

A comparative study of the Gini coefficient estimators based on the regression approach

Shahryar Mirzaei^a, Gholam Reza Mohtashami Borzadaran^{1,b}, Mohammad Amini^b,
Hadi Jabbari^b

^aDepartment of Statistics, Payame Noor University, Iran;
^bDepartment of Statistics, Ferdowsi University of Mashhad, Iran

Abstract

Resampling approaches were the first techniques employed to compute a variance for the Gini coefficient; however, many authors have shown that an analysis of the Gini coefficient and its corresponding variance can be obtained from a regression model. Despite the simplicity of the regression approach method to compute a standard error for the Gini coefficient, the use of the proposed regression model has been challenging in economics. Therefore in this paper, we focus on a comparative study among the regression approach and resampling techniques. The regression method is shown to overestimate the standard error of the Gini index. The simulations show that the Gini estimator based on the modified regression model is also consistent and asymptotically normal with less divergence from normal distribution than other resampling techniques.

Keywords: bootstrap technique, Gini coefficient, jackknife method, Lorenz curve, modified regression model, resampling techniques

1. Introduction

Measures of inequality are used to analyzing incomes, welfare, and poverty issues. They can also be helpful to measure the level of social stratification and polarization. Among many inequality measures, the Gini index is of the most widely known measures of income inequality due to its easy interpretation.

The Gini coefficient is the popular measure of income inequality; however, it is usually reported without any information about variance and ideally, with a standard error (SE). The reason for this is that most of the formulations of its proposed SE are mathematically complex or require considerable numerical computation.

Different methods to compute the SE of Gini index have been shown in various research on statistics and economics. Some related references are Woo and Yoon (2001), Xu (2007), Davidson (2009), Langel and Tillé (2013), and Yitzhaki and Schechtman (2013). Many authors have proposed using resampling approaches such as jackknife and bootstrap techniques to compute variance for the Gini estimator (Kang and Cho, 1996; Mills and Zandvakili, 1997; Moran, 2006; Palmitesta and Provasi, 2006; Yitzhaki, 1991).

Some authors have demonstrated that estimates of the Gini coefficient can be obtained from ordinary linear regression based on data and their ranks, thereby providing an exact analytic SE. Lerman

¹ Corresponding author: Department of Statistics, Ordered and Spatial Data Center of Excellence, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: grmohtashami@um.ac.ir

and Yitzhaki (1984) derived a convenient method to calculate the Gini coefficient, using the covariance between data and corresponding ranks. In this way, Shalit (1985) proposed a regression model based on a ranked variable for the set of natural numbers to compute the Gini index. Ogwang (2000) provided a method to compute the Gini index by an ordinary least square (OLS) regression, as well as discussed how to use the regression to simplify the computation of the jackknife SE. Giles (2004) later showed that the OLS SE from this regression could be used directly in order to compute the SE of the Gini index. Modarres and Gastwirth (2006) with simulations showed that the regression method overestimates the SE of the Gini estimator. They noted that the defect is due to the dependency between error terms in the proposed regression model. They also recommend complex or computationally intensive methods. However, Ogwang (2006) and Giles (2006) disagreed with this suggestion.

In this paper, according to simplicity of the regression approach method, we evaluate and compare this way with other resampling techniques. In this regard, we examine some special situations where the underlying distribution follows log-logistic (LL) and exponential (Exp) distributions as particular cases of generalized Beta II distribution (McDonald, 1984). In the next section, we introduce the concept of the Gini index which is discussed the income inequality. Section 3 then deals with resampling techniques such as bootstrap and jackknife, to measure the SE and confidence interval of the Gini estimator. In Section 4, with discussing shortcoming of the regression approach, we study this method for the variance estimation of the Gini estimator. Section 5 provides simulation evidence that proves the main conclusions of the paper and compares the some inferential statistics of Gini estimators such as consistency, divergence from normal distribution and asymptotic properties among the methods. Some graphical comparisons have also been done. Section 6 illustrated the results of the paper for real data on income inequality in Britain between 1994/1995. Conclusions are left for the last section.

2. The Gini coefficient

The most well-known member of the income inequality family is the Gini coefficient. It is widely used to measure income inequality because of its clear economic interpretation. This measure can be defined in various ways (Xu, 2003; Yitzhaki, 1998).

