A New Inequality Measure that is Sensitive to Extreme Values and Asymmetries[1]

Michael McAleer[2], Hang K. Ryu[3] and Daniel J. Slottje[4]

EI2017-28

Abstract

There is a vast literature on the selection of an appropriate index of income inequality and on what desirable properties such a measure (or index) should contain. The Gini index is, of course, the most popular. There is a concurrent literature on the use of hypothetical statistical distributions to approximate and describe an observed distribution of incomes. Pareto and others observed early on that incomes tend to be heavily right-tailed in their distribution. These asymmetries led to approximating the observed income distributions with extreme value hypothetical statistical distributions, such as the Pareto distribution. But these income distribution functions (IDFs) continue to be described with a single index (such as the Gini) that poorly detect the extreme values present in the underlying empirical IDF. This paper introduces a new inequality measure to supplement, but not to replace, the Gini that measures more accurately the inherent asymmetries and extreme values that are present in observed income distributions. The new measure is based on a third-order term of a Legendre polynomial from the logarithm of a share function (or Lorenz curve). We advocate using the two measures together to provide a better description of inequality inherent in empirical income distributions with extreme values.

JEL Classification: D31, D63

Keywords: Inequality Index, Extreme value distributions, Maximum entropy method, Orthonormal basis, Legendre polynomials.

---

[1] This research was supported by the National Research Foundation of Korea (2017S1A3A2066657), National Science Council, Ministry of Science and Technology (MOST), Taiwan, and the Australian Research Council.

[2] Department of Quantitative Finance, National Tsing Hua University, Taiwan; Discipline of Business Analytics, University of Sydney Business School, Australia; Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands; Department of Quantitative Economics, Complutense University of Madrid, Spain; Institute of Advanced Sciences, Yokohama National University, Japan.
Email: michael.mcaleer@gmail.com

[3] Department of Economics, Chung Ang University, Seoul, Korea, 156-756, Tel.: +82-11-253-6500;
Email: hangryu@cau.ac.kr

[4] Department of Economics, SMU, Dallas, TX 75275, Tel: 214-732-9170,
Email: dan.slottje@fticonsulting.com

I. Introduction

Income inequality research has experienced a resurgence after losing some momentum in the late 1990s and the first decade of the Twenty-first Century. Piketty (1995, 2014) and Boushey et al. (2017) reignited some interest in the field; Piketty did so with his 2014 tome on "polarization." There is a vast literature on the measurement of income inequality, cf. Cowell (2011) for an excellent bibliography of much of this work. This literature contains hundreds of papers on an appropriate index of income inequality and on what desirable properties such a measure (or index) should possess. We present and review some of this discussion below.

There is also a concurrent literature on the use of hypothetical statistical distributions to approximate and describe an observed distribution of incomes. Pareto (1896) and others observed early on that incomes tend to be heavily right-tailed in their distribution. These asymmetries led researchers to approximating the observed income distributions with extreme value hypothetical statistical distributions, such as the Pareto distribution. Statisticians have done considerable work on extreme value distributions in other applications. The generalized extreme value distribution (GEV) and its family members, including the Weibull, Gumbel, Frechet and others, have been extensively explored by statisticians and inequality researchers alike (cf. Coles (2001) and Cowell and Flachaire (2007)). James McDonald has been a leading researcher in the area of functional forms of hypothetical statistical distributions to describe IDFs for a long time (cf. McDonald (1984), McDonald et al. (2013) and Slottje (1987)).

Interestingly, even with the recognition of the fact that incomes are distributed with asymmetric higher moments, inequality indices constructed to capture the level of inequality inherent in these observed income distributions (with a single number) are generally based on the mean and variance of the

observed data. Cowell and Flachaire (2002, 2007) is the only work that seems to discuss the two concepts (that is, extreme values in the IDF and detecting it with an inequality index) in the same place. They do not introduce a new index or measure to deal with the issue, but note that the two most popular classes of measures, the Gini and Entropy-based measures, have different sensitivities to the problem in their first paper (cf. Cowell and Flachaire (2002)).

In their second paper, the authors are primarily concerned about how sensitive commonly used inequality measures are to extreme values in the underlying distributions, and suggest some semi-parametric specifications of the commonly used measures to account for the extreme values (cf. Cowell and Flachaire (2007)). The Gini coefficient and Theil's entropy measure (frequently generalized) are two very popular inequality indices, among others, that have not always performed well in describing some of the tail behavior in observed income distributions. Specifically, both measures fall short in detecting changes in various group's share (cf. Ryu(2013) and Ryu and Slottje (2017))[5].

Another way to approach the problem is to realize that there are many income distribution functions which will produce the same value of a Gini coefficient. The overall shape of the income share function may be well described by the Gini coefficient (or by Theil's entropy measure), but the poorest group's share and the precise details of the richest group's share generally are not described well by these measures. In this paper, a second inequality measure is introduced and added to the Gini coefficient to describe movements of the extreme values and asymmetries of observed income distributions as they change over time.

[5] See Maasoumi (1986, 1989) for excellent work on the generalized entropy class of measures.

In the next section we discuss desirable properties an inequality measure should possess. In Section 3 and 4 we introduce the new measure, which is based on the expansion of the logarithm of the share function (or Lorenz curve) with a Legendre polynomial expansion. Section 5 of the paper discusses an application by fitting the new measure to CPS data. Section 6 concludes the paper.

II. Desirable Properties of an Income Inequality Index, I(y)[6]

There is significant consensus among inequality researchers that any income inequality index, I(y), should possess statistical properties that allow it to reasonably describe the inequality inherent in an observed IDF. Given the inherent difficulty in describing the characteristics of an entire IDF with one number, the following properties are desirable:

- **Anonymity or symmetry**

  The inequality measure should not depend on how individuals in an observed distribution are labeled. Another words, it doesn't matter who receives the income, all that matters is the distribution of income. This is generally expressed mathematically as:

$$I(P(y)) = I(y) \tag{1}$$

  where *P(y)* is any permutation of income y;

[6] This list is a collection whose individual properties are discussed in many places, including Cowell (2011), Ryu and Slottje (1998), Basmann and Slottje (1987), and Basmann, Hayes and Slottje (1991), among others.

