Article

A New Logit-Based Gini Coefficient

Hang K. Ryu ^{1,*}, Daniel J. Slottje ² and Hyeok Y. Kwon ³¹ Department of Economics, Chung Ang University, Seoul 156-756, Korea² Department of Economics, Southern Methodist University, Dallas, TX 75275, USA; dan.slottje@fticonsulting.com³ Department of Political Science, Korea University, Seoul 136-701, Korea; hkwon@korea.ac.kr

* Correspondence: hangryu@cau.ac.kr; Tel.: +82-10-5253-6500

Received: 21 March 2019; Accepted: 9 May 2019; Published: 13 May 2019

Abstract: The Gini coefficient is generally used to measure and summarize inequality over the entire income distribution function (IDF). Unfortunately, it is widely held that the Gini does not detect changes in the tails of the IDF particularly well. This paper introduces a new inequality measure that summarizes inequality well over the middle of the IDF and the tails simultaneously. We adopt an unconventional approach to measure inequality, as will be explained below, that better captures the level of inequality across the entire empirical distribution function, including in the extreme values at the tails.

Keywords: projection of share function; logit function; maximum entropy method; inequality measure

MSC: 62E17; 62P20; 91B15

1. Introduction

This paper introduces a new index of inequality that better captures information about the extreme values (the tail behavior) of observed distribution functions. The conventional approach to measuring inequality is to begin with an empirical distribution and to measure inequality in it by applying an index such as the Gini coefficient post facto (cf. Cowell [1] for an excellent discussion). We take the opposite approach; we begin with a given summary measure of inequality and then derive properties of the income distribution function (IDF) underlying it as expressed in the “shares” of the distribution. Ryu [2] was the first to use the “share function” as a vehicle to analyze inequality in empirical distribution functions. Following Ryu’s [2] work, Tanak et al. [3] and Rad et al. [4] cleverly apply the concept of the share function to an extended Gini measure (Tanak et al. [3]) and by introducing new specifications of the Lorenz curve (Rad et al. [4]), respectively. Here, we show that the income share ratio of the richest group to the poorest group of a given observed empirical income distribution can be derived accurately using the income distribution function derived from our new measure. We call our new measure the logit-based Gini coefficient (LBGC).

Several statisticians have examined the issue of measuring and describing inequality in statistical distribution functions (cf. Arnold and Sarabia [5], Arnold, Castillo, and Sarabia [6], Giorgi [7], Giorgi and Gubbiotti [8], and Nygard and Sandstrom [9,10] among others, for a broad discussion of many of these issues). For traditional income inequality measurement, the Gini coefficient is by far the most widely used index even though other measures exist. The seminal work of Lerman and Yitzhaki [11] and Yitzhaki [12] demonstrated that the Gini coefficient can be derived from the first moment of the observed income shares for a given batch of income data. During this filtering process, the Gini in effect only projects a linear trend and throws out the remaining information contained in the observed share values. This information loss problem can be remedied, and the major contribution of this paper is that it introduces one way to do so by projecting the income shares with the logit function.

The choice of a weighing function in the projection uniquely determines the inequality measure. The range of the logit function varies from minus infinity to plus infinity and it can detect and project changes in the very small shares of the IDF's poorest group's income shares and the IDF's richest group's income shares.

This approach can be applied to higher moments in an IDF as well. In Appendix B, projections with other moments such as the second moment, third moment, square root moment, and exponential moment are explained. Each of these other forms is a legitimate inequality index. However, the performance of these derived functions from their moments was not quite satisfactory and each alternative measure more or less provides the same level of information as contained in the original Gini coefficient. The reason for these relatively poorer performances comes from the finite range of the weighting function used in these other projections. The ranges of the first moment, second moment, and square root moment are restricted between zero and one. As noted above, Appendix B of this paper discusses the details of these other measures, but this paper focuses on the LBGC measure.

Because this paper introduces a new LBGC summary inequality measure, it is necessary to check to confirm the new inequality measure possesses desirable properties of an income inequality measure. These well-known properties are listed in Appendix A where it is also shown that the properties are satisfied by the new measure. For practical purposes, we also proffer the following properties, which are shown to be satisfied. (1) The income inequality measure should be sensitive to the movements of income shares of the poorest and richest groups. (2) The underlying IDF can be derived directly from the given summary measure and it is hoped the estimates will be accurate at all ranges of income groups. (3) If an economic interpretation can be given to the definition of the summary measure, it provides guidelines indicating how best to use the measure.

In the next section, we discuss the linear projection method utilized in applying the well-known Gini coefficient. The underlying income distribution function is derived by the maximum entropy method subject to the given Gini coefficient. In Section 3, we introduce the new summary inequality measure and the LBGC measure, and the corresponding underlying distribution function is derived. In Section 4, Current Population Survey data are utilized to show the usefulness and limitations of analysis made with the Gini coefficient and with the new LBGC measure. Section 5 concludes the paper.