The best known definition of the Gini index of inequality is as twice the area between the 45° -line (equality line) and the Lorenz curve, as demonstrated in Figure 1. Therefore, it can be expressed as

$$G = 2 \int_0^1 (p - L(p))dp, \quad (2.1)$$

such that $p = F(x)$ is a cumulative distribution function (cdf) of income and $L(p)$ is the Lorenz function given by $(1/\mu) \int_0^p F^{-1}(t)dt$ (Gastwirth, 1971), where $\mu = E(X) > 0$ and $F^{-1}(p) = \inf\{x|F(x) \geq p; p \in [0, 1]\}$.

The Gini index takes values between 0 and 1. The value 0 corresponds to complete equality and 1 corresponds to complete inequality. This index is also defined as

$$G = \frac{\Delta}{2\mu},$$

where $\Delta = \int_0^\infty \int_0^\infty |x - y|dxdy$ is the population mean difference.

Using Expression (2.1), integration by parts and applying a change of variable $p = F(x)$, it can be found that (Xu, 2003):

$$G = \frac{2}{\mu} \left[\int_0^1 xF(x)dF(x) - \frac{\mu}{2} \right]. \quad (2.2)$$

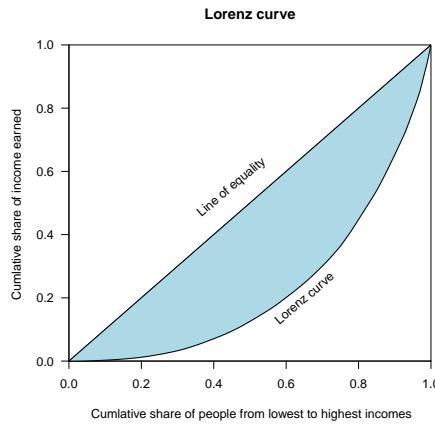


Figure 1: The area between the equality-line and the Lorenz curve.

This formula clearly shows an interpretation of the Gini coefficient in terms of covariance via

$$G = \frac{2}{\mu} \text{Cov}(X, F(X)). \quad (2.3)$$

Suppose that an identical independent distribution (iid) sample of size n is drawn randomly from the population, and let its empirical distribution function denoted as \hat{F}_n . It can be noticed that using (2.2), the natural plug in estimator of G by estimating F with \hat{F}_n defied as

$$\hat{G} = \frac{2}{\hat{\mu}} \left[\int_0^1 x \hat{F}_n(x) d\hat{F}_n(x) - \frac{\hat{\mu}}{2} \right].$$

In the context of discrete income distribution, Xu (2003) showed that the Gini index can be estimated by

$$\hat{G} = \frac{2 \sum_{i=1}^n i Y_i}{n^2 \bar{Y}} - \frac{n+1}{n},$$

where $0 \leq Y_1 \leq \dots \leq Y_n$ are the ordered income data.

Davidson (2009) proposed an approximate expression for the bias of \hat{G} , from which he subsequently derived the following bias-corrected estimator of the Gini coefficient, denoted \tilde{G} , which is given by:

$$\tilde{G} = \frac{n}{n-1} \hat{G}, \quad (2.4)$$

while the estimator is still biased but its bias is of order n^{-1} . Sometimes using this estimator is recommended because the properly bias-corrected estimator is not only even easier to compute rather than the other estimators, but also its bias converges to 0 as $n \rightarrow \infty$ faster.

3. Resampling techniques

Different methods with complicated formula to variance estimation of the Gini index have prompted significant research in statistics and economics. Most of the formulations of the variance for the

Gini estimator proposed in the literature are mathematically complex or require considerable numerical computation. To avoid these mathematical difficulties, various authors have proposed using the resampling techniques such as jackknife and bootstrap methods as follows.

3.1. The jackknife method

The jackknife provides a general purpose approach to estimating the bias and variance of an estimator. The jackknife is particularly useful when standard methods for computing bias and variance cannot be applied or are difficult to apply. Suppose that \hat{G} is an estimator of the Gini coefficient (G) based on sample of iid random variables X_1, \dots, X_n . Let $\hat{G}^{(i)}$ be the Gini estimator for the subsample of the initial sample where the i^{th} observation has been deleted, then the jackknife estimator to measure the Gini coefficient based on the n values of $\hat{G}^{(i)}$ is defined as:

$$\hat{G}_J = \hat{G} + \frac{n-1}{n} \sum_{i=1}^n (\hat{G} - \hat{G}^{(i)}), \quad (3.1)$$

and

$$b_n^J = \frac{n-1}{n} \sum_{i=1}^n (\hat{G} - \hat{G}^{(i)}), \quad (3.2)$$

which is the jackknife bias estimator (Yitzhaki, 1991). It follows that \hat{G}_J will be asymptotically unbiased (Jiang, 2010).

