Spring 5-10-2014

Influence Function-based Empirical Likelihood Inferences for Lorenz Curve

Bing Liu
Georgia State University

INFLUENCE FUNCTION-BASED EMPIRICAL LIKELIHOOD INFERENCES FOR LORENZ CURVE

by

BING LIU

Under the Direction of Dr. Gengsheng Qin

ABSTRACT

In this thesis, an empirical likelihood method based on influence function is developed and used to construct confidence intervals for the Lorenz ordinates. This method is defined under the simple random sampling and the limiting distribution of the proposed empirical likelihood ratio statistic is a standard Chi-square distribution. Extensive simulation studies are conducted to evaluate the proposed empirical likelihood-based confidence intervals for the Lorenz ordinates. Finally, this method is used on a real income data as an application.

INDEX WORDS: Empirical likelihood, Influence function, Income distribution, Lorenz curve, Chi-square distribution.

INFLUENCE FUNCTION-BASED EMPIRICAL LIKELIHOOD INFERENCES FOR LORENZ CURVE

by

BING LIU

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2014

INFLUENCE FUNCTION-BASED EMPIRICAL LIKELIHOOD INFERENCES FOR LORENZ CURVE

by

BING LIU

Committee Chair: Dr. Gengsheng Qin

Committee: Dr. Xin Qi
Dr. Ruiyan Luo

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2014

DEDICATION

This thesis is dedicated to Georgia State University.

ACKNOWLEDGEMENTS

This thesis work would not have been possible without the support of many people. First, I want to express my gratitude to my advisor Dr. Gengsheng Qin, who gave me lots of suggestions and guidance while I was doing this thesis research. His insight and knowledge always inspired me and gave me strength when I faced a new challenge.

Furthermore, I want to thank my dear parents and my relatives who always support me once I make my decision to do something. This power is big enough to keep me carrying on. I love you all!

What's more, my gratitude also goes to our Chair Dr. Guantao Chen, who offered me advice and help. To Dr. Xin Qi and Dr. Ruiyan Luo, my committee members who gave me advice on this thesis. To my undergraduate advisor, Dr. Danhui Yi, who led me into the field of statistics and gave me opportunities to practice while learning new stuff in different views and solving real problems using statistical tools. To Dr. Yichuan Zhao, who gave wonderful lectures on different topics on statistics and biostatistics from which I benefit a lot.

Last but not least, I would like to thank my friends who helped me when I faced difficulties while I finished writing this thesis. The friends include but not limited to: Chenxue Li, Yanan Yin, Yuan Liu and Jin Zhao.

Thank you all who helped me and let's make our outlook brighter!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

- GSU - Georgia State University
- EL - Empirical Likelihood
- s.r.s - simple random sampling
- IF - Influence Function
- CI - Confidence Interval

CHAPTER 1

INTRODUCTION

In economics, the Lorenz curve is a graphical representation of income distribution, where it shows the percentage of the total income that the bottom (100*t)% of households have. It was developed by Max O. Lorenz in 1905 [1] for representing inequality of the wealth distribution. If we put the income data in order, the Lorenz curve illustrates the percentage of total income earned by different proportions of the whole population. Let X be a non-negative random variable with a cumulative distribution function $F(x)$ and assume that $F(x)$ is differentiable, Gastwirth (1971)[2] provided a definition of Lorenz curve as below:

$$\eta(t) = \frac{1}{\mu}\int_0^{\xi_t} x dF(x), \ t \in [0,1], \tag{1.1}$$

where $\mu = \int_0^\infty x dF(x)$ is the mean of the distribution F, and $\xi_t = F^{-1}(t)$ 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 householders, and, the mean income of total householders . Figure 1.1 is a graph describing Lorenz curve [3].

The Lorenz curve has been widely used in different disciplines. In the field of economics and social sciences, it provides a way for the partial ordering of income distributions (Atkinson, 1970)[4], and analyzing income and earning inequality (Doiron and Barrett, 1996[5]; Sen, 1973[6]). The analysis of Lorenz curve have also been applied in industrial concentration (Hart, 1971)[7], reliability (Gail and Gastwirth, 1978)[8], and medical and health services research (Chang and Halfon, 1997[9]; Hallas and Stovring,2006[10]).

However, the income distribution F is rarely known in practice, so we need to estimate the Lorenz curve from the sample income data. Let X_1, X_2, ..., X_n be an independent sample from F, then the Lorenz curve can be empirical estimated by

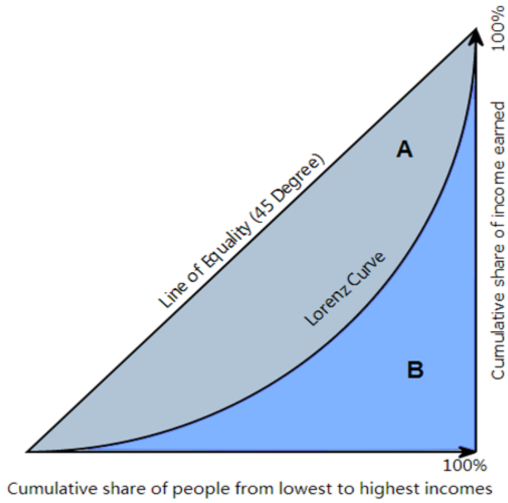

Figure 1.1 Lorenz Curve

$$\hat{\eta}(t) = \frac{1}{\hat{\mu}} \int_0^{\hat{\xi}_t} x dF_n(x), \tag{1.2}$$

where $\hat{\mu}$ is the sample mean, F_n is the empirical distribution function of the sample, $\hat{\xi}_t = inf\ \{y : F_n(y) \geq t\}$ is the t-th sample quantile.

Beach and Davison (1983)[11] has developed the asymptotic theory for $\hat{\eta}$. However, the existing normal approximation-based inferential methods have poor performance when the population distribution F is skewed and t falls in the tails of the Lorenz curve.

Empirical likelihood (EL), introduced by Owen (1988[12], 1990[13])), is a powerful non-parametric method. Hall and La Scala (1990)[14] discussed its advantage over the normal approximation-based methods. Over last two decades, EL has been widely applied in many areas such as survey sampling, medical studies, and econometrics. In the area of survey sampling, Chen and Qin (1993)[15] proposed EL-based inferences for finite populations with auxiliary information. In the area of health care, Zhou et al. (2006)[16] developed EL-based inferences in censored cost regression models and showed EL-based method outperforms the existing methods. Chen and Qin (2003)[17] developed EL-based inferences for data containing observations that are zero. Qin, Yang and Belinga (2013)[18] observed that most income data are skewed or highly skewed in economics study, so they developed new EL-based methods to make inferences for the Lorenz curve like hybrid bootstrap and EL approach.