- **Scale independence or homogeneity**

  As Cowell (2011, 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. This may be stated as:

$$I(ay) = I(y) \tag{2}$$

  where a is a positive real number.

- **Population independence**

  Similarly, the inequality measure should be independent of the level of population. Cowell (2011, p. 63) notes the inequality of the cake distribution should not depend on the number of cake-receivers. This is generally written as:

$$I(y \cup y) = I(y) \tag{3}$$

  where $\cup$ is the union of x with itself.

- **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. In its strong form, the transfer principle says the measured level of inequality should decrease. As will be shown below in

our paper, our new second measure satisfies this condition if it is considered together with the Gini coefficient (see the Appendix for proof).

- **Non-negativity**

  The inequality index *I(y)* must be greater than or equal to zero.

- **Egalitarian zero**

  The index *I(y)* is zero when everyone has the same income, meaning when all values y_i are equal.

- **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).

In the discussion to follow, we introduce a new measure that will be shown to satisfy these properties.

III. New Measure of Inequality that Supplements the Gini Coefficient

Given our objective to find a new income inequality measure which is sensitive to extreme values, we propose to describe the income distribution with two summary measures rather than a single measure. The Gini coefficient, Theil's entropy measure, and other well-known measures are useful in describing the overall state of income inequality, but these measures do not provide precise information about the presence of extreme values in an underlying IDF, or in how change in the extreme values over time impact the level of inequality as reflected in the summary index over time.

In this paper, we conceptualize a complete set of distributions all having the same Gini value. A function derived using only the Gini coefficient will be called the basic model in the paper. This basic model is known to be imprecise in describing the presence of extreme values. A second inequality measure will supplement the Gini, and is designed to describe the movements of the poorest group's income share and the extreme values of the richest income group.

The choice of the second inequality measure is extremely important. The basic model can be derived using the first inequality measure, such as the Gini coefficient, Theil's entropy measure, and others. The basic model used in this paper is the Gini coefficient-based model. When the second inequality measure is added, it is desirable to derive the functional form corresponding to this second measure and to add this part to the basic model. In the applications section, the income distribution of the basic model and the distribution of the extended model will be compared.

To introduce the second inequality measure, two functional forms are considered in this paper. The first functional form is the expansion of the logarithm of the share function in terms of the Legendre polynomial series. The second functional form is the expansion of the Lorenz curve in terms of the Legendre polynomial series. For the first functional form, the parameter of the first order polynomial term can be derived from the Gini coefficient, and the parameter of the third order polynomial term will be used as the second inequality measure. Note that the second-order term of the Legendre polynomial series is a symmetric function, so that it cannot be used in describing the monotonic increasing function. Both forms will be explained below.

For the second functional form where the Lorenz curve is expanded in Legendre polynomials, the parameter of the zero-th Legendre polynomial term corresponds to the Gini coefficient, and the parameter of the first Legendre polynomial term can be used as the second inequality measure.

3.1 Orthonormal basis expansion of the logarithm of income share function

For the given income observations, there are many ways to approximate the functional form of the data generating model. If an orthonormal basis (ONB) expansion is applied, the parameter calculation is unaffected by the size of the series. In comparison, the estimated parameters of the ordinary least squares regression method change their values when a new term is added in the regression series.

The addition of higher-order terms in the series will allow the approximated function to converge to the data generating model. These functions with different series lengths form a complete set of income distributions corresponding to the basic model derived from the Gini coefficient. Orthonormal basis expansion allows us to superpose new terms on the basic model without disturbing the basic model.

Suppose we have a continuous share function $s(z)$ for $0 \le z \le 1$, where the poorest person is located at $z = 0$ and the richest at $z = 1$. We can approximate the logarithm of the share function with a sequence of orthonormal functions, $P_0(z), P_1(z), P_2(z), P_3(z), \ldots$. Arfken (1985) presents an explanation of the ONB method:

$$log\ s_N(z) = \sum_{n=1}^{N} a_n P_n(z) \tag{4}$$

An orthonormal sequence satisfies:

$$\int_Z P_n(z)\,P_m(z)\,dz = \delta_{nm}, \qquad n,m,=0,1,2,\cdots \tag{5}$$

where $\delta_{nm}=1$ if $n=m$ and zero otherwise. The parameters of (4) can be found with:

$$a_m = \int P_m(z)\ log\ s_N(z)\,dz = \int P_m(z)\left[\sum_{n=1}^{N} a_n P_n(z)\right] dz \tag{6}$$

(see Ryu (1993) for the continuous version of ONB, and Ryu and Slottje (1996) and Milne (1949) for a discussion of the discrete version of ONB). The orthogonal sequence $\{P_n\}$ in the space $L^2(Z)$ is called complete if there is no element $f\neq 0$ of $L^2(Z)$ which is orthogonal to all the elements of P_n. If:

$$\int_Z f(z)\,P_n(z)\ dz = 0 \qquad \text{for} \qquad n=0,1,2,\cdots \tag{7}$$

it follows $f(z)=0$ for almost all $z\in Z$.

Suppose the Legendre polynomials are used for $0\le z\le 1$:

$$
\begin{aligned}
P_0(z) &= 1 \\
P_1(z) &= \sqrt{3}\,(2z-1) \\
P_2(z) &= \sqrt{5}\left(6z^2-6z+1\right) \\
P_3(z) &= \sqrt{7}\left(20z^3-30z^2+12z-1\right) \\
P_4(z) &= \sqrt{9}\left(70z^4-140z^3+90z^2-20z+1\right) \\
P_5(z) &= \sqrt{11}\left(252z^5-630z^4+560z^3-210z^2+30z-1\right)
\end{aligned}
\qquad (8)
$$

Fig.1 shows $P_0(z)$ is flat and $P_1(z)$ is a linear function but $P_n(z)$ has $n-1$ peak values. To approximate the logarithm of the share function, the Legendre polynomials with degrees of even numbers seem to be less useful because they have peak values at $z=0$. Those functions with degrees of odd numbers will be useful as they have their lowest values at $z=0$ and their largest values at $z=1$.