The jackknife method can also be used to estimate the SE of the Gini estimator. This was first noted by Yitzhaki (1991), whose SE estimate has the following expression:

$$\hat{\sigma}_J = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\hat{G}^{(i)} - \hat{G}_\bullet)^2}, \quad (3.3)$$

where

$$\hat{G}_\bullet = \frac{1}{n} \sum_{i=1}^n \hat{G}^{(i)}.$$

3.2. The Bootstrap method

The bootstrap method is an alternative approach to the variance estimation of Gini estimator. This technique is relatively straightforward, yet analytically powerful. Mathematical justifications can be quite sophisticated, the bootstrap method requires no theoretical calculations, applies identically to any income inequality measure, and is available no matter how mathematically complicated the parameter estimate or its asymptotic SE may be (Mills and Zandvakili, 1997). The bootstrap procedure is:

- Given a sample X_1, \dots, X_n of size n and an estimator \hat{G} .
- Draw B bootstrap samples of size n with replacement from X_1, \dots, X_n .
- Calculate the estimator for each one of them and obtain B values of the estimator, denoted by $\hat{G}_1^*, \dots, \hat{G}_B^*$.

Now, these values are used in order to estimate the variance of the original estimator. Namely, the sample variance of $\hat{G}_1^*, \dots, \hat{G}_B^*$ is used as the bootstrap variance estimator of the variance of the original statistic \hat{G} . The bootstrap SE of \hat{G} can then be estimated as:

$$\hat{\sigma}_{\text{Boot}} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{G}_b^* - \bar{G}^*)^2}, \quad (3.4)$$

where

$$\bar{G}^* = \frac{1}{B} \sum_{b=1}^B \hat{G}_b^*.$$

4. The regression method

The regression approach is the simplest way of computing a SE for the Gini coefficient. The computational difficulties or mathematical complexities associated with conventional formulas to compute the variance of the Gini coefficient make the use of simpler regression-based approach seems attractive. At first Lerman and Yitzhaki (1984) stated the Gini coefficient based on a covariance between the variable and its rank. Shalit (1985) proposed a regression model to compute the Gini index in the following form:

$$y_i = \alpha + \beta i + \xi_i, \quad i = 1, 2, \dots, n, \quad (4.1)$$

such that $0 \leq y_1 \leq y_2 \leq \dots \leq y_n$ and $\xi_1, \xi_2, \dots, \xi_n$ are errors with zero mean and homogenous variance σ^2 . He calculated the Gini estimator based on the estimated slope of this regression model as follows:

$$\hat{G} = \left(\frac{n^2 - 1}{6n} \right) \left(\frac{\hat{\beta}}{\bar{y}} \right), \quad (4.2)$$

where \bar{y} is the sample mean, $\hat{\beta}$ is the OLS estimator of β . To see an alternative regression interpretation of the Gini index, Ogwang (2000) also showed that \hat{G} can be written as:

$$\hat{G} = \frac{2\hat{\theta}}{n} - \frac{n+1}{n}, \quad (4.3)$$

such that $\hat{\theta} = \sum_{i=1}^n iy_i / \sum_{i=1}^n y_i$, is the weighted least squares estimator of θ in the following regression model:

$$i\sqrt{y_i} = \theta\sqrt{y_i} + v_i, \quad i = 1, 2, \dots, n, \quad (4.4)$$

where v_1, v_2, \dots, v_n are errors with mean zero and homogenous variance σ^2 . Davidson (2009) proposed regression modified model in (4.4) as:

$$\left(\frac{2i-1}{n} - 1 \right) \sqrt{y_i} = \theta^* \sqrt{y_i} + \xi_i, \quad i = 1, 2, \dots, n. \quad (4.5)$$

In this model, the Gini coefficient has an estimator of the slope of the regression model directly obtained and the SE of the Gini estimate can be computed.