In this thesis, a new approach is proposed by using empirical likelihood methods based on influence function (Zheng, Zhao and Yu, 2012)[19] to make inference for Lorenz curve. In Chapter 2, the profile empirical likelihood (EL) ratio statistic for Lorenz ordinate under simple random sampling (s.r.s) is reviewed, and the asymptotic distribution of the statistic is shown to be a scaled Chi-square distribution (Qin et al, 2013)[18]. In Chapter 3, the empirical likelihood based on influence function (IF) for Lorenz curve with simple random sampling is introduced and the asymptotic distribution of the empirical likelihood ratio statistic can be shown to be a standard Chi-square distribution, which is easier to be applied to real problems than those methods in Qin et al (2013)[18]. And in Chapter 4, my simulation study shows that the proposed influence function-based empirical likelihood confidence interval has good coverage probability under different nominal levels. In Chapter 5, a real data analysis is conducted. Finally, there are some discussions and conclusions, as well as some possible future work.

CHAPTER 2

A REVIEW OF EMPIRICAL LIKELIHOOD FOR THE LORENZ CURVE WITH SIMPLE RANDOM SAMPLE

2.1 Empirical Likelihood

Let $(X_1, ..., X_n)$ be a simple random sample from the population of X with cumulative distribution function F. For a fixed $t \in (0, 1)$, the Lorenz ordinate $\eta(t)$ satisfies:

$$E[X(I(X \leq \eta_t) - \eta(t))] = 0. \tag{2.1}$$

Thus, the empirical likelihood for $\eta(t)$ can be defined as:

$$\tilde{L}_1(\eta(t)) = \sup_{\boldsymbol{p}} \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i-1}^{n} p_i D_i(t) = 0 \right\}, \tag{2.2}$$

where $\boldsymbol{p} = (p_1, ..., p_n)$ is a probability vector, $D_i(t) = X_i[I(X_i \leq \xi_t) - \eta(t)]$, $i = 1, ..., n$.

2.2 Profile Empirical Likelihood

After substituting $\hat{\xi}_t = X_{[nt]}$, the t-th quantile estimated from sample, we have the profile empirical likelihood for $\eta(t)$:

$$L_1(\eta(t)) = \sup_{\boldsymbol{p}} \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i-1}^{n} p_i \hat{D}_i(t) = 0 \right\}, \tag{2.3}$$

where $\hat{D}_i(t) = X_i[I(X_i \leq \hat{\xi}_t) - \eta(t)]$, $i = 1, ..., n$.

A unique maximum for $\boldsymbol{p}$ exists if $\eta(t)$ is inside the convex hull of $\Big\{ X_1[I(X_1 \leq \hat{\xi}_t) - \eta(t)], ...$ $, X_n[I(X_n \leq \hat{\xi}_t) - \eta(t)] \Big\}$. By Lagrange multiplier method, the supremum is achieved at

$p_i = \frac{1}{n}\left\{1 + \nu(t)\hat{D}_i(t)\right\}^{-1}$, $i = 1, ..., n$, where $\nu(t)$ is the solution to:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{D}_i(t)}{1+\nu(t)\hat{D}_i(t)} = 0. \tag{2.4}$$

So the profile empirical likelihood ratio for $\eta(t)$ can be defined as:

$$R_1(\eta(t)) = \prod_{i=1}^{n}(np_i) = \prod_{i=1}^{n}\{1+\nu(t)\hat{D}_i(t)\}^{-1}. \tag{2.5}$$

And the corresponding profile empirical log-likelihood ratio for $\eta(t)$ is:

$$l_1(\eta(t)) = -2logR_1(\eta(t)) = 2\sum_{i=1}^{n}log\{1+\nu(t)\hat{D}_i(t)\}. \tag{2.6}$$

We have the following theorem from Qin et al (2013)[18].

***Theorem* 2.1** If $E(X^2) < \infty$, and $\eta(t_0) = E[XI(X \leq \xi_{t_0})]/E(X)$ for a given $t = t_0 \in (0, 1)$, then the limiting distribution of $l_1(\eta(t_0))$ is a scaled Chi-square distribution with degree of freedom 1.

That is,

$r_1 l_1(\eta(t_0)) \to^{\mathcal{L}} \chi_1^2$,

$r_1 = s_p^2(t_0)/s_d^2(t_0)$,

$s_p^2(t_0) = \int_0^\infty \{x[I(x \leq \xi_{t_0}) - \eta(t_0)]\}^2 dF(x)$,

$s_d^2(t_0) = \int_0^\infty \{(x - \xi_{t_0})[I(x \leq \xi_{t_0}) - x\eta(t_0)]\}^2 dF(x) - (t_0\xi_{t_0})^2$.

This Theorem can be used to construct confidence intervals for the Lorenz ordinates.

CHAPTER 3

INFLUENCE FUNCTION-BASED EMPIRICAL LIKELIHOOD FOR THE LORENZ CURVE WITH SIMPLE RANDOM SAMPLE

In the previous chapter, we can see if we want to construct confidence interval for the Lorenz ordinates, we need to find the unknown scale r_1, which is complicated. If we take advantage of the influence function when using empirical likelihood, the limiting distribution of the log-likelihood ratio statistic would be a standard chi-square distribution. Thus, we do not need to estimate the unknown scale. So in this chapter, influence function-based empirical likelihood for the Lorenz curve with simple random sample is introduced.