Consider the following basic model, which can be derived from the given Gini coefficient:

$$
log\ s_{Gini}(z) = a_0 + a_1 P_1(z) \qquad \text{or} \qquad s_{Gini}(z) = \exp[a_0 + a_1 P_1(z)] \qquad (9)
$$

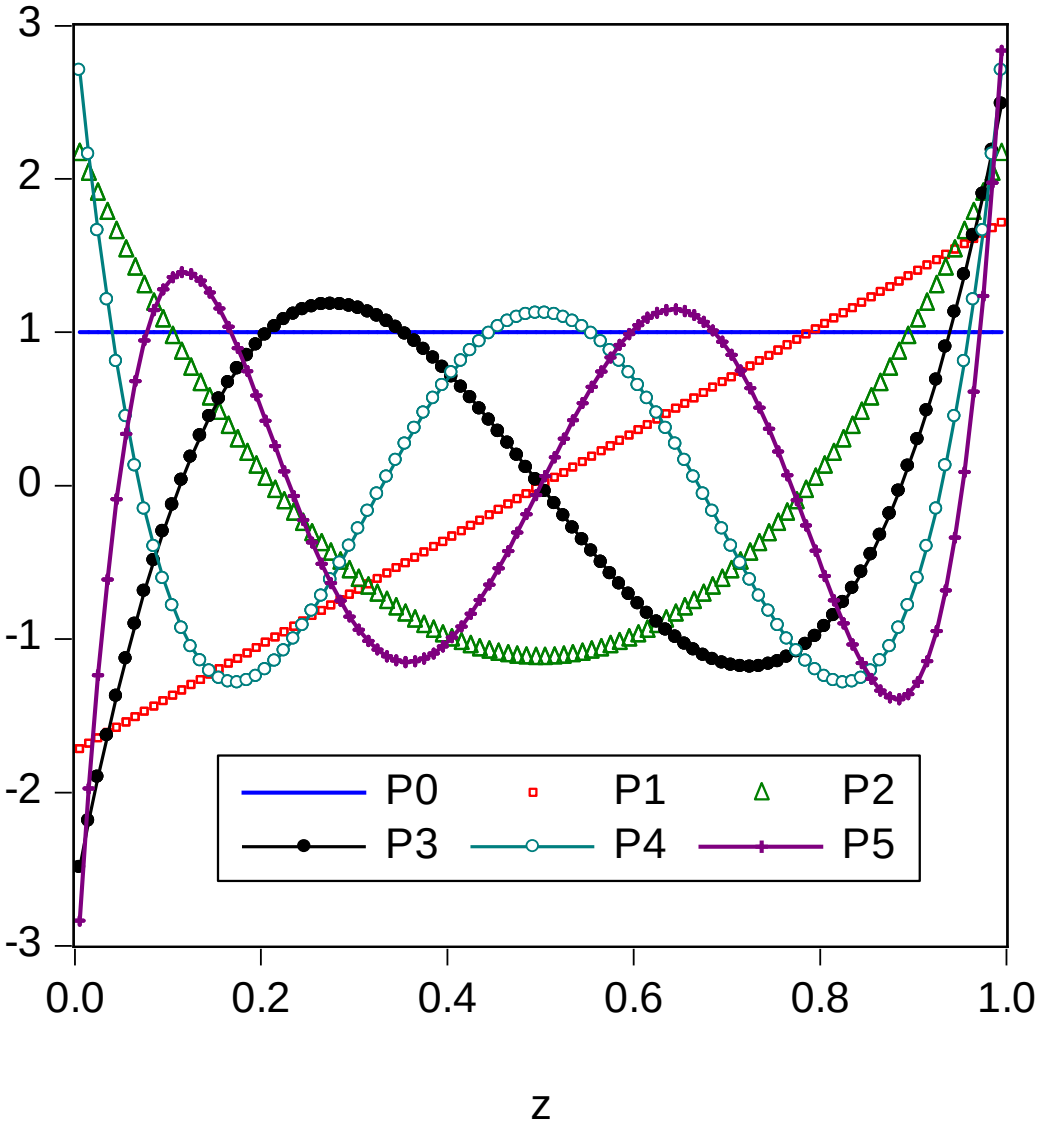
Yitzhaki (2013) has shown that knowledge of the Gini coefficient is equivalent to knowledge of the first moment of the share function. To find the parameters of (9) from the Gini coefficient, consider:

$$
a_0 + a_1 P_1(z) = a_0 + a_1\sqrt{3}\,(2z-1) = a_0 - \sqrt{3}\,a_1 + 2\sqrt{3}a_1 z = A + Bz \qquad (10)
$$

$$\mu_1 = \int z\, s(z)\, \mathrm{dz} = \int z \exp[\mathrm{A} + Bz]\, dz$$
$$= \left[\frac{B}{e^B - 1}\right] \int z \exp[\mathrm{B}\, z]\, dz = \frac{1 + \mathrm{Gini}}{2} \qquad (11)$$

where the parameter A is removed with normalization of the share function. Knowledge of the Gini allows us to find B, a_0 and a_1 of (10). Therefore, the basic model is derived from the given Gini coefficient.

Fig.1 Plots of Legendre Polynomials



To consider the extreme values at the fat right tail of the share function, the following extended functional forms can be applied:

Basic model: $$log\ s_{Gini}(z) = a_0 + a_1 P_1(z) \quad (12)$$

Second order: $$log\ s_2(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) \quad (13)$$

Third order: $$log\ s_3(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) + a_3 P_3(z) \quad (14)$$

Fourth order: $$log\ s_4(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) + a_3 P_3(z) + a_4 P_4(z) \quad (15)$$

Fifth order: $$log\ s_5(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) + a_3 P_3(z) + a_4 P_4(z) + a_5 P_5(z) \quad (16)$$

The parameters can be found with:

$$a_m = \int P_m(z)\ log\ s_N(z)\, dz \quad (17)$$

The parameter values calculated by (17) do not depend on the length of the series. For example, the a_2 parameters of (13), (14), (15), and (16) are the same. This is the benefit of the orthonormal function expansion. In comparison, the parameters estimated using a least squares method will fluctuate when we increase the length of series. Therefore, we can superpose another function derived with the additional parameter to the basic Gini model without damaging the basic model.