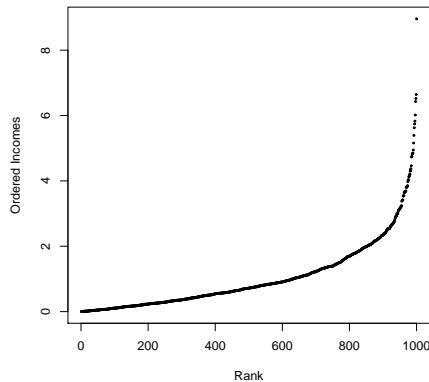


Figure 2: The dependency between ordered incomes under exponential distribution.

4.1. Shortcomings of the regression model approach

The regression model to estimating the SE of the Gini estimator cannot produce reliable results because it does not account for potential shortcomings introduced in the following:

- The regression model takes no account of the sampling design. It is used only for a random sampling technique.
- The independent variable in ordinary linear regression model is measured with no error; however, this model assumes that the independent variable is random.
- The normality of the error terms may be not satisfied and it should be tested.
- The relationship between ordered income and its rank is convex as demonstrated in Figure 2. Therefore, the error terms are dependent because the variance-covariance matrix of the errors is not diagonal. It has nonnegative element σ_{ij}/n , such that

$$\sigma_{ij} = \frac{p_i(1-p_j)}{f(\zeta_{p_i})f(\zeta_{p_j})}, \quad i \leq j, \quad i, j = 1, 2, \dots, n,$$

and

$$\sigma_{ij} = \sigma_{ji}, \quad i > j,$$

where ζ_{p_j} is the p_j^{th} population quantile and $f(\zeta_{p_j})$ is the positive and continuous density-quantile function for $0 < p_j < 1$ (David and Nagaraja, 2003).

It is important to note that actual data have these defects. Therefore, the method is used even though the defects are true. We must be cautious when using a regression-based approach to construct a SE for the Gini coefficient; in addition, tests for the validity of assumptions in the regression model must be formally conducted.

4.2. Asymptotic behavior based on regression approach

We can suppose that the estimator of Gini coefficient is unbiased and the asymptotic properties can be concluded if the regression model in (4.5) is true. Based on modified regression model in (4.5), the

slope estimate of the regression model is equal to the Gini estimator is:

$$\hat{G} = \frac{\sum_{i=1}^n \left(\frac{2i-1}{n} - 1 \right) y_i}{\sum_{i=1}^n y_i}. \quad (4.6)$$

The Gini estimator in (4.6) is the proportion of two functions that are linear combination of order statistics (L-statistics); therefore, asymptotical normality and consistency can be obtained with the theory of L-statistics (Sendler, 1979). \hat{G} is normally distribution based on theory of U-statistics and under some regularity conditions (Xu, 2007).

5. Simulation study

To comparative study of the Gini coefficient estimators based on bootstrap, jackknife and the regression approach in equation (4.5), we examined a special situation where the underlying distributions follow generalized Beta II (GB2) distribution. The most general distribution which has been proposed for fitting income data is the GB2 introduced in McDonald (1984). Its density is:

$$f_{GB2}(x|a, b, p, q) = \frac{ax^{ap-1}}{b^{ap} B(p, q)[1 + (x/b)^a]^{p+q}},$$

with $x > 0, a, b > 0$ and where $B(p, q)$ is the Beta function given by

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^\infty \frac{t^{p-1}}{(1+t)^{p+q}} dt, \quad p, q > 0.$$

The GB2 has the advantage that many densities can be obtained as a particular case and therefor constitutes a nice framework for discussion. Because of the complexity of the mathematical expression, we concentrate on Exp distribution with cdf:

$$F(x) = 1 - e^{-x}, \quad x > 0, \quad (5.1)$$

and LL distribution with cdf:

$$F(x) = 1 - \frac{1}{1+x^a}, \quad x > 0, \quad (5.2)$$

where $a > 0$, is a shape parameter ($a > 2$ leads to existence of the second moment of X).

In each stage, the simulations were undertaken by drawing 10,000 independent samples of size $n = 10, 20, 50, 100, 500, 1,000$ from the underlying distributions.