From results in Qin et al (2013)[18], we can derive:

$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_i(t_0)$

$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i[I(X \le \xi_{t_0}) - \eta(t_0)])$

$= \sqrt{n}\{\frac{1}{n} \sum_{i=1}^{n} [(X_i - \xi_{t_0}) I(X_i \le \xi_{t_0}) + t_0 \xi_{t_0} - X_i \eta(t_0)]\} + o_p(1)$

$= \sqrt{n}\{\frac{1}{n} \sum_{i=1}^{n} g(X_i, \eta(t_0))\} + o_p(1)$

where $g(X_i, \eta(t_0))$ is called the influence function, and

$$g(X_i, \eta(t_0)) = (X_i - \xi_{t_0}) I(X_i \le \xi_{t_0}) + t_0 \xi_{t_0} - X_i \eta(t_0). \tag{3.1}$$

Then, based on the estimated influence function $\hat{g}(X_i, \eta(t_0))$, where $\hat{g}(X_i, \eta(t_0)) = (X_i - \hat{\xi}_{t_0}) I(X_i \le \hat{\xi}_{t_0}) + t_0 \hat{\xi}_{t_0} - X_i \eta(t_0)$, we can define the influence function-based empirical likelihood for $\eta(t_0)$ as:

$$L_{IF}(\eta(t_0)) = \sup_{\boldsymbol{p}} \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i-1}^{n} p_i \hat{g}(X_i, \eta(t_0)) = 0 \right\}. \tag{3.2}$$

And the empirical likelihood ratio based on the estimated influence function can be defined as:

$$R_{IF}(\eta(t_0)) = \prod_{i=1}^{n}(np_i) = \prod_{i=1}^{n}\{1+\nu_{IF}(t_0)\hat{g}(X_i,\eta(t_0))\}^{-1}, \quad (3.3)$$

where ν_{IF} is the solution to:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}(X_i,\eta(t_0))}{1+\nu_{IF}(t_0)\hat{g}(X_i,\eta(t_0))} = 0. \quad (3.4)$$

And the corresponding influence function-based empirical log-likelihood ratio for $\eta(t_0)$ is:

$$l_{IF}(\eta(t_0)) = -2logR_{IF}(\eta(t_0)) = 2\sum_{i=1}^{n} log\{1+\nu_{IF}(t_0)\hat{g}(X_i,\eta(t_0))\}, \quad (3.5)$$

Then the following result gives the limiting distribution of the empirical likelihood based on influence function:

***Theorem* 3.1** If $E(X^2) < \infty$, and $\eta(t_0) = E[XI(X \leq \xi_{t_0})]/E(X)$ for a given $t = t_0 \in (0,1)$, then the limiting distribution of $l_{IF}(\eta(t_0))$ is a standard Chi-square distribution, i.e., $l_{IF}(\eta(t_0)) \to \chi_1^2$ as $n \to \infty$.

This theorem makes it much easier to construct confidence interval for the Lorenz ordinates than the profile empirical likelihood method. And Theorem 3.1 can be proved by using Theorem 2.1 in Zhang, Zhao and Yu (2012)[19]. Based on Theorem 3.1, a $(1-\alpha)$ level asymptotic confidence interval for Lorenz ordinate $\eta(t)$ can be constructed as $R_{IF} = \{\eta(t) : l_{IF}(\eta(t)) \leq \chi^2_{1,1-\alpha}\}$.

CHAPTER 4

SIMULATION STUDY

In this chapter, we're going to use simulation study to evaluate our method. First, recall the $(1-\alpha)$ confidence interval is defined as $R_{IF} = \{\eta(t) : l_{IF}(\eta(t)) \leq \chi^2_{1,1-\alpha}\}$, equivalently, we have $p(\eta(t) \in R_{IF}) = (1-\alpha) + o(1)$. In the simulation study, the coverage probabilities and the average lengths of the confidence intervals at different $t's$ are calculated to evaluate our method.

The coverage probabilities are calculated following these procedures:

1. Generate X_1, X_2, ... , X_n from a distribution F, n is the sample size.

2. Calculate $\eta(t_0)$ for fixed t_0, solve $\nu_{IF}(t_0)$ and get the value of log-likelihood $l_{IF}(\eta(t_0)) = -2logR_{IF}(\eta(t_0)) = 2\sum_{i=1}^{n} log\{1 + \nu_{IF}(t_0)\hat{g}(X_i, \eta(t_0))\}$.

3. Repeat 1 and 2 for B (a large number) times, then calculate the coverage probability:

$$\frac{1}{B}\sum_{b=1}^{B} I(\eta(t_0) \in R_{IF,b}) = \frac{1}{B}\sum_{b=1}^{B} I(l_{IF,b}(\eta(t_0)) \leq \chi^2_{1,1-\alpha}). \tag{4.1}$$

At the same time, the average lengths of the 95% confidence intervals are calculated from the generated samples.

When generating samples, we should notice that most income distributions are positively skewed, so the choice of underlying distribution F should be a positively skewed distribution.

Then we choose six different simulation settings as follows:

$n = 100,\ B = 10000,\ X_i \sim Weibull(1, 2)$

$n = 200,\ B = 10000,\ X_i \sim Weibull(1, 2)$

$n = 400,\ B = 10000,\ X_i \sim Weibull(1, 2)$

$n = 100,\ B = 10000,\ X_i \sim beta(2, 5)$

$n = 200,\ B = 10000,\ X_i \sim beta(2, 5)$

$n = 400,\ B = 10000,\ X_i \sim beta(2, 5)$

$t = (0.1, 0.15, 0.2, 0.25, ..., 0.85, 0.9)$, $\alpha = 0.1$, $\alpha = 0.05$

After the simulation, results are shown in the Appendices. In Appendix A, Figure A.1 - Figure A.12 show the coverage probabilities. From these results, we observe that the coverage probabilities of those confidence intervals for Lorenz ordinates are closer to the nominal levels (0.9, 0.95) as sample size increases. When t_0 falls in the lower and upper tails of Lorenz curve, the coverage probabilities are relatively lower than the nominal level. But as the sample size increases, this problem is less severe. In general, these confidence intervals have good coverage probabilities. These results are also shown in the Appendix B, Table B.1 and Table B.2. In Table B.3, it shows the average lengths of the 95% confidence intervals of Lorenz ordinates. To summary, the proposed method is a good way to construct confidence intervals for the Lorenz ordinates.

CHAPTER 5

A REAL EXAMPLE

In this chapter, I apply the influence function-based empirical likelihood method to make inference for Lorenz curve with a real income data.

Income inequality is a significant economic problem. The rising of the inequality is the highest in the United States among most developed countries (Weeks, 2007)[20]. By constructing confidence intervals for Lorenz ordinates at different t, we can discuss how the income inequality is in the United States and have a general view of the inequality.