We have assumed knowledge of a continuous function $s(z)$ and expanded the logarithmic transformation with an orthonormal basis (4), so that

the parameters were found with (6) using the orthogonality of the Legendre functions. As an alternative method, suppose we do not know the functional form of the underlying share function $s(z)$. If nothing is known, the share function can be assumed to be a flat function. Suppose the moments of the share function are known, as follows:

$$\mu_m = \int z^m \; s(z)\,dz \quad \text{for} \quad m = 0,1,2,...,N \tag{18}$$

Then the following moments can be calculated based on (8):

$$\lambda_m = \int P_m(z) \; s(z)\,dz \quad \text{for} \quad m = 0,1,2,...,N \tag{19}$$

Zellner and Highfield (1988) and Ryu (1993) solved an entropy maximization problem:

$$Max_s W = -\int s(z) \log \, s(z)\,dz \tag{20}$$

satisfying:

$$\lambda_m = \int P_m(z) \; s(z)\,dz \quad \text{for} \quad m = 0,1,2,...,N \tag{19}$$

Then:

$$s(z)=\exp\left[\sum_{n=0}^{N} c_n P_n(z)\right] \text{ satisfying } \lambda_m=\int P_m(z)\ s(z)\,dz \text{ for } m=0,1,2,...,N \qquad (21)$$

If the Gini coefficient is known, this is equivalent to knowledge of λ_0 and λ_1, and so we have:

$$s(z)=\exp\left[c_0 P_0(z)+c_1 P_1(z)\right] \qquad (22)$$

which is equivalent to (12). The parameters of (22) can be determined from the given Gini coefficient, as derived in Ryu and Slottje (2017b). Two alternative methods to approximate the share function are now explained. The first method assumes knowledge of the continuous $s(z)$, which is expanded with a Legendre series. The second method does not assume the functional form of $s(z)$ but maximizes entropy subject to known values of moments. The derived functional forms are the same, but the parameter calculation methods are different.

As we add more terms to the series, the approximated function approaches $\log s_N(z)$:

$$\int\left[log\ s_N(z)\right]^2 dz=\int\left[\sum_{n=1}^{N} a_n P_n(z)\right]^2 dz=a_0^2+a_1^2+a_2^2+\cdots+a_N^2 \qquad (23)$$

Using 2016 CPS data (which will be discussed below in detail), we have:

$$a_0^2 = 27.921, \quad a_1^2 = 1.190, \quad a_2^2 = 0.0376, \quad a_3^2 = 0.1340, \quad a_4^2 = 0.0146, \quad a_5^2 = 0.0740 \quad (24)$$

where a_0 is used for normalization and a_1 is the slope term corresponding to the Gini coefficient. If we have to choose a term in addition to the basic model, then we can choose a term with the largest parameter squared value. In our case, a_3^2 has the largest value among the remaining terms.

Now suppose we wish to introduce a second inequality measure as a supplement to the Gini coefficient. There are a few choices suitable for this purpose. Consider the following:

Typical model: $$log\ sh_N(z) = a_0 + a_1 P_1(z) + a_N P_N(z) \quad (25)$$

Basic model: $$log\ s_{Gini}(z) = a_0 + a_1 P_1(z) \quad (12)$$

Second order model: $$log\ s_2(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) \quad (13)$$

Third order model: $$log\ sh_3(z) = a_0 + a_1 P_1(z) + a_3 P_3(z) \quad (26)$$

Fourth order model: $$log\ sh_4(z) = a_0 + a_1 P_1(z) + a_4 P_4(z) \quad (27)$$

Fifth order model: $$log\ sh_5(z) = a_0 + a_1 P_1(z) + a_5 P_5(z) \quad (28)$$

An approximated share function with the additional third-order term will be a monotonic increasing function if its slope is nonnegative for the given values of positive a_1 and a_3:

$$\frac{\partial log\ sh_3(z)}{\partial z} = \frac{\partial a_0 + a_1 P_1(z) + a_3 P_3(z)}{\partial z} = 2\sqrt{3}\ a_1 + \sqrt{7}\, a_3 (60z^2 - 60z + 12) > 0 \qquad (29)$$

If a monotonicity test is passed for (26), then the third-order parameter a_3 can be used as the second inequality measure. A similar monotonicity test can be performed for (28):

$$\frac{\partial log\ sh_5(z)}{\partial z} = \frac{\partial a_0 + a_1 P_1(z) + a_5 P_5(z)}{\partial z} > 0 \qquad (30)$$

IV. Lorenz dominance and expansion of the basic model

Another way to understand the intuition behind our new measure is to think about it in terms of Lorenz dominance. There are many Lorenz curves which can generate the same Gini coefficient. If we expand the Lorenz curve with a Legendre polynomial series, the zero-th order parameter can be determined from the Gini coefficient. The basic model will be the second-order Legendre polynomial series with three parameters, which can be determined from two boundary conditions, $L(z=0)=0$ and $L(z=1)=1$, and the Gini coefficient. Inclusion of higher-order Legendre functions will modify the basic Lorenz curve, but all these Lorenz functions will have the same Gini coefficient due to the orthogonality of the Legendre series. A related discussion can be found in Choo and Ryu (1994).