2. Motivation for Our Analysis

To begin, we define an income inequality measure as a projection of the income share function with some weighting function. The Gini coefficient is derived by utilizing a linear weighting function, Theil's entropy measure utilizes the logarithm of the share as a weighting function, and the new LBGC measure is derived by imposing a logit function as the weighting function. To understand our introduction of the logit function, consider the following: suppose the income observations have a lognormal distribution. This means the logarithm of the income observations is normally distributed. It is well known that the poorest segments of the population across virtually all countries have (as groups) income shares very near to zero. Of course, the logarithm of a value approaching zero is a large negative number. Concomitantly, the normal distribution possesses a point of "high contact" in classical statistics, meaning the poorest segment in the left tail has a very small accumulated probability. With the inversion of the accumulated normal distribution, a small far left tail above corresponds to a small z value (z is near zero in our paper and is explicated below). However, the accumulated normal distribution does not have an explicit inverse functional form; if it did, it could have been used as a projection function. Therefore, we use the logit function because it does have an explicit functional form and the shapes of the probit and logit functions are similar.

Therefore, we use the logit function as a projection function in (13) below, because both the logit function and share function become very large for z near one. For the left tail of the distribution, when z approaches zero, the logit approaches negative infinity but the share function approaches zero. Therefore, based on the logit projection upon the share function in (13), entropy maximization produces a share function (17). The logarithm of this share function includes a logit form and converges to a

lognormal distribution. The difference, again, is that we now have an explicit functional form of share function based on the logit projection because we cannot derive an explicit functional form for the probit case. The imposition of other weighting functions will also yield other legitimate inequality measures. In Appendix B, a second moment, a third moment, a square root moment, and exponent moment are used as weighting functions. We present them for the reader's consideration only. The LBGC measure describes inequality in the overall IDF better, but these other measures are interesting.

2.1. Derivation of the Gini Coefficient with a Linear Projection Function

In order to show that the usual derivation of a Gini approximation neglects all the remaining information contained in observed share values, one can first introduce and then expand the share function with a Legendre polynomial series. Suppose we have an IDF (income distribution function) and we have income observations for 100 individuals $q_1, q_2, \dots, q_{100}$. We can plot a histogram with these 100 income observations. Alternatively, if these observations are normalized, the income share of the i -th person is $s_i = q_i / (q_1 + q_2 + \dots + q_{100})$. The share variable is a random variable. As an example, when a dollar bill falls from the sky, the probability it will land on the i -th person is s_i . Different people have different capabilities of picking up the dollar bill as it falls from the sky. For a continuous share function $s(z)$ for $0 \leq z \leq 1$, the income shares of the poorest and richest are

$$s_1 = \int_0^{0.01} s(z) dz \text{ and } s_{100} = \int_{0.99}^1 s(z) dz.$$

We can show that the Gini coefficient can be derived from the first moment of the share function. Suppose n is the number of persons and assume $n-1$ is approximately equal to n . Using the Lorenz curve definition,

$$Gini = 1 - \frac{2}{n-1} \left(n - \frac{\sum_{i=1}^n i q_i}{\sum_{i=1}^n q_i} \right) = 1 - \frac{2}{n-1} \left(n - \frac{\sum_{i=1}^n i s_i}{\sum_{i=1}^n s_i} \right) \approx 1 - 2(1 - \mu)$$

The first moment becomes $\mu = \sum_{i=1}^n i s_i / n \approx \int_0^1 z s(z) dz$.

The share function is related to the underlying income distribution by virtue of the fact that a Lorenz curve underlies (can be constructed from) any income distribution function and the first derivative of the Lorenz curve is the share function. Conversely, if one were to accumulate the share function piecewise, this accumulation would become a Lorenz curve. Tanak et al. [3] and Rad et al. [4] also recognized the issue of the Gini index being a poor indicator of movements in the tails of the empirical distribution and how using entropy and the attendant share function can be specified in a way that can more accurately capture the tail behavior of the observed distribution. It should be noted that Gastwirth [13] dissented from this view and showed a situation where the Gini coefficient is more sensitive to changes in the lower and upper parts of the distribution than in the middle. However, our point is not about the sensitivity of the Gini coefficient per se to changes in an income distribution. Rather, the income shares of the top 5% of incomes can be reproduced very accurately by our new LBGC measure whereas those of the Gini were not very good. We are concerned less with the direction and sensitivity of the reproduced results by the Gini coefficient but rather about the magnitude and direction of reproduced results relative to those reproduced by the LBGC measure. Figure 1 also illustrates the relatively poorer performance of the Gini for the wealthiest segment of the distribution.