The income data was selected from the database - The Panel Study of Income Dynamics (PSID) - from the University of Michigan. It is called the 2011 PSID Main Family Data, which contains two variables: 2011 Family Interview (ID) Number and Total Family Income-2010. The income reported here was collected in 2011 about tax year 2010. Please note that this variable can contain negative values. Negative values indicate a net loss, which in waves prior to 1994 were bottom-coded at $1, as were zero amounts. These losses occur as a result of business or farm losses. There are in total 8907 households in this dataset.

A brief summary of this income data is shown below:

Table 5.1 Summary of Income Data (amount in US$)

$Min.$	$1stQu.$	$Median$	$Mean$	$3rdQu.$	$Max.$
-70000	22620	45880	64870	83000	2420000

The histogram (Figure 5.1) shows that this income data is severely positively skewed. We applied our method and calculated the 95% confidence intervals for Lorenz ordinates at different t's. The results are shown in Table B.4. For example, when $t = 0.1$, the 95% confidence interval is $(0.003349568, 0.01156908)$, it means the ratio of the mean income of the lowest 10% households and the mean income of total households is greater than 0.0033, but smaller than 0.0116. In other words, the percentage of the income of the lowest 10% households out of the total households is between 0.33% and 1.16%. On the other hand, at $t = 0.9$, the 95% confidence interval is $(0.648726902, 0.66399601)$, so the the ratio of the mean income of the lowest 90% households and the mean income of the total households is greater than 0.6487, but smaller than 0.6640. Similarly, the percentage of the income of the lowest 90% households out of total households is between 64.87% and 66.40%. That means, the top 10% households owned almost 35% percent of the total income in total households in 2010. This is a huge difference between the lower 10% and the upper 10% households, which indicates the severity of the income inequality.

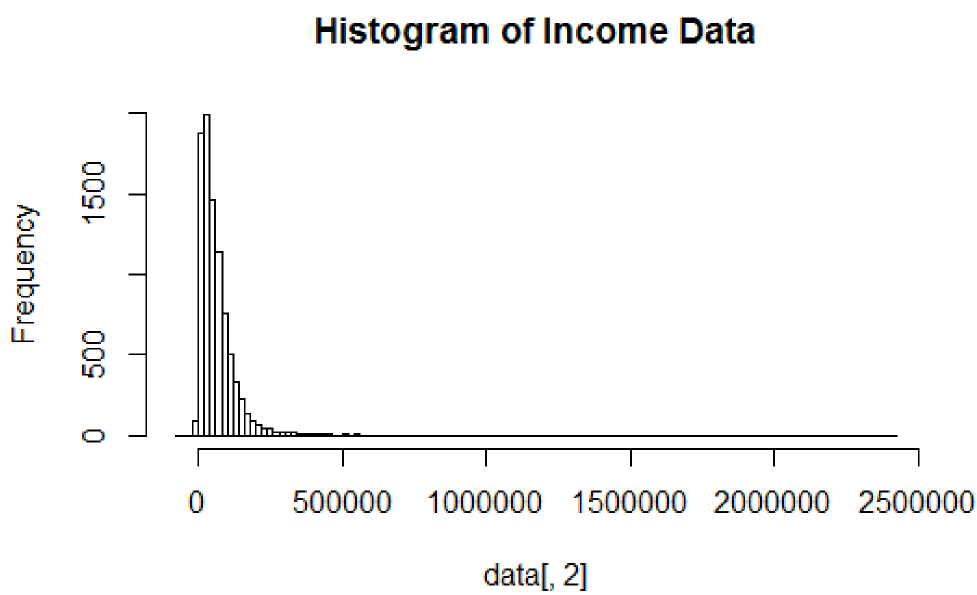

Figure 5.1 Histogram of Income Data

CHAPTER 6

CONCLUSIONS

In this thesis, an empirical likelihood method based on influence function is developed and used to construct confidence intervals for the Lorenz ordinates. This method is defined under the simple random sampling and the limiting distribution of the proposed empirical likelihood ratio statistic is a standard Chi-square distribution. Comparing with the profile empirical likelihood, we get rid of the constant scale which is complicated to calculate.

Simulation results shows good coverage probability and average lengths of the confidence intervals. This interval also has good coverage probabilities even at the lower and upper tail of the Lorenz curve, which gives us a better way to make inferences on Lorenz curve.

The real data example shows the confidence intervals for Lorenz ordinates at different t's, which give us a general view of the income inequality and how severe it is in the United States.

In the future, I will compare my method with the plug-in empirical likelihood method and the Jackknife empirical likelihood inferences for Lorenz curve to find the best one among these methods to make inferences for the Lorenz curve.

REFERENCES

[1] M. Lorenz, "Methods of measuring the concentration of wealth," *Journal of the American Statistical Association*, vol. 9, pp. 209–219, 1905.

[2] J. Gastwirth, "A general defination of lorenz curve," *Econometrika*, vol. 39, pp. 1037–1039, 1971.

[3] Wikipedia, "Plagiarism — Wikipedia, the free encyclopedia," 2009, [Online; accessed 14-Feburary-2014]. [Online]. Available: \url{http://en.wikipedia.org/wiki/File:Economics_Gini_coefficient2.svg}

[4] A. Atkinson, "On the measurement of inequality," *Journal of Economic Theory*, vol. 2, pp. 244–263, 1970.

[5] D. Doiron and G. Barrett, "Inequality in male and female earnings: the roles of hours and earnings," *Review of Economics and Statistics*, vol. 78, pp. 410–420, 1996.

[6] A. Sen, *On Economic Inequality.* New York: Norton, 1973.

[7] P. Hart, "Entropy and other measures of concentration," *Journal of the Royal Statistical Society: Series A*, vol. 134, pp. 73–89, 1971.

[8] M. Gail and J. Gastwirth, "A scale-free goodness-of-fit test for the exponential distribution based on the lorenz curve," *Journal of the American Statistical Association*, vol. 73, pp. 229–243, 1978.

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

[10] J. Hallas and H. Stovring, "Templates for analysis of individual-level prescription data," *Basic and Clinical Pharmacology and Toxicology*, vol. 98, pp. 260–265, 2006.

[11] C. Beach and R. Davison, "Distribution-free statistical inference with lorenz curves and income shares," *Review of Economic Studies*, vol. 50, pp. 723–735, 1983.

[12] A. Owen, "Empirical likelihood ratio confidence intervals for single functional," *Biometrika*, vol. 75, pp. 237–249, 1988.

[13] A. B. Owen, "Empirical likelihood ratio confidence regions," *The Annals of Statistics*, vol. 18, pp. 90–120, 1990.