Suppose the Lorenz curve can be expanded through Legendre functions:

$$L_N(z) = \sum_{n=1}^{N} b_n P_n(z) \qquad (31)$$

The parameters can be found from the following relation:

$$b_m = \int P_m(z)\ L_N(z)\,dz = \int P_m(z)\left[\sum_{n=1}^{N} b_n P_n(z)\right]dz \qquad (32)$$

The Gini coefficient determines the zero-th order parameter:

$$\frac{1-\text{Gini}}{2} = \int_0^1 L(z)dz \;\square\; \int_0^1 L_N(z)dz = b_0 \qquad (33)$$

Notice the above relation does not depend on the size of the series N and all $L_N(z)$ will share the same Gini coefficient. The Lorenz curve should satisfy two boundary conditions:

$$L_N(z=0) = 0 \qquad \text{and} \qquad L_N(z=1) = 1 \qquad (34)$$

Now using:

$$P_n(\text{z}=0) = (-1)^n\sqrt{2n+1} \qquad \text{and} \qquad P_n(\text{z}=1) = \sqrt{2n+1} \qquad (35)$$

the second-order polynomial series, which we label as the basic model, is given

as follows:

$$L_2(z) = b_0 P_0(z) + b_1 P_1(z) + b_2 P_2(z) \tag{36}$$

Suppose the Gini coefficient is known, that is, b_0 is known. Using the boundary conditions, $L_2(z=0)=0$ and $L_2(z=1)=1$, the parameters b_1 and b_2 can be calculated for the given Gini coefficient:

$$L_2(z) = \left(\frac{1-\text{Gini}}{2}\right) + \frac{1}{2\sqrt{3}} P_1(z) + \frac{\text{Gini}}{2\sqrt{5}} P_2(z) = 3\text{Gini}\, z^2 + (1-3\text{Gini})z \tag{37}$$

This function becomes a nonnegative convex function if Gini < 1/3 because the convexity is satisfied if $\partial^2 L_2(z)/\partial z^2 \geq 0$ for all z.

(i) If the Gini coefficient is greater than 1/3, (37) will not be a convex function.

(ii) If the Gini coefficient is zero, $L(z) = z$;

(iii) If the Gini coefficient is 1/3, then $L(z) = z^2$.

The third-order polynomial series is:

$$L_3(z) = b_0 P_0(z) + b_1 P_1(z) + b_2 P_2(z) + b_3 P_3(z) \tag{38}$$

If we apply the boundary conditions $L_3(z=0)=0$ and $L_3(z=1)=1$, we have the following

$$L_3(z)=\left(\frac{1-\text{Gini}}{2}\right)+b_1 P_1(z)+\frac{\text{Gini}}{2\sqrt{5}}P_2(z)+\frac{(1-2\sqrt{3}\,b_1)}{2\sqrt{7}}P_3(z) \tag{39}$$

if $B=(1-2\sqrt{3}\,b_1)/2$, rewrite (39) as:

$$L_3(z)=(1-3\text{Gini}+5B)z+3(\text{Gini}-5B)z^2+10Bz^3 \tag{40}$$

Sufficient conditions to make (40) a positive convex function are:

$$B\geq 0,\qquad \text{Gini}\geq 5B,\quad 1-3\text{Gini}+5B\geq 0 \tag{41}$$

These conditions can be simplified as:

$$0\leq 5B<\text{Gini}\leq\frac{1+5B}{3} \tag{42}$$

This condition limits the range of $0\leq B\leq 0.1$ and $\text{Gini}\leq 0.5$. If the given data do not satisfy the above conditions, then the Lorenz curve derived by (40) may not be a nonnegative convex function. If the Gini coefficient is 0.5 and $B=0.1$, then $L(z)=z^3$.

V. Applications

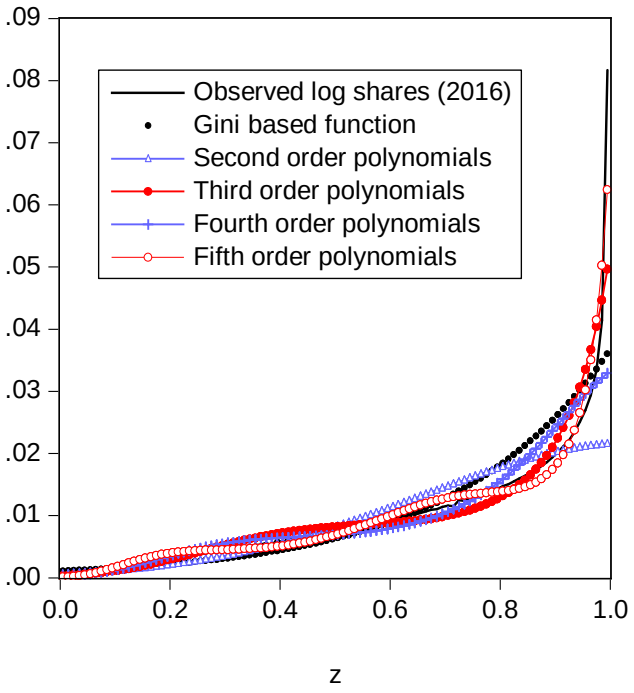
In order to illustrate the usefulness of the new measure, we present examples using Current Population Survey (CPS) data from 2000-2016. The CPS is sponsored jointly by the U.S. Bureau of the Census and the U.S. Bureau of the Census. The CPS produced a technical paper, TP66, which describes the design and methodology of the CPS, cf. www.bls.census.gov/cps/tp66.htm.

We use CPS household income data disaggregated into centiles for the years 2000-2016.[7] The distribution of the data for each year can be summarized by the Gini index. Now using the logarithmic share function given in (26), we can calculate a secondary measure to supplement the Gini index.

In Fig.2, the approximated function converges to the observed income shares for 2016 as we increase the number of expansion terms. The Gini-based model in (12) is a basic model, and it performs poorly for the very richest income group. Even-order polynomials of the second-order in (13) and fourth-order in (15) performed badly because the even power terms of the Legendre polynomial terms are symmetric functions, and do not fit well for the monotonically increasing function. The third-order model in (14) seems to perform well, but the fifth-order model in (16) produced minor fluctuations in the middle range of the IDF.