The generalized Gini index is defined as (see papers of Rad et al. [4] at p. 2911 and Tanak et al. [3] at p. 281)

$$G_v = 1 - v(v-1) \int_0^1 (1-p)^{v-2} L_x(p) dp, v > 1$$

where $L_x(p)$ is the Lorenz curve of population proportion p . If $v = 2$, the generalized Gini index is equivalent to the classical Gini index. The primary difference between the generalized Gini index from the classical Gini index is the following. The former uses an additional curvature parameter v but the “cost” is the loss of simplicity in describing income inequality with a single measure (either the Gini, or Theil, or others). For different data sets, a researcher would need to select different values of v . In our work, the income share functions can be derived based on a given set of conditions using the maximum entropy method. As will be shown below, our paper is based on a procedure that only requires knowledge of a single parameter Equation (13) or equivalently Equation (16), whereas the procedures of Rad et al. [4] and Tanak et al. [3] require knowledge of both v and G_v . Therefore, the derived share functions in their work become functions of z and v , instead of just on z as in our work. Arfken [14] and Milne [15] present a good explanation of the orthonormal basis method and the use of Legendre polynomials.

$$s(z) \approx s_N(z) = \sum_{n=1}^N a_n P_n(z) \quad (1)$$

Figure 1. LBGC estimated shares and Gini estimated shares.

The Legendre polynomials are the following for $0 \leq z \leq 1$:

$$\begin{aligned} P_0(z) &= 1 \\ P_1(z) &= \sqrt{3}(2z - 1) \\ P_2(z) &= \sqrt{5}(6z^2 - 6z + 1) \\ P_3(z) &= \sqrt{7}(20z^3 - 30z^2 + 12z - 1) \\ P_4(z) &= \sqrt{9}(70z^4 - 140z^3 + 90z^2 - 20z + 1) \\ P_5(z) &= \sqrt{11}(252z^5 - 630z^4 + 560z^3 - 210z^2 + 30z - 1) \end{aligned} \quad (2)$$

An orthonormal sequence satisfies

$$\int_z P_n(z)P_m(z)dz = \delta_{nm}, n, m, = 0, 1, 2, \dots \quad (3)$$

where $\delta_{nm} = 1$ if $n = m$ and zero otherwise. The parameters of (1) can be found with

$$a_m = \int P_m(z)s_N(z)dz = \int P_m(z)\left[\sum_{n=1}^N a_n P_n(z)\right]dz \quad (4)$$

Lerman and Yitzhaki [11] argued that knowledge of the Gini coefficient is equivalent to the knowledge of the first moment. This assertion is proved above. They stated in their paper, “Using the regression coefficient yields a general graphical interpretation of the Gini. Consider a case of an honor guard of soldiers in which the soldiers are ordered by height from the shortest to the tallest and are equidistant from adjacent soldiers. Assume that the entire length of the honor guard is a distance of 1. Using this ordering and running a regression of height against the soldier’s position will yield a regression line that passes through the mean value (height) at the mean rank (or the 50th percentile). The slope of this regression line is the value of the absolute Gini times a constant.” (Lerman and Yitzhaki [11], p. 365)

If the Gini coefficient utilizes only the first moment,

$$\mu_1 = \int zs(z)dz = \frac{1 + Gini}{2} \approx \int z s_N(z)dz = \int z \sum_{n=1}^N a_n P_n(z)dz \quad (5)$$

Use $z = \frac{P_1(z)}{2\sqrt{3}} + \frac{P_0(z)}{2}$.

$$\mu_1 \approx \int \left[\frac{P_1(z)}{2\sqrt{3}} + \frac{P_0(z)}{2} \right] \sum_{n=1}^N a_n P_n(z)dz = \left[\frac{a_1}{2\sqrt{3}} + \frac{a_0}{2} \right] \quad (6)$$

Therefore, the Gini coefficient uses only a linear approximation of the share function (1) and it neglects all the remaining information contained in the raw data. To see why this is so, consider that we can derive the same Gini coefficient value even if we assume the raw data generating function is

$$s^*(z) = a_0 P_0(z) + a_1 P_1(z) \quad (7)$$

where the parameters a_0 and a_1 are the same as those of (1).

2.2. Derivation of a Share Function from the Given Gini Coefficient

If raw income data are unavailable, information about the underlying income distribution function can still be uncovered by extracting information from the summary index that is available, such as that contained in the given Gini coefficient. This paper now explains how that is done. This section

begins with a review of Ryu and Slottje [16,17], which is the first step towards our final objective of presenting more information-rich inequality measures. Solving an entropy maximization problem of Shannon [18] and Ryu [2] subject to the given first moment:

$$\text{Max}_s W \equiv - \int s(z) \log s(z) dz \quad (8)$$

satisfying

$$\int z s(z) dz = \mu_1 \quad (9)$$

where $s(z)$ is the share function of a person located at z . The Lagrangian method produces

$$s_{\text{Gini}}(z) = \exp[a + bz] = \left[\frac{b}{e^b - 1} \right] \exp[bz] \quad (10)$$

where the normalization condition of the share function is used to remove a . Now the first moment condition (5) produces

$$\mu_1 = \left[\frac{b}{e^b - 1} \right] \int_0^1 z \exp[bz] dz = \frac{1 + \text{Gini}}{2} \quad (11)$$

Since the integration is a function of b ,

$$\mu_1 \equiv -\frac{1}{b} + \frac{e^b}{e^b - 1} = \frac{1 + \text{Gini}}{2} \quad (12)$$

Then b approaches zero if the Gini = 0 and b approaches infinity if the Gini = 1. Since the center part of (12) is a monotonic increasing function of b , a given Gini coefficient will uniquely determine b and the income share function $s(z)$.