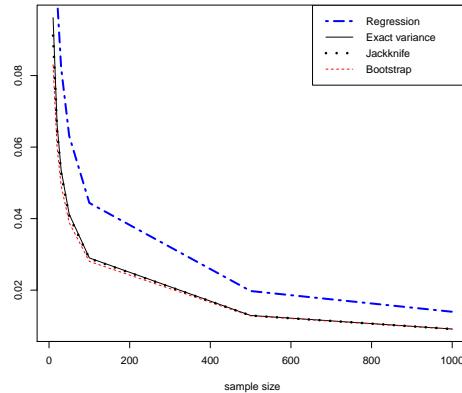
5.1. Variance inflation of Gini estimator in regression approach

It is important to have a correct method available to compute the SE of the Gini coefficient. Using a hypothesis test, it is possible when we use the SE of one method; the difference is significant but it is not when using another method. In Table 1, we write the values of the SE for Gini estimator with the regression model and other resampling techniques under Exp and LL distributions with parameter $a = 5$ as a benchmark. From these values, it is clear that the regression method underestimate the true value of Gini index and also overestimates the SE of the Gini estimator rather than other resampling techniques.

Table 1: Comparison the standard errors of Gini estimates

<i>n</i>	Exp			LL		
	Boot	Jack	Reg	Boot	Jack	Reg
10	0.0825	0.0904	0.1478	0.0391	0.0465	0.1867
20	0.0597	0.0636	0.1016	0.0314	0.0363	0.1292
50	0.0385	0.0403	0.0630	0.0225	0.0241	0.0807
100	0.0278	0.0288	0.0443	0.0168	0.0175	0.0568
500	0.0127	0.0128	0.0197	0.0081	0.0081	0.0253
1,000	0.0090	0.0091	0.0139	0.0058	0.0058	0.0178

LL = loglogistic distribution; Exp = exponential distribution; Boot = bootstrap; Jack = jackknife; Reg = regression.

**Figure 3:** Comparison of the standard errors under exponential distribution.

Three methods, the bootstrap, jackknife and the regression approach, provide comparable results. SE estimates using the bootstrap and jackknife are close to the real values, while those of the suggested regression model are noticeably inflated. The weakness of the regression based approach is essentially a finite sample matter, and its importance should diminish as $n \rightarrow \infty$. Figure 3 shows corresponding results for Exp distribution.

Table 2 reports the bias values, SE and mean square error (MSE) for the Gini estimator with the regression model underlying the LL distribution with parameter $a = 3, 10, 20$. It is clear that the change of these properties is influenced by the sample size and the values of a .

5.2. Consistency of Gini estimator in regression model

Table 3 presents the values of the MSE for Gini index with regression model underlying Exp and LL distributions with $a = 3$ as a benchmark. The results are shown in the following Table obtained by drawing 10,000 independent samples for the Gini index and $n = 10, 20, 50, 100, 500, 1,000$. The simulations show that the Gini coefficient based on parameter estimate in the modified regression model is consistent.

5.3. Asymptotic behavior

First, in order to see whether the asymptotic normality assumption yields a good approximation, simulations are undertaken with drawings 10,000 independent samples of size $n = 100$ for the Gini index from the Exp distribution, with cdf in (5.1). Using normality tests, it is evident that the consistency of

Table 2: Summary of regression approach under loglogistic distribution

a	n	Bias	Standard error	Mean square error
3	10	-0.04496	0.1780	0.0337
	20	-0.02211	0.1236	0.0157
	50	-0.00986	0.0774	0.0060
	100	-0.00369	0.0545	0.0029
	500	-0.00029	0.0243	0.0005
	1,000	-0.00017	0.0172	0.0002
10	10	-0.01057	0.1902	0.0363
	20	-0.00489	0.1314	0.0173
	50	-0.00186	0.0820	0.0067
	100	-0.00065	0.0577	0.0033
	500	-0.00013	0.0257	0.0006
	1,000	-0.00006	0.0181	0.0003
20	10	-0.00521	0.1911	0.0365
	20	-0.00242	0.1320	0.0174
	50	-0.00089	0.0823	0.0067
	100	-0.00032	0.0579	0.0033
	500	-0.00007	0.0258	0.0006
	1,000	-0.00003	0.0182	0.0003

Table 3: Consistency of Gini estimator in regression model for exponential distribution

MSE		n							
		10	20	50	100	500	1,000		
		Exp	LL	0.0243	0.0109	0.0048	0.0019	0.0003	0.0001
				0.0337	0.0157	0.0061	0.0030	0.0005	0.0002

MSE = mean square error; Exp = exponential distribution; LL = loglogistic distribution.