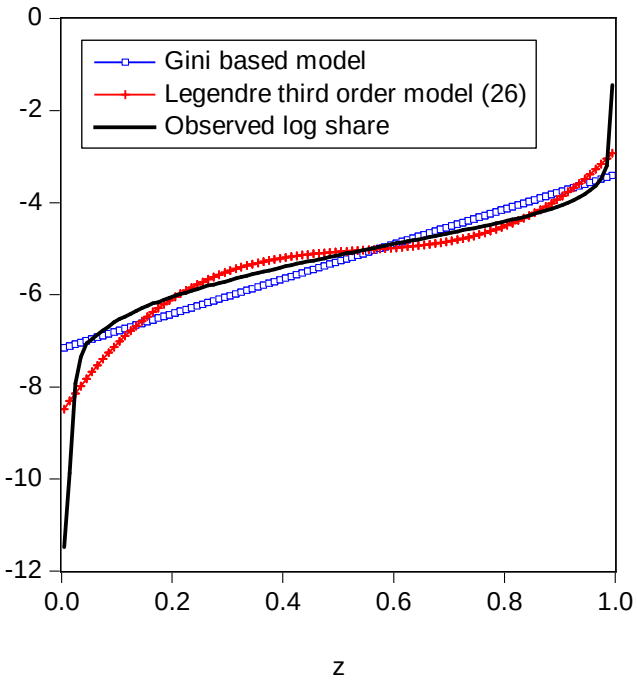
[7] We are grateful to Martha Starr for providing these data to us.

Fig.2 Converg. of Legendre polynomials to obs. log shares



In Fig.3, the Gini-based model produced a straight line and could not approximate the share values for the very poor and very rich groups properly. In comparison, if the third-order term is added, (26) showed an improved result for the poorest and very richest group. In the middle ranges, slight improvements were observed.

Fig.3 Approx. log shares with Gini and third order models



In Fig. 4, the performance of the third-order model of (26) is shown. Except for the very rich group, this model provided a relatively good performance. In Fig. 5, the performance of the fifth-order model of (28) is shown. Here, there is a small fluctuation around $z = 0.7$, but it produced a better performance for the richest group.

Fig.4 Approximated observed shares with third order model

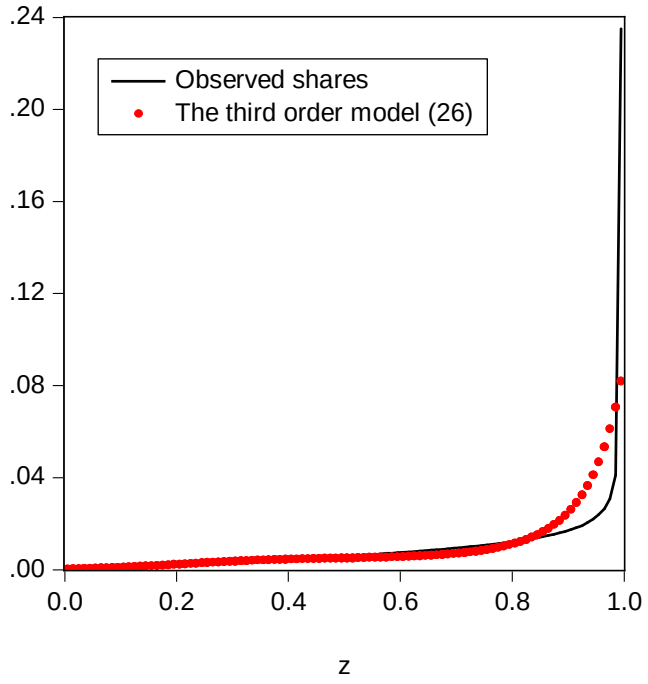
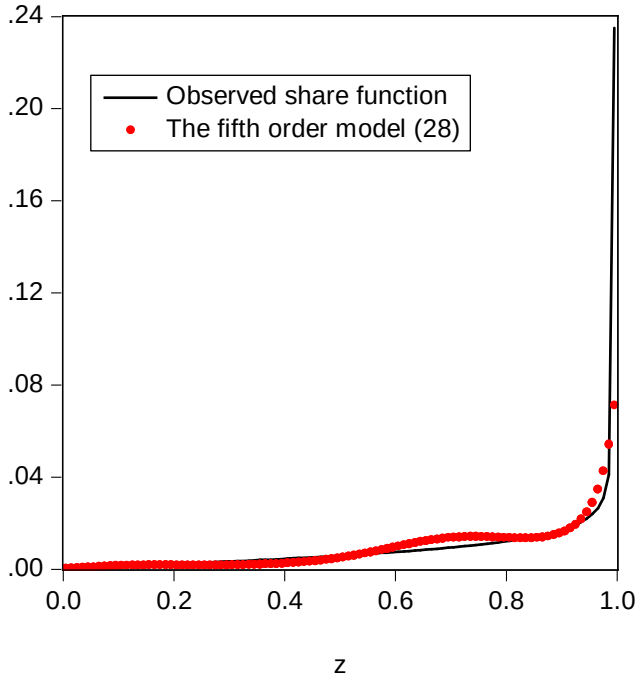


Fig.5 Approximate observed shares with fifth order model



In Fig. 6, we used the CPS data from the year 2000 and examined the performance of the Legendre polynomial series expansion of the Lorenz curve. To impose the convexity of an approximated Lorenz curve of a third-order polynomial series, the Gini coefficient should not be larger than 0.5, as stated below (42). The Gini coefficient for CPS data in 2000 is 0.490. The CPS data for the years 2012~2016 have Gini coefficients greater than 0.5. If the Gini coefficient is larger than 0.5, we need a higher-order Legendre polynomial series expansion instead of relying only on (39). In comparison, to impose the convexity of the approximated Lorenz curve of the second-order, the Gini coefficient should be less than 1/3, as stated below (37).