3. Derivation of the LBGC and the Corresponding IDF

Suppose we define a summary measure with the projection of a share function $s(z)$ with a weighting function $\psi(z)$, i.e.,

$$\phi = \int_0^1 \psi(z) s(z) dz \quad (13)$$

Suppose the logit function is defined as

$$\psi(z) = \text{logit}(z) = \log\left(\frac{z}{1-z}\right) \quad (14)$$

The projection of a share function with the above logit function is

$$\phi = \int \log\left(\frac{z}{1-z}\right) s(z) dz \quad (15)$$

For the uniform distribution with equal shares or for the extremely unequal shares where only one person has all the income, ϕ will have the values

$$B = \int_0^1 \log\left(\frac{z}{1-z}\right) dz \text{ or } \log(2n)$$

where n is the sample size. Now, define the inequality measure LBGC as

$$LBGC = \frac{\phi - B}{\log(2n) - B} \quad (16)$$

Then LBGC will be zero for the extremely equal or uniform distribution, and LBGC will be one for the extremely unequal distribution, where (say) all the mass is at one point in the tail.

The Lagrangian method of (8) subject to ϕ and normalization produces

$$L = \int s(z) \log s(z) dz + \lambda \left[\int s(z) dz - 1 \right] + \gamma \left[\int \log\left(\frac{z}{1-z}\right) s(z) dz - \phi \right]$$

Maximization with respect to $s(z)$ produces

$$\begin{aligned} \log s(z) + 1 + \lambda + \gamma \log\left(\frac{z}{1-z}\right) &= 0 \\ s(z) &= \exp\left[-\left(1 + \lambda + \gamma \log\left(\frac{z}{1-z}\right)\right)\right] \\ s_L(z) &= C(\gamma) \left(\frac{z}{1-z}\right)^\gamma \end{aligned} \quad (17)$$

For the chosen γ , $C(\gamma)$ is found with normalization. Choose γ such that the estimated logit moment of (15) will be the same as the observed logit moment.

$$\phi_\gamma = \int \log\left(\frac{z}{1-z}\right) C(\gamma) \left(\frac{z}{1-z}\right)^\gamma dz$$

Suppose percentile share data are given for an economy. Then one can summarize inequality with (15) and get ϕ^* . For various values of γ s, we can find ϕ_γ by numerically adding values at the midpoint of $z = [0, 0.01], \dots, [0.99, 1.0]$. Since ϕ_γ is an increasing function of γ , a unique γ^* is found corresponding to ϕ^* .

4. Applications

To demonstrate the usefulness of this new approach, we utilize Current Population Survey (CPS) data for the years 2000 through 2016. Because the Gini coefficient and the LBGC are both inequality measures that are in fact functions of underlying income shares, an underlying share function can be uniquely determined for each (as was shown in Equation (10) for the Gini and Equation (17) for the LBGC). These underlying share functions afford us the opportunity to directly compare how the Gini and LBGC measures perform in fitting actual income distribution share data. That is, we can compare apples to apples by seeing how various expected shares from the Gini and LBGC compare to actual shares of an actual empirical distribution, across various parts of the empirical distribution. Using CPS data divided into income centiles, we can examine directly how the two measures perform in fitting actual observed shares over the entire actual empirical distribution, divided into centiles (100th percentiles). Concomitantly, we can also perform the exercise by isolating various regions of the actual observed centile shares to focus on the poorest groups and richest groups. Recall that our hypothesis is that while the Gini is an excellent intuitive measure of inequality, it doesn't perform as well in capturing the level of inequality inherent in the tails because it loses information as it is "first moment focused"; since the empirical distribution tails may contain extreme values and the share function derived from the Gini coefficient may not detect that the tail has extremely high or low values, the Gini index will not accurately capture that fact. However, because the LBGC measure captures information over the entire distribution and the share functions underlying it take into account all values in the distribution including in the extreme tails, the share function derived from the LBGC is quite accurate. To see this, consider Figure 1.

We used Shazam version 11 for computer calculations and Eviews 10 for plotting the graphs.