[14] P. Hall and B. L. Scala, "Methodology and algorithms of empirical likelihood," *International Statistical Review*, vol. 58, pp. 109–127, 1990.

[15] J. Chen and J. Qin, "Empirical likelihood estimation for finite populations and the effective usage of auxiliary information," *Biometrika*, vol. 80, pp. 107–116, 1993.

[16] X. Zhou, G. Qin, H. Lin, and G. Li, "Inferences in censored cost regression models with empirical likelihood," *Statistica Sinica*, vol. 16, pp. 1213–1232, 2006.

[17] S. Chen and J. Qin, "Empirical likelihood-based confidence intervals data with possible zero observations," *Statistics Probability Letters*, vol. 65, pp. 29–37, 2003.

[18] G. Qin, B. Yang, and N. E. Belinga-Hall, "Empirical likelihood-based inferences for the lorenz curve," *Annals of the Institute of Statistical Mathematics*, vol. 65, no. 1, pp. 1–21, 2013.

[19] M. Zheng, Z. Zhao, and W. Yu, "Empirical likelihood methods based on influence functions," *Statistics and Its Interface*, vol. 5, pp. 355–366, 2012.

[20] J. Weeks, "Inequality trends in some developed oecd countries," *United Nations DESA Woking Paper*, vol. 6, pp. 1–14, 2007.

Appendix A

FIGURES FOR SIMULATION STUDY

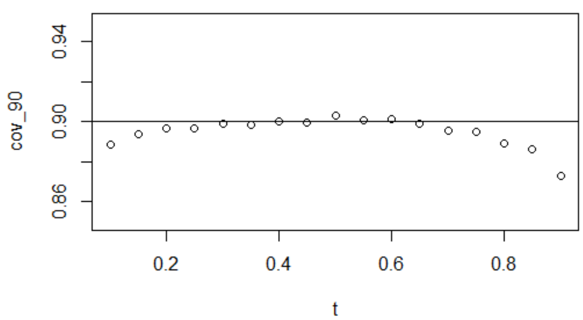

Figure A.1 Coverage probabilities of 90% level confidence intervals for Lorenz ordinates (n=100,Weibull(1,2))

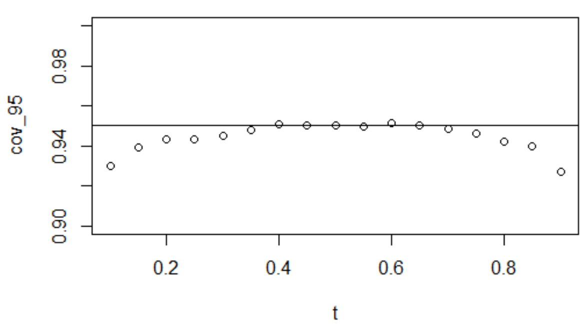

Figure A.2 Coverage probabilities of 95% level confidence intervals for Lorenz ordinates (n=100,Weibull(1,2))

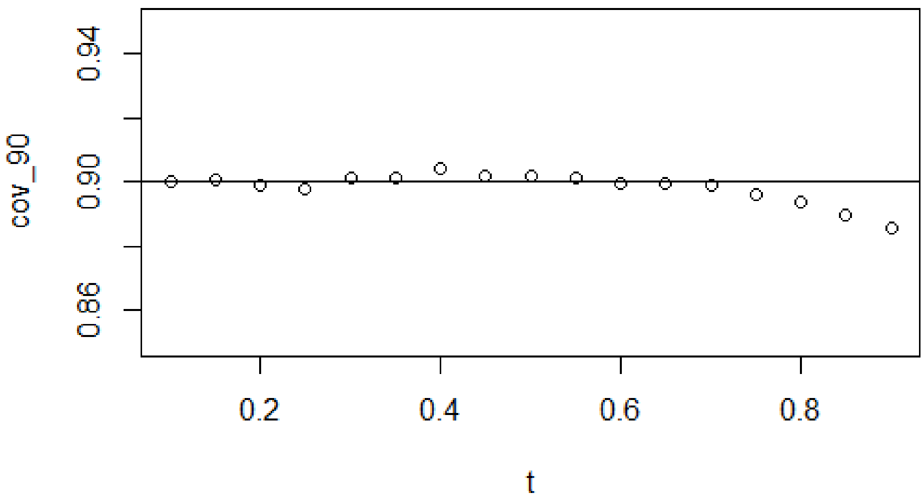

Figure A.3 Coverage probabilities of 90% level confidence intervals for Lorenz ordinates (n=200,Weibull(1,2))

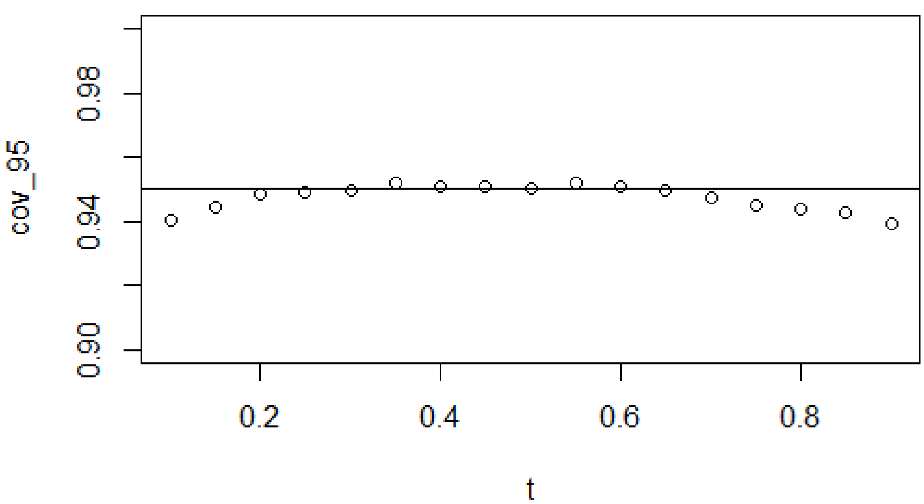

Figure A.4 Coverage probabilities of 95% level confidence intervals for Lorenz ordinates (n=200,Weibull(1,2))