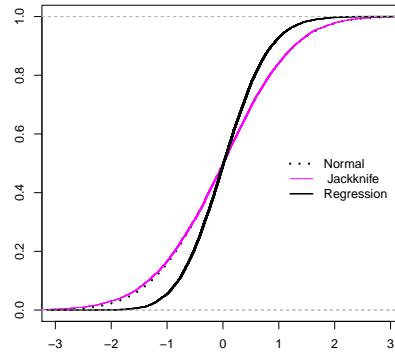


Figure 4: Comparison of the empirical distributions of jackknife and regression statistics.

the Gini estimator with the normal distribution is high.

In comparison normality of estimators based on jackknife and the regression approach, Figure 4 shows the empirical distributions for $n = 100$ of the statistics $\tau_{\text{Reg}} = (\hat{G} - G_0)/\hat{\sigma}_{\text{Reg}}$ and $\tau_{\text{Jack}} = (\hat{G}_J - G_0)/\hat{\sigma}_{\text{Jack}}$. Here $\hat{\sigma}_{\text{Reg}}$ is the regression SE estimator from regression (4.5) and $\hat{\sigma}_{\text{Jack}}$ is the jackknife SE estimator from (3.3) and \hat{G}_J is given by (3.1). Note that G_0 , the true value of the Gini index, for Exp distribution is 0.5.

Both of these statistics have distributions that are close to the standard normal distribution, but the jackknife estimator has better behavior of normality than the regression estimator.

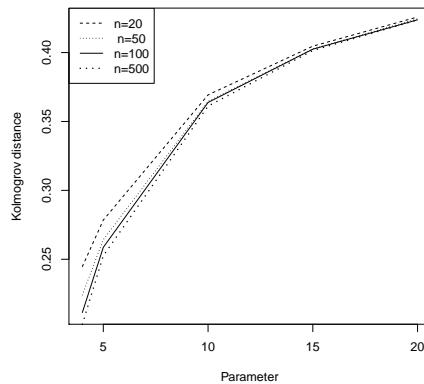


Figure 5: The trend of Kolmogorov distances with respect to the values of parameter a .

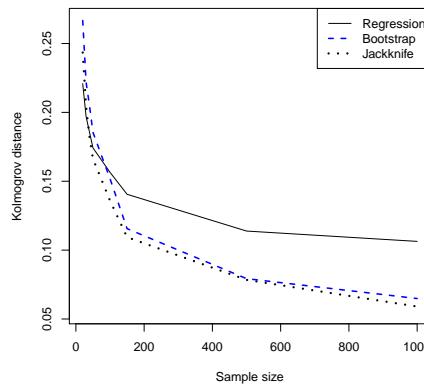


Figure 6: The divergence from $N(0, 1)$ for jackknife, bootstrap, and regression approach.

5.4. Gini regression estimate deviation from the normal distribution

In this section, we examine the divergence of Gini estimator in regression approach to normal distribution using the Kolmogorov distance under LL distribution with parameter a . The results are shown in Figure 5, where we show the Kolmogorov distances of sampling indices from a normal distribution influenced by parameter values of the distribution and sample size values. The concave curves express that the parameter value has the opposite effect.

Here, we compare the deviation of Gini estimates from normal distribution. Figure 6 explains divergence from normal distribution for jackknife, bootstrap and the regression approach estimators under Exp distribution. It is evident that the deviation of these estimates are asymptotically equivalent (especially for jackknife and bootstrap). It can be seen that the Gini estimate based on the jackknife method has a good performance in deviation from normal distribution.

5.5. Comparison of confidence intervals

In this section, we proposed three 95% confidence intervals for the Gini estimates of Exp distribution. Recall that the true value of the Gini index for this distribution is 0.5. The first column of Table 4 uses SE of the jackknife with $N(0, 1)$ critical values, the second is based on the percentile-t bootstrap confi-