In Figure 1, we present a mapping of the estimated Gini-coefficient-derived shares and the estimated logit-function-derived shares against the actually observed centile income shares for 2016, for the entire observed empirical distribution function. The black line represents the percent of actual observed income held by each centile share against the shares predicted by the Gini (red line) and logit (LBGC, the blue line). A perfect fit would mean the lines lie on top of each other as one moves from lower income centiles to successively higher ones. As can be seen from Figure 1, it is evident that the Gini estimates do not perform well at the very highest income centile shares while the LBGC function does a very good job of tracking with the highest centiles of actual income shares in 2016. While not as evident from the graph in Figure 1, the lower tails of the actual empirical distribution are also approximated better by the LBGC functional than by the Gini coefficient. Figure 2 illustrates that when the poorest 5% income group shares are examined, having been derived from both the Gini coefficient and the LBGC measure, and are then compared with the observed shares for the CPS data for the years 2000–2016, the LBGC-measure-derived shares more accurately approximate the actual centiles and also capture movement across the lowest centiles of income shares more accurately. As can be seen by viewing the blue line (LBGC) and the red line (Gini), it is evident that estimation with the LBGC-based shares is more precise in comparison to the Gini coefficient as the blue line is closer to the actual observed shares represented in the black line. In Figure 3, the richest 5% of income group shares are compared in the same way as in Figure 2. The richest 5% of income shares are estimated well using the LBGC-derived shares as the blue line lies almost on top of the black line. The Gini-derived shares fit relatively poorly. The purpose of these graphs is to show precisely these facts. In fact, the same holds true across all of the empirical distributions for all of the years for which we have data.

Figure 2. Poorest 5% shares estimated from Gini and LBGC.

Figure 3. Richest 95% shares estimated with Gini and LBGC.

To summarize in an intuitive manner how the shares derived from the two inequality models performed against each other for the period from 2000 to 2016, we examined graphically how observed income shares have changed from the CPS data over time and compared graphically the fit of those to the estimated shares from our underlying projections. Of course, it is possible to assess the relative fit of the shares derived from our inequality models more explicitly. To do so, we calculate the sum of squared residuals. We found that the sum of squared residuals (RSS) produced by various models produced very clear results. For example, recalling that the estimated share based on the observed Gini model is presented in Equation (10) and the LBGC-measure-derived share is presented in (17), we can dispense with the heuristic graphical presentations to show the actual fits for each model explicitly. The RSS is a standard fit criterion,

$$RSS_{Gini}(t) = [\text{Observed share}(t)_i - \text{Gini Estimated share}(t)_i]^2 \quad (18)$$

As Figure 4 illustrates, one can see that year 2013 was somewhat of an anomalous year with more income inequality than prior years or years that followed. The share of income of the richest 1% of earners was 29% in 2013. In comparison, the richest 1% share of income in 2015 was 23%. Figure 4 shows the LBGC model has an absolute advantage relative to the Gini model in that it yields a much lower RSS. It is frequently the case that a policymaker does not have access to the underlying data and only has access to or wants to focus on a single measure or indicator of inequality. In those instances, the logit variant of the Gini might be seen as most useful in gauging how overall inequality has changed over time.

Figure 4. Comparison of residual sum of squares.

5. Conclusions

This paper introduced a new income inequality measure as a projection of the share function with a particular weighting function. The choice of weighting function uniquely determines the summary measure. By choosing a logit function as a weighting function, a summary measure is determined which we labeled the LBGC measure. The underlying income distribution function (IDF) is estimated using the maximum entropy method subject to the given LBGC. The movements of extreme income share changes are described well with the estimated IDF. Movements of the vanishing middleclass income share are also detected well with the LBGC measure. The Gini coefficient moves in a parallel path with the LBGC and it gives good directional information on worsening inequality but it could not provide quantitative information on the rapidly increasing income share of the richest group. The particular shape of the logit function is sensitive to movements in income shares of the poorest and richest groups. This logit measure and the derived share function provide a more useful tool for policy makers who actually have to be concerned with how (say) tax policy will affect the poorest and richest segments of a society.

Author Contributions: H.K.R. and D.J.S.; methodology, H.K.R.; software, H.K.R.; validation, H.K.R., D.J.S. and H.Y.K.; formal analysis, H.K.R.; investigation, H.Y.K. and D.J.S.; resources, D.J.S.; data curation, H.K.R.; writing—original draft preparation, H.K.R.; writing—review and editing, H.K.R. and D.J.S.; visualization, H.K.R.; supervision, H.Y.K.; project administration, H.Y.K.; funding acquisition, H.K.

Funding: This research was supported by the National Research Foundation of Korea (NRF-2017S1A3A2066657).

Acknowledgments: We thank three referees and the editor for comments that greatly improved the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Desirable Properties of an Income Inequality Index

This list is a collection whose individual properties are discussed in many places, including Cowell [1], Sen [19], and Ryu and Slottje [17], among others. They are enumerated in an excellent appendix by Sen and Foster, contained in Sen [19].

Given the inherent difficulty in describing the characteristics of an entire IDF with one number, the following properties for an income inequality index are *desirable*. Professor Sen's seminal book in its third edition in 2003 added an appendix with Professor James Foster that rigorously laid out these axioms:

Anonymity or symmetry

The inequality measure should not depend on how individuals in an observed distribution are labeled (cf. Sen and Foster [19], Appendix A3.2).

Scale independence or homogeneity

As Cowell ([1], p. 63) notes, the measured inequality of the slices of the cake should not depend on the size of the cake. This property says that if (say) every person's income in an economy is increased by some constant, then the overall metric of inequality should not change.