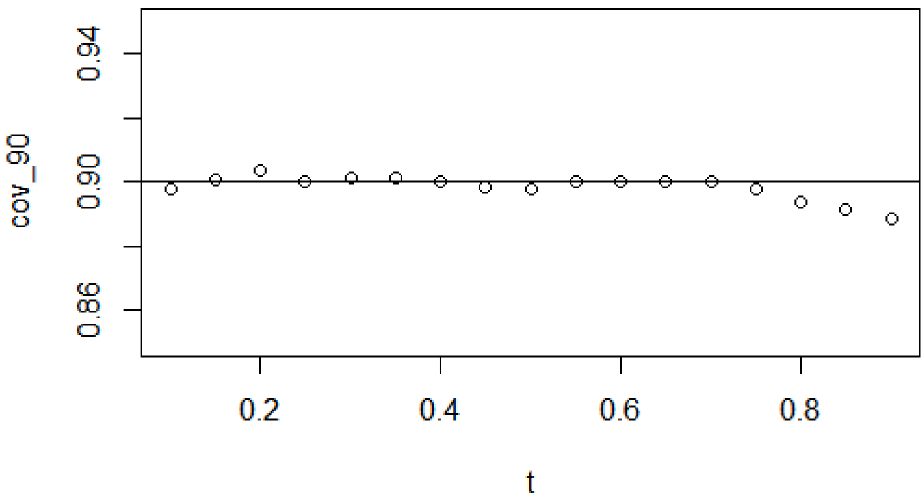

Figure A.5 Coverage probabilities of 90% level confidence intervals for Lorenz ordinates (n=400,Weibull(1,2))

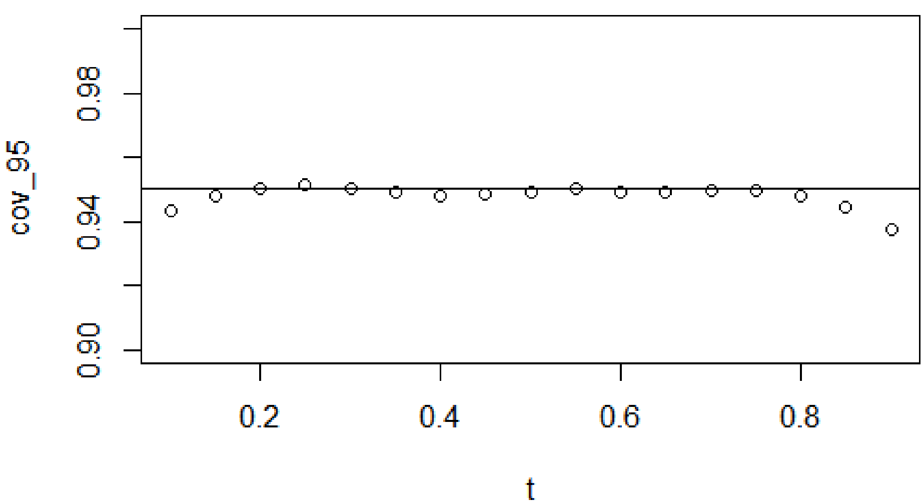

Figure A.6 Coverage probabilities of 95% level confidence intervals for Lorenz ordinates (n=400,Weibull(1,2))

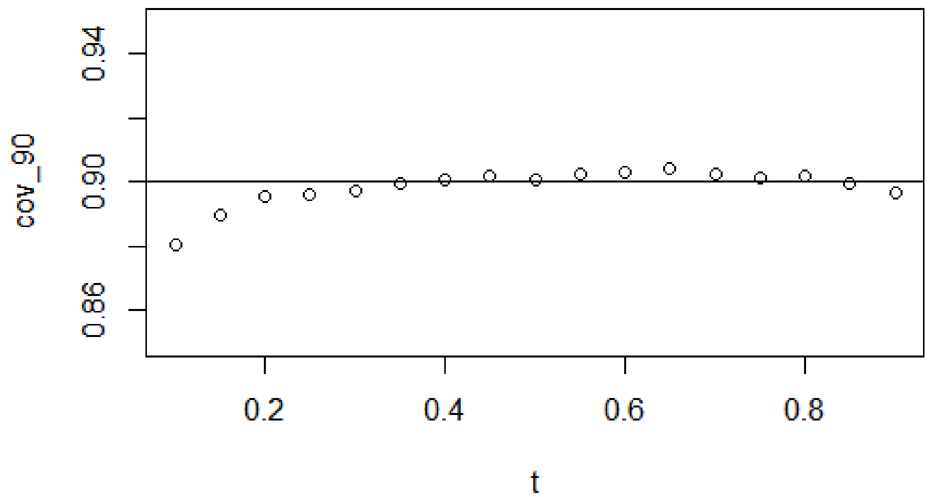

Figure A.7 Coverage probabilities of 90% level confidence intervals for Lorenz ordinates (n=100,beta(2,5))

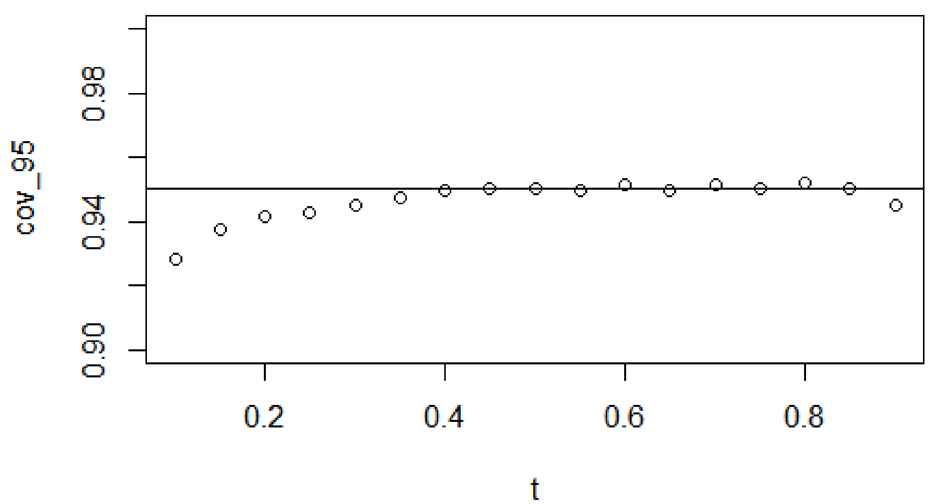

Figure A.8 Coverage probabilities of 95% level confidence intervals for Lorenz ordinates (n=100,beta(2,5))

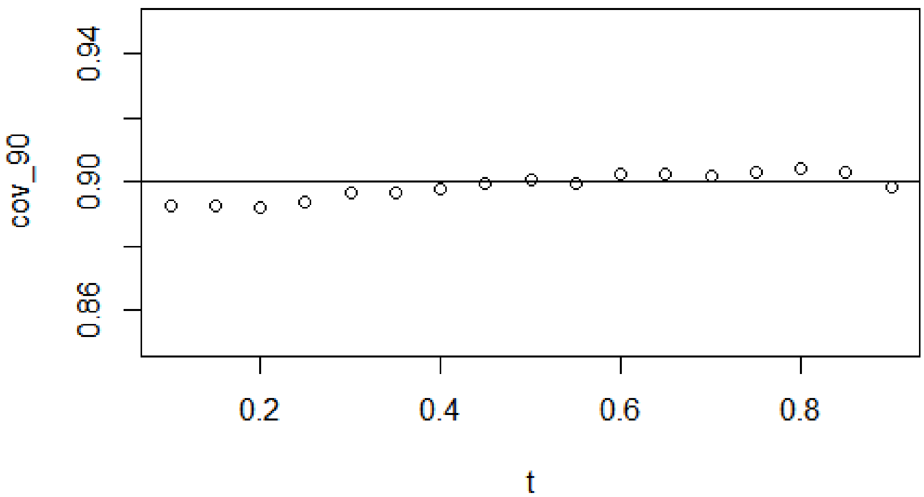

Figure A.9 Coverage probabilities of 90% level confidence intervals for Lorenz ordinates (n=200,beta(2,5))

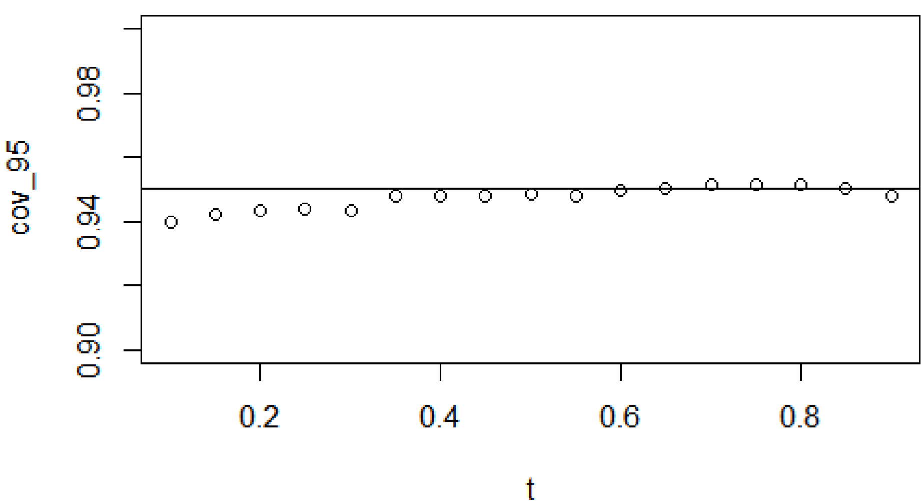

Figure A.10 Coverage probabilities of 95% level confidence intervals for Lorenz ordinates (n=200,beta(2,5))

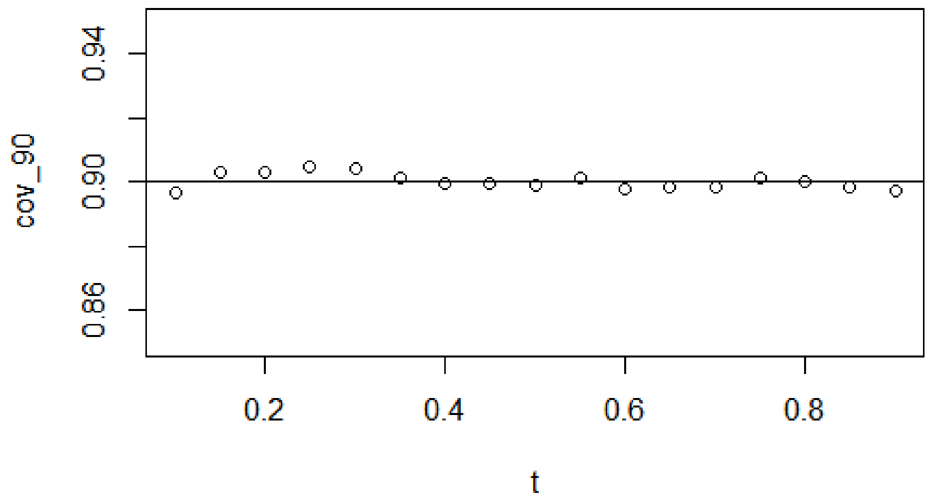

Figure A.11 Coverage probabilities of 90% level confidence intervals for Lorenz ordinates (n=400,beta(2,5))

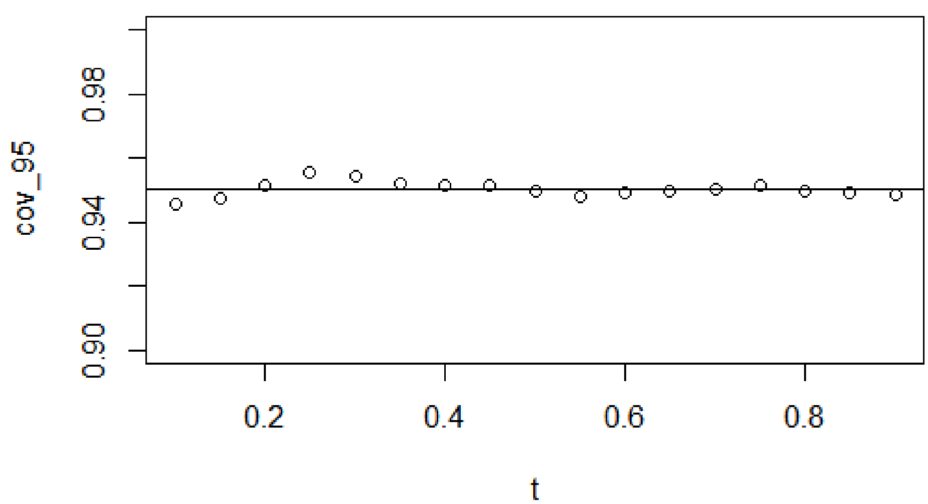

Figure A.12 Coverage probabilities of 95% level confidence intervals for Lorenz ordinates (n=400,beta(2,5))