Fig.6 Approximate the Lorenz Curve for 2000

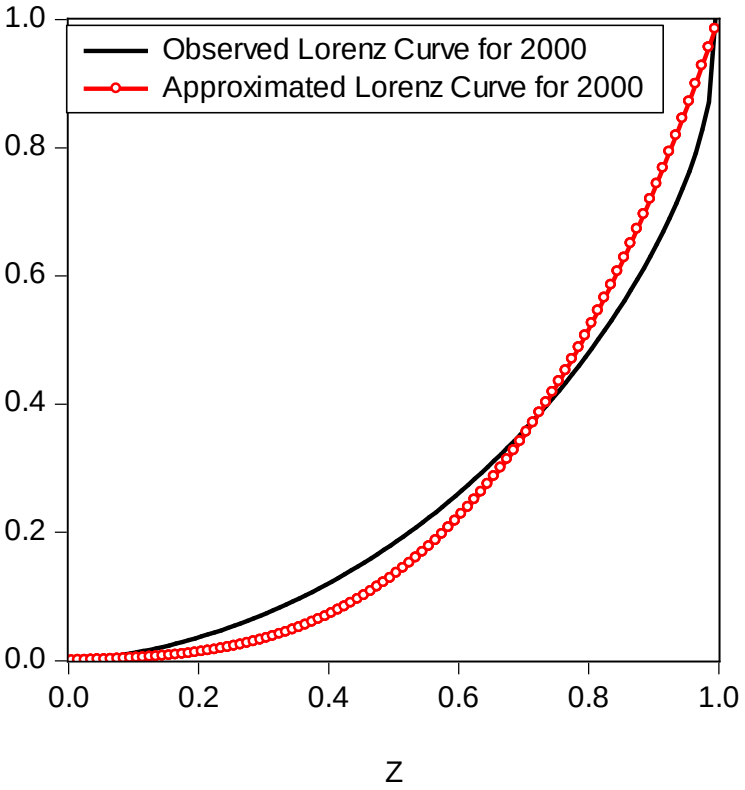
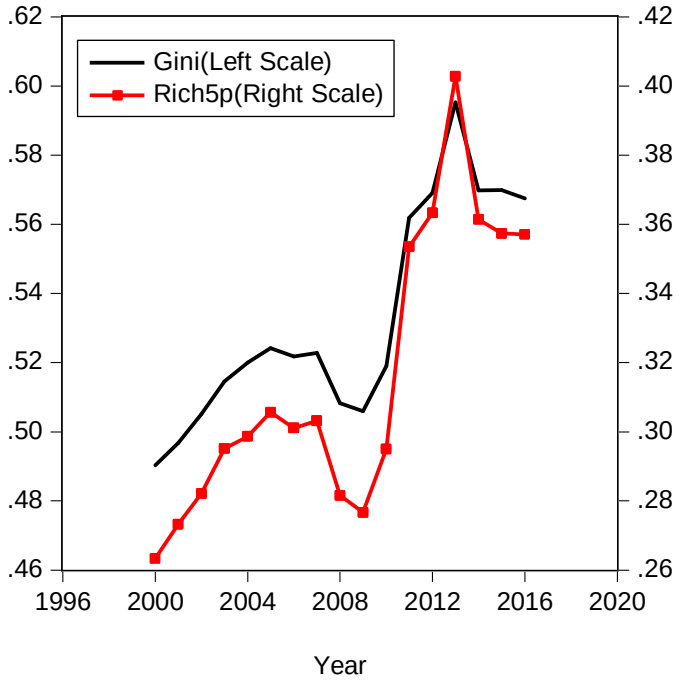


Fig.7 Comparison of Gini and richest 5% movements



In Fig. 7, the movements of the Gini coefficient and income shares of the richest 5% are compared. They move more or less in the same directions, though the gap between the two curves decreased after 2012. This means the Gini coefficient is not as sensitive to extreme movement in the highest percentiles of income earners.

Fig.8 Comparison of a3(ONB), rich 5P, and poor 5P

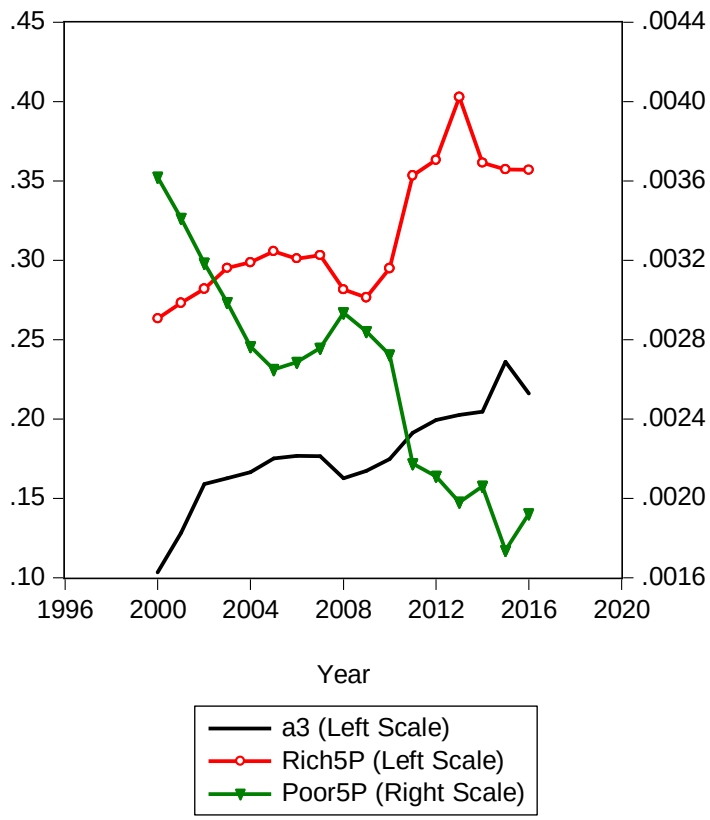


Fig. 8 shows the third order parameter (a_3) of an ONB expansion of the log share in (26). This parameter (a_3) moves in an opposite direction relative to the movements of the poorest 5 percent of income earners (poor 5P) curve. In 2015, the poorest 5P faced a significant loss in income share but recovered in 2016. The parameter (a_3) shows the opposite movements, indicating more inequality as the poorest group suffered a loss in income share. For movement of the richest 5P and parameter (a_3), a similar trend is observed but more refined details are different. Here, the (a_3) measure goes up as the richest share increases and goes down as the richest share decreases.