Population independence

Similarly, the inequality measure should be independent of the level of population. Cowell ([1], p. 63) notes the inequality of the cake distribution should not depend on the number of cake receivers.

Transfer principle

The Pigou–Dalton, or transfer principle, states, in its weak form, that if income is transferred from a rich person to a poor person while still preserving the order of income ranks, then the inequality measurement should not increase (cf. Cowell [1], p. 62).

Non-negativity

The inequality index $I(y)$ must be greater than or equal to zero (cf. Ryu and Slottje [17] and Sen and Foster [19], Appendix A3.3).

Egalitarian zero

The index $I(y)$ is zero when everyone has the same income, meaning when all values y_i are equal (cf. Ryu and Slottje [17] and Sen and Foster [19], Appendix A3.2).

Bounded above by maximum inequality

The index $I(y)$ attains its maximum value of one, reflecting the maximum level of inequality (all y_i are zero except one) (cf. Ryu and Slottje [17] and Sen and Foster [19], Appendix A3.1).

In this paper, we utilize the Gini coefficient and introduce a new inequality measure based on the logit type function. Since we use only 100 percentile income shares, anonymity, scale independence, and population independence are satisfied for the summary measure used in this paper. Non-negativity, egalitarian zero, and “bounded by maximum inequality” are also satisfied for the summary measure (LBGC) used in this paper. The remaining property is the transfer principle. The projection function, the logit function, is a monotonic increasing function. Therefore, if a small portion of income is transferred from a rich person to a poorer person, then the projection with a monotonic increasing function will decrease and inequality will decrease. Thus, the transfer principle is satisfied.

Appendix B. Various Weighting Functions for Inequality Measures

In this paper, we introduced a projection of a share function with a weighting function of a logit function. This functional form was one of several moment-based weighting function candidates considered. Suppose we label the inequality measure derived from the second moment as G_2 , the inequality measure derived from the third moment as G_3 , the inequality measure derived from the square root moment as G_{sqrt} , and the inequality measure derived from the exponential moment (μ_{exp}) as G_{exp} . The exact meaning of these derived Ginis will be explicated below. However, at this

juncture, we simply want to note that each of these “derived Ginis” are derived from the underlying share function and subsequently, we can derive a share function (s_2) using the maximum entropy method from the second moment (μ_2), derive a share function (s_3) from the third moment (μ_3), derive a share function ($s_{\text{sqr}}t$) from the square root moment ($\mu_{\text{sqr}}t$), and derive a share function (s_{exp}) from the given exponent moment μ_{exp} .

Once these derived share functions are determined, these approximated shares will be compared to the observed shares used to calculate the original Gini. Finally, using US CPS data for the period from 2000 through 2016, we can compare the performance of the Gini, G_2 , G_3 , $G_{\text{sqr}}t$, G_{exp} , and LBGC of (16) with respect to movements of the income quantiles over time by examining changes in the observed IDF by honing in on the poorest 5% shares, Q1, Q2, Q3, Q4, Q5, and the richest 5% income shares of the observed shares. This exercise will allow us to glean more information about how movement within the observed IDF (and its attendant summary Gini coefficient) has fluctuated over time. We will also examine our proffered measures with respect to desirable properties that an income inequality index should possess, such as symmetry, scale independence, population independence, transfer principle, non-negativity, egalitarian zero, and bounded above by maximum inequality. The examination will be performed for G_2 , G_3 , $G_{\text{sqr}}t$, G_{exp} , and LBGC and the proofs that these measures satisfy desirable properties that any inequality measure should possess can be explained with the same logic listed in Appendix A above.

Appendix B.1. Derivation of an Inequality Measure and Corresponding Share Function from the Given Second Moment

If only the second moment is utilized,

$$\mu_2 = \int z^2 s(z) dz \approx \int z^2 s_N(z) dz = \int z^2 \sum_{n=1}^N a_n P_n(z) dz \quad (\text{A1})$$

Since $\mu_2 = 1/3$ for the uniform shares and $\mu_2 = 1$ for extreme inequality, define

$$G_2 = \frac{3\mu_2 - 1}{2} \quad (\text{A2})$$

To derive the share function from the second moment, use

$$z^2 = \frac{P_2(z)}{6\sqrt{5}} + \frac{\sqrt{3}P_1(z)}{6} + \frac{P_0(z)}{3}$$

$$\mu_2 = \int \left[\frac{P_2(z)}{6\sqrt{5}} + \frac{\sqrt{3}P_1(z)}{6} + \frac{P_0(z)}{3} \right] \sum_{n=1}^N a_n P_n(z) dz = \left[\frac{a_2}{6\sqrt{5}} + \frac{a_1}{6} + \frac{a_0}{3} \right] \quad (\text{A3})$$

The Lagrangian method subject to the second moment produces

$$s_2(z) = C \exp(dz^2) \quad (\text{A4})$$

For the d selected, C is found through normalization. Choose d such that the estimated second moment is the same as the observed second moment.