Table 4: Confidence intervals for Gini estimates of exponential distribution

n	Jackknife	Bootstrap	Regression
10	(0.3222, 0.6363)	(0.2626, 0.6794)	(0.1596, 0.7399)
20	(0.3735, 0.6012)	(0.3497, 0.6265)	(0.2758, 0.6747)
50	(0.4199, 0.5682)	(0.4115, 0.5799)	(0.3662, 0.6138)
100	(0.4437, 0.5503)	(0.4395, 0.5570)	(0.4085, 0.5822)
500	(0.4748, 0.5239)	(0.4739, 0.5251)	(0.4605, 0.5377)
1,000	(0.4820, 0.5174)	(0.4816, 0.5177)	(0.4723, 0.5270)

Table 5: Comparison the Gini estimates in Britain in fiscal years 1994–1995

Method	\hat{G}	Standard error(\hat{G})	Confidence Interval
Bootstrap	0.33185	0.00272	[0.32578, 0.33715]
Jackknife	0.33186	0.00190	[0.32623, 0.33750]
Regression	0.33185	0.00337	[0.32524, 0.33846]

Table 6: Divergence of Gini estimates from $N(0, 1)$ based on fitted distribution

n	Jackknife	Bootstrap	Regression
10	0.2321	0.2716	0.2297
20	0.2061	0.2389	0.1932
50	0.1618	0.1876	0.1809
100	0.1259	0.1372	0.1441
500	0.0612	0.1079	0.1304
1,000	0.0548	0.0663	0.1268

dence interval (Mills and Zandvakili, 1997), the third are confidence intervals based on the regression model.

It is evident that the asymptotic bootstrap and jackknife intervals are very similar and both are very much narrower than those computed with the SEs based on regression approach.

6. Empirical illustration

Here, we refer to the real data based on income inequality in Britain on the fiscal years 1994–1995. Estimation is based on the unit record data used to calculate the official income statistics, derived from Family Resources Surveys available from the UK Data Service's web site (<https://www.ukdataservice.ac.uk>). We first performed an analysis on a comparison of the SEs and three 95% confidence intervals. Table 5 provides the corresponding results.

According to probability plots and quantile plots, we fit the GB2 distribution to real data with scale parameter equal to $b = 227.84$ and shape parameters equal to $a = 2.99$, $p = 1$, and $q = 1$. This result has also been reported in Jenkins (2009). The parameters considered are the maximum likelihood estimates of the GB2 distribution based on incomes. The distribution was well-characterized by a Fisk distribution with $a = 2.99$ and $b = 227.84$. Here, we refer to Monte Carlo samples drawn from a Fisk distribution as well as performed an analysis on a comparison of the divergence of Gini estimates from normal distribution by using the Kolmogorov distance. Table 6 summarizes the results for better interpretation.

Table 7 presents the empirical MSEs of the Gini estimator based on resampling and regression model. It is clear that all of these methods have asymptotically equal MSE values.

Table 8 reports the relative frequencies of the 95% confidence intervals containing the true value of the inequality measure (coverage probability) and the sizes of confidence intervals (average size)

Table 7: MSEs of the Gini estimates based on fitted distribution

<i>n</i>	MSE		
	Jackknife	Bootstrap	Regression
10	0.007451	0.005592	0.033719
20	0.003862	0.002952	0.015768
50	0.001964	0.001498	0.006093
100	0.001020	0.000861	0.002993
500	0.000257	0.000247	0.000593
1,000	0.000132	0.000133	0.000296

MSE = mean square error.

Table 8: CP and AS of fitted distribution to real data

<i>n</i>	Jackknife		Bootstrap		Regression	
	CP	AS	CP	AS	CP	AS
10	0.7971	0.2864	0.7085	0.2358	0.6955	0.6219
20	0.8366	0.2298	0.7683	0.1944	0.7502	0.4804
50	0.8705	0.1655	0.8305	0.1471	0.8043	0.3033
100	0.8916	0.1253	0.8741	0.1158	0.8647	0.2136
500	0.9223	0.0619	0.9111	0.0605	0.9012	0.0954
1,000	0.9313	0.0451	0.9233	0.0442	0.9189	0.0675

CP = coverage probability; AS = average size.

of 10,000 confidence intervals, for different methods of the fitted distribution to real data.

The results show that the coverage accuracy of the resampling techniques confidence intervals is reasonably close to the nominal confidence level for a large sample size. As expected, there was no substantial difference in coverage probability (CP) and in average size (AS) for the two resampling techniques. The jackknife confidence interval performs best in terms of CP and AS at the cost of providing larger confidence intervals.