Appendix B

COVERAGE PROBABILITIES AND AVERAGE LENGTHS OF CONFIDENCE INTERVALS

Table B.1 Coverage probabilities (Weibull(1,2))

$1-\alpha$	0.9			0.95		
n	100	200	400	100	200	400
t=0.10	0.8889	0.9001	0.8977	0.9299	0.9401	0.9435
t=0.15	0.8936	0.9008	0.9010	0.9393	0.9444	0.9478
t=0.20	0.8970	0.8990	0.9038	0.9434	0.9486	0.9502
t=0.25	0.8967	0.8982	0.9001	0.9433	0.9491	0.9513
t=0.30	0.8993	0.9014	0.9016	0.9452	0.9499	0.9503
t=0.35	0.8987	0.9012	0.9014	0.9481	0.9519	0.9493
t=0.40	0.9000	0.9041	0.9002	0.9507	0.9506	0.9481
t=0.45	0.8994	0.9021	0.8986	0.9502	0.9508	0.9485
t=0.50	0.9029	0.9019	0.8978	0.9501	0.9504	0.9491
t=0.55	0.9011	0.9016	0.9003	0.9497	0.9517	0.9501
t=0.60	0.9013	0.8999	0.9004	0.9513	0.9506	0.9488
t=0.65	0.8988	0.8997	0.9005	0.9501	0.9494	0.9491
t=0.70	0.8957	0.8990	0.9001	0.9483	0.9471	0.9495
t=0.75	0.8949	0.8964	0.8977	0.9460	0.9449	0.9498
t=0.80	0.8890	0.8941	0.8936	0.9421	0.9436	0.9480
t=0.85	0.8861	0.8899	0.8917	0.9400	0.9426	0.9444
t=0.90	0.8732	0.8858	0.8888	0.9270	0.9391	0.9373

Table B.2 Coverage probabilities (Beta(2,5))

$1-\alpha$	0.9			0.95		
n	100	200	400	100	200	400
t=0.10	0.8806	0.8927	0.8968	0.9283	0.9396	0.9455
t=0.15	0.8896	0.8925	0.9030	0.9372	0.9420	0.9474
t=0.20	0.8958	0.8922	0.9034	0.9414	0.9433	0.9516
t=0.25	0.8964	0.8936	0.9048	0.9426	0.9437	0.9556
t=0.30	0.8973	0.8970	0.9044	0.9452	0.9433	0.9540
t=0.35	0.8998	0.8965	0.9015	0.9475	0.9477	0.9518
t=0.40	0.9007	0.8980	0.8999	0.9496	0.9479	0.9512
t=0.45	0.9022	0.8998	0.8999	0.9504	0.9477	0.9514
t=0.50	0.9006	0.9010	0.8988	0.9504	0.9483	0.9495
t=0.55	0.9026	0.8998	0.9015	0.9497	0.9478	0.9480
t=0.60	0.9034	0.9023	0.8981	0.9513	0.9494	0.9491
t=0.65	0.9045	0.9025	0.8985	0.9497	0.9500	0.9498
t=0.70	0.9026	0.9021	0.8983	0.9512	0.9512	0.9504
t=0.75	0.9016	0.9031	0.9013	0.9504	0.9513	0.9515
t=0.80	0.9022	0.9043	0.9005	0.9522	0.9514	0.9496
t=0.85	0.8998	0.9030	0.8985	0.9503	0.9502	0.9493
t=0.90	0.8966	0.8985	0.8976	0.9452	0.9477	0.9486

Table B.3 Average lengths of 95% confidence intervals

Distn.	*Weibull*			*Beta*		
n	100	200	400	100	200	400
t=0.10	0.00945357	0.00692476	0.0535361	0.0323008	0.0264231	0.0268866
t=0.15	0.01454205	0.01002360	0.0537866	0.0351350	0.0283931	0.0276706
t=0.20	0.02128002	0.01450925	0.0539807	0.0385226	0.0309470	0.0286824
t=0.25	0.02901595	0.01989264	0.0546131	0.0424714	0.0335878	0.0300119
t=0.30	0.03727772	0.02568766	0.0550189	0.0466611	0.0360278	0.0315837
t=0.35	0.04603719	0.03179192	0.0560467	0.0511770	0.0382981	0.0332935
t=0.40	0.05489759	0.03816266	0.0571385	0.0555562	0.0405042	0.0348211
t=0.45	0.06401218	0.04462963	0.0557727	0.0595336	0.0427314	0.0359971
t=0.50	0.07303239	0.05117023	0.0527874	0.0626508	0.0447654	0.0368143
t=0.55	0.08177513	0.05760901	0.0523597	0.0650283	0.0464122	0.0373429
t=0.60	0.09041622	0.06378781	0.0510167	0.0664787	0.0474715	0.0376683
t=0.65	0.09833928	0.06958508	0.0504515	0.0668277	0.0477620	0.0377743
t=0.70	0.10517143	0.07477144	0.0530669	0.0659313	0.0471970	0.0376009
t=0.75	0.11056774	0.07896507	0.0560403	0.0636315	0.0456557	0.0371167
t=0.80	0.11345419	0.08166692	0.0580766	0.0596822	0.0430609	0.0362429
t=0.85	0.11281494	0.08199173	0.0585355	0.0537300	0.0396800	0.0344566
t=0.90	0.10596554	0.07808131	0.0568427	0.0458065	0.0360526	0.0314706

Table B.4 Pointwise 95% Confidence Intervals for Lorenz Ordinates

	Real Data lower	$n = 8907$ upper
t=0.10	0.003349568	0.01156908
t=0.15	0.007994029	0.02480974
t=0.20	0.028432227	0.03147866
t=0.25	0.043483721	0.04775264
t=0.30	0.060762805	0.06709955
t=0.35	0.083463703	0.09009548
t=0.40	0.108935447	0.11638391
t=0.45	0.138160473	0.14754500
t=0.50	0.175101618	0.18135262
t=0.55	0.211608438	0.21973006
t=0.60	0.254046440	0.26303637
t=0.65	0.294915464	0.31141553
t=0.70	0.349759799	0.36487000
t=0.75	0.410361397	0.42512349
t=0.80	0.479237680	0.49364741
t=0.85	0.557037935	0.57158249
t=0.90	0.648726902	0.66399601