Appendix B.2. Derivation of an Inequality Measure and Corresponding Share Function from the Given Third Moment

If only the third moment is utilized,

$$\mu_3 = \int z^3 s(z) dz \approx \int z^3 s_N(z) dz = \int z^3 \sum_{n=1}^N a_n P_n(z) dz \quad (\text{A5})$$

Since $\mu_3 = 1/4$ for the uniform shares and $\mu_3 = 1$ for extreme inequality, define

$$G_3 = \frac{4\mu_3 - 1}{3} \quad (\text{A6})$$

To derive the share function from the third moment, use

$$\begin{aligned} z^3 &= \frac{P_3(z)}{20\sqrt{7}} + \frac{\sqrt{5}P_2(z)}{20} + \frac{3\sqrt{3}P_1(z)}{20} + \frac{P_0(z)}{4} \\ \mu_3 &= \int \left[\frac{P_3(z)}{20\sqrt{7}} + \frac{\sqrt{5}P_2(z)}{20} + \frac{3\sqrt{3}P_1(z)}{20} + \frac{P_0(z)}{4} \right] \sum_{n=1}^N a_n P_n(z) dz \\ &= \left[\frac{a_3}{20\sqrt{7}} + \frac{\sqrt{5}a_2}{20} + \frac{3\sqrt{3}a_1}{20} + \frac{a_0}{4} \right] \end{aligned} \quad (\text{A7})$$

The Lagrangian method subject to the third moment produces

$$s_3(z) = F \exp(gz^3) \quad (\text{A8})$$

For the chosen g , F is found with normalization. Choose g such that the estimated third moment is the same as the observed third moment.

Appendix B.3. Derivation of an Inequality Measure and Corresponding Share Function from the Given Square Root Moment

If only the square root moment is utilized,

$$\mu_{0.5} = \int \sqrt{z}s(z)dz \approx \int \sqrt{z}s_N(z)dz = \int \sqrt{z} \sum_{n=1}^N a_n P_n(z) dz \quad (\text{A9})$$

Since $\mu_{0.5} = 2/3$ for the uniform shares and $\mu_{0.5} = 1$ for extreme inequality, define

$$G_{sqr} = 3\mu_{0.5} - 2 \quad (\text{A10})$$

To derive the share function, use the Lagrangian method subject to the square root moment, which produces

$$s_{sqr}(z) = N \exp(m\sqrt{z}) \quad (\text{A11})$$

For the chosen m , N is found with normalization. Choose m such that the estimated square root moment is the same as the observed square root moment.

Appendix B.4. Derivation of an Inequality Measure and Corresponding share Function from the Given Exponent MOMENT

If only the exponent moment is utilized,

$$\psi = \int e^z s(z) dz \approx \int e^z s_N(z) dz = \int e^z \sum_{n=1}^N a_n P_n(z) dz \quad (\text{A12})$$

Since $\psi = e - 1$ for the uniform shares and $\psi = e$ for extreme inequality, define

$$G_{\exp} = \psi + 1 - e \quad (\text{A13})$$

The Lagrangian method subject to the exponent moment produces

$$s_{\exp}(z) = H \exp[\exp(kz)] \quad (\text{A14})$$

For the chosen k , H is found with normalization. Choose k such that the estimated exponent moment is the same as the observed exponent moment.

Figure A1. Estimated shares from various moments.

Figure A2. Comparison of log shares from various moments.

Figure A3. Comparison of various methods for bottom 5% share.

Figure A4. Comparison of various methods for Q1 shares.

Figure A5. Comparison of various methods for Q2.

Figure A6. Comparison of various methods for Q3.

Figure A7. Comparison of various methods for Q4.

Figure A8. Comparison of various methods for Q5.

Figure A9. Comparison of various methods for top 5% share.

Figure A10. Comparison of Residual Sum of Squares.

Figure A1 compares estimated shares from various candidates. Only the LBGC measure indicates a good approximation while the other candidate measures seemingly fit less well. Figure A2 shows another comparison. The LBGC distribution approximates the very poor and rich groups well. Figures A3–A9 compare estimated shares for the bottom 5% share, 20% share (Q1), next 20% share (Q2), next 20% share (Q3), next 20% share (Q4), top 20% share (Q5), and the top 5% group shares. The LBGC method indicates a robust performance for the bottom 5% of income earners, Q1, Q2, and for the top 5% share, but for Q3, Q4, and Q5, the other moment-based measures indicate a better performance. Interestingly, the performance of the actual Gini itself was relatively weak. In Figure A6, the square root model and the actual Gini model did a bit better than the logit model across the third quintile,

the Q3 region. The other models such as the exponent model, the second moment model, and the third moment model showed relatively bad performances in this portion of the IDF. In Figure A7, the second moment model seems to be doing well in the fourth quintile, Q4 region. The logit is the next best performer while the square root model and the Gini model indicate relatively poorer performances. In Figure A8, the Gini model, the exponent moment model, and the second moment model performed better than the logit model for the upper portions of the IDF. In Figure A9, the logit model showed a strong performance that was far better than the other Gini variants. In particular, the Gini model and the square root model performed worse than the other models over the fourth quintile. In summary, the logit model demonstrated the best performance for the bottom 5% share of income earners and for the Q1, Q2, and top 5% share income shares. For the other regions of the IDF, the logit model performed as the second or the third highest ranking measure. In other regions, the square root model was strong for Q3, the second moment model was strong for Q4, and the third moment model was the best for Q5.