7. Conclusion

The regression approach is the simplest way to estimate the Gini coefficient and its SE; however, this analysis of the Gini estimator can produce weaker results because it does not account for potential shortcomings introduced in the proposed regression model. It is important to note that actual data has defects when this method is used, despite the known shortcomings. The Gini estimator based on the regression model is consistent and asymptotically normal with less divergence from normal distribution than other resampling techniques. This method does not require the grouping of individual data to economize on computations. In addition, the estimator can be analyzed by using standard statistical software. The weakness of this method decreases as the sample size grows; therefore, we should be cautious when using regression-based approach to analyze the Gini coefficient in small samples size.

References

- David HA and Nagaraja HN (2003). *Order Statistics*, John & Wiley, New York.
- Davidson R (2009). Reliable inference for the Gini index, *Journal of Econometrics*, **150**, 30–40.
- Gastwirth JL (1971). A general definition of the Lorenz curve. *Econometrica*, **39**, 1037–1039.
- Giles DEA (2004). Calculating a standard error for the Gini coefficient: some further results, *Oxford Bulletin Economics and Statistics*, **66**, 425–433.
- Giles DEA (2006). A cautionary note on estimating the standard error of the Gini index of inequality:

- comment, *Oxford Bulletin Economics and Statistics*, **68**, 395–396.
- Jenkins SP (2009). Distributionally-sensitive inequality indices and the GB2 income distribution, *Review of Income and Wealth*, **55**, 392–398.
- Jiang J (2010). *Large Sample Techniques for Statistics*, Springer Science, New York.
- Kang SB and Cho YS (1996). Estimation of Gini index of the exponential distribution by bootstrap method, *Communications for Statistical Applications and Methods*, **3**, 291–297.
- Langel M and Tillé Y (2013). Variance estimation of the Gini index: revisiting a result several times published, *Journal of the Royal Statistical Society Series A (Statistics in Society)*, **176**, 521–540.
- Lerman RI and Yitzhaki S (1984). A note on the calculation and interpretation of the Gini index, *Economics Letters*, **15**, 363–368.
- McDonald JB (1984). Some generalized functions for the size distribution of income, *Econometrica*, **52**, 647–665.
- Mills JA and Zandvakili S (1997). Statistical inference via bootstrapping for measures of inequality, *Journal of Applied Econometrics*, **12**, 133–150.
- Modarres R and Gastwirth JL (2006). A cautionary note on estimating the standard error of the Gini index of inequality, *Oxford Bulletin Economics and Statistics*, **68**, 385–390.
- Moran TP (2006). Statistical inference for measures of inequality with a cross-national bootstrap application, *Sociological Methods & Research*, **34**, 296–333.
- Ogwang T (2000). A convenient method of computing the Gini index and its standard error, *Oxford Bulletin Economics and Statistics*, **62**, 123–129.
- Ogwang T (2006). A cautionary note on estimating the standard error of the Gini index of inequality: comment, *Oxford Bulletin Economics and Statistics*, **68**, 391–393.
- Palmitesta GMGP and Provaci C (2006). Asymptotic and bootstrap inference for the generalized Gini indices, *Metron*, **64**, 107–124.
- Sendler W (1979). On statistical inference in concentration measurement, *Metrika*, **26**, 109–122.
- Shalit H (1985). Calculating the Gini index of inequality for individual data, *Oxford Bulletin Economics and Statistics*, **47**, 185–189.
- Woo JS and Yoon GE (2001). Estimations of Lorenz curve and Gini index in a Pareto distribution, *Communications for Statistical Applications and Methods*, **8**, 249–256.
- Xu K (2003). How has the literature on Gini's index evolved in the past 80 years?, *China Economic Quarterly*, **3**.
- Xu K (2007). U-statistics and their asymptotic results for some inequality and poverty measures, *Econometric Reviews*, **26**, 567–577.
- Yitzhaki S (1991). Calculating jackknife variance estimators for parameters of the Gini method, *Journal of Business and Economic Statistics*, **9**, 235–239.
- Yitzhaki S (1998). More than a dozen alternative ways of spelling Gini, *Research in Economic Inequality*, **8**, 13–30.
- Yitzhaki S and Schechtman E (2013). *The Gini Methodology: A Primer on a Statistical Methodology*, Springer Science, New York.

Received December 3, 2016; Revised May 1, 2017; Accepted June 28, 2017