZ09] B. Y. Jing, J. Yuan, and W. Zhou. Jackknife empirical likelihood. *Journal of the American Statistical Association*, 104(487):1224–1232, 2009.

[KT92] Y. Kobayashi and H. Takaki. Geographic distribution of physicians in japan. *Lancet.*, 340:1391–1393, 1992.

[Lor05] M. C. Lorenz. Method of measuring the concentration of wealth. *J. Ame. Statist.*, 9:209–219, 1905.

[LW18] Z. Li and Y. Wei. Statistical inference for the difference of two lorenz curves. *Statistics*, 52(5):1128–1155, 2018.

[Owe88] A. B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249, 1988.

[Owe90] A. B. Owen. Empirical likelihood confidence regions. *The Annals of Statistics*, 18:90–120, 1990.

[Que56] M. H. Quenouille. Notes on bias in estimation. *Biometrika*, 43(3/4):353–360, 1956.

[RWAR23] S. Ratnasingam, S. Wallace, I. Amani, and J. Romero. Non-parametric confidence intervals for generalized lorenz curve using modified empirical likelihood. *(submitted)*, 2023.

[Shi84] X. Shi. The approximate independence of jackknife pseudo-values and the bootstrap methods. *J. Wuhan Inst. Hydra-Elect. Eng.*, 2:83–90, 1984.

[Xu97] K. Xu. Asymptotically distribution-free statistical test for generalized lorenz curves: An alternative approach. *Journal of income distribution*, pages 45 – 62, 1997.

[YQBH12] B. Y. Yang, G. S. Qin, and N. E. Belinga-Hill. Non-parametric inferences for the generalized lorenz curve. *Sci Sin Math*, 42(3):235–250, 2012.