To summarize the overall fitting performance of these models for the period from 2000 to 2016, we examined how observed income shares have changed from the CPS data over time and compared those observed shares to the estimated shares from our underlying projections. We once again found that the sum of squared residuals produced by various models produced very clear results. For example, the estimated share based on the observed Gini model was presented in Equation (10).

$$RSS_{Gini}(t) = [\text{Observed share}(t)_i - \text{Gini Estimated share}(t)_i]^2$$

Similarly, the estimated shares using the second moment, third moment, square root moment, exponent moment, and logit type moment are presented in (A4), (A8), (A11), (A14) and (17). Year 2013 was somewhat of an anomalous year with more income inequality than prior years or years that followed. The share of income of the richest 1% of earners was 29% in 2013. In comparison, the richest 1% share of 2015 was 23%. Figure A10 shows the logit type moment has an absolute advantage relative to the other moment projections over the entire IDF. By focusing only on results produced by the Gini coefficient, we are throwing out other information contained in the nonlinear part of the underlying IDF, as we utilized only the first moment projection of the share function. In comparison, the logit type projection removed a significant portion of the estimation error.

References

1. Cowell, F. *Measuring Inequality*, 3rd ed.; Oxford University Press: Oxford, UK, 2011.
2. Ryu, H. Maximum Entropy Estimation of Density and Regression Functions. *J. Econom.* **1993**, *56*, 397–440. [\[CrossRef\]](#)
3. Tanak, A.K.; Borzadaran, G.R.M.; Ahmadi, J. New Functional Forms of Lorenz Curves by Maximizing Tsallis Entropy of Income Share Function under the Constraint on Generalized Index. *Physica A* **2018**, *511*, 280–288. [\[CrossRef\]](#)
4. Rad, N.; Borzadaran, G.R.M.; Yari, G.H. Maximum Entropy Estimation of Income Share Function from Generalized Gini Index. *J. Appl. Stat.* **2016**, *43*, 2910–2921.
5. Arnold, B.; Sarabia, J.M. *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics*; Springer: New York, NY, USA, 2018.
6. Arnold, B.; Castillo, E.; Sarabia, J.M. *Conditional Specification of Statistical Models*; Springer: New York, NY, USA, 1999.
7. Giorgi, G.M. A Fresh Look at the Topical Interest of the Gini Concentration Ratio. *Econometrics* **1993**, *51*, 83–98.
8. Giorgi, G.M.; Gubbiotti, S. Celebrating the memory of Corrado Gini: A personality out of the ordinary. *Int. Stat. Rev.* **2017**, *85*, 325–339. [\[CrossRef\]](#)
9. Nygard, F.; Sandstrom, A. *Measuring Inequality*; Almqvist and Wiksell: Stockholm, Sweden, 1981.
10. Nygard, F.; Sandstrom, A. Income Inequality Measures based on Sample Surveys. *J. Econom.* **1989**, *42*, 81–96. [\[CrossRef\]](#)

11. Lerman, R.; Yitzhaki, S. A Note on the Calculation and Interpretation of the Gini Index. *Econom. Lett.* **1984**, *15*, 363–368. [[CrossRef](#)]
12. Yitzhaki, S. More than a dozen ways of spelling Gini. *Res. Income Inequal.* **1998**, *8*, 13–33.
13. Gastwirth, J. Is the Gini Index Overly Sensitive to changes in the Middle of the Income distribution? *Stat. Public Policy* **2017**, *4*, 1–11. [[CrossRef](#)]
14. Arfken, G. *Mathematical Methods for Physicists*, 3rd ed.; Academic Press, Inc.: San Diego, CA, USA, 1985.
15. Milne, W. *Numerical Calculations*; Princeton University Press: Princeton, NJ, USA, 1973.
16. Ryu, H.; Slottje, D. Income Inequality vs Utility Inequality. *Commun. Stat.* **2017**, *46*, 3631–3640.
17. Ryu, H.; Slottje, D. Maximum Entropy Estimation of Income Distributions from Basmann’s WGM Class. *J. Econ.* **2017**, *199*, 221–231. [[CrossRef](#)]
18. Shannon, C. The Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, reprinted in Shannon, C.E.; Weaver, W.; Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949; pp. 3–91. [[CrossRef](#)]
19. Sen, A. *On Economic Inequality*, 3rd ed.; Oxford University Press: Oxford, UK, 2003.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).