Fig.9 Compare Gini, Theil, a3(ONB), Rich 5P, and Poor 5P

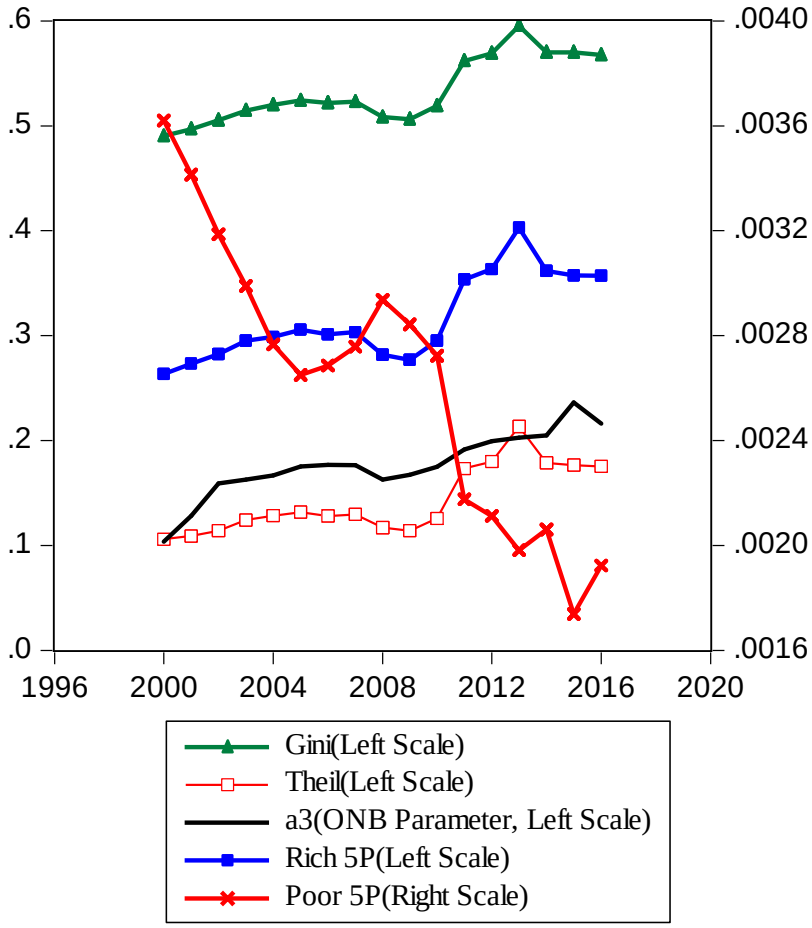


Fig.9 shows the usefulness of the Gini coefficient, Theil's entropy measure, and the third order parameter (a_3) in describing the movements of the poorest 5P and the richest 5P.

The Gini coefficient and Theil's measure are more or less the same in that they are both are reasonably good at describing the movement of the richest 5P. As explained in the discussion of Fig. 7, the third parameter (a_3) was stronger in describing the movement of the poorest 5P group's share.

To check the performance of the Gini, Theil, and the third parameter a_3, a curve-fitting exercise is performed where least squares estimation results are compared:

$$P5 = 0.01124_{(0.0007662)} - 0.01616_{(0.001435)}\text{Gini} + u_1, \quad R^2 = 0.8943 \tag{43}$$

$$P5 = 0.004793_{(0.0002582)} - 0.01523_{(0.001768)}\text{Theil} + u_2, \quad R^2 = 0.8319 \tag{44}$$

$$P5 = 0.008385_{(0.0004459)} - 0.007416_{(0.001144)}\text{Gini} - 0.01025_{(0.001157)}a_3 + u_3, \quad R^2 = 0.9840 \tag{45}$$

$$R5 = -0.3677_{(0.01404)} + 1.2824_{(0.02630)}\text{Gini} + u_4, \quad R^2 = 0.9937 \tag{46}$$

$$R5 = 0.1371_{(0.003108)} + 1.2544_{(0.02128)}\text{Theil} + u_5, \quad R^2 = 0.9957 \tag{47}$$

$$R5 = -0.4111_{(0.01355)} + 1.4155_{(0.03475)}\text{Gini} - 0.1559_{(0.03514)}a_3 + u_6, \quad R^2 = 0.9974 \tag{48}$$

Equations (45) and (48) show that the poorest group and the richest group are both described well if the Gini coefficient and the third parameter a_3 are used simultaneously, as these combinations provide the best fit of the data.

VI. Conclusion

This paper introduced a new inequality measure to supplement the better known Gini Index, where the new measure is sensitive to the asymmetries and extreme values in the underlying IDF that the index is intended to measure. The inequality measurement literature contains hundreds of papers on an appropriate index of income inequality, and on what desirable properties such a measure (or index) should contain.

There is a concurrent literature on the use of hypothetical statistical distributions to approximate and describe an observed distribution of incomes. Even with the recognition by some of the fact that incomes are distributed with asymmetric higher moments, inequality indices constructed to capture the level of inequality inherent in these observed income distributions (with a single number) are generally based on the mean and variance of the observed data. This paper introduced a new inequality measure to supplement, but not to replace, the Gini coefficient that measures more accurately the inherent asymmetries and extreme values that are present in observed income distributions.

The new measure is based in a third-order term of a Legendre polynomial from the logarithm of a share function (or a first-order term of a Lorenz curve). In this paper, we advocated using the two measures together to provide a better description of inequality inherent in empirical income distributions with extreme values.

We applied the new measure to examine inequality in U.S. CPS household income data for 2000-2016 in income centiles. The new measure was shown to be an excellent supplement to the Gini coefficient. The Gini index provides an intuitive overall measure of the inequality inherent in an IDF. Changes in the level of inequality inherent in the empirical IDF (particularly for the extreme portions of the IDF) were detected more accurately by the new measure than by simply calculating the Gini index alone.

References

Arfken, George, 1985, Mathematical methods for physicists, third edition, Academic Press, Inc. San Diego.

Basmann, R. and D. Slottje, (1987), "A new index of income inequality," *Economics Letters* 24: 385-389.

Basmann, R., K. Hayes, and D. Slottje, (1991), "The Lorenz curve and the mobility function," *Economics Letters,* 35: 105-111.

Boushey, H., J. Delong, and M. Steinbaum, (2017), *After Piketty*, Harvard University, Cambridge, MA.

Choo, Hakchung, and Hang Ryu, 1994, Gini coefficient, Lorenz curves, and Lorenz dominance effect: An application to Korean income distribution data, *Journal of Economic Development* 19, No.2, 47-65.

Coles, S. (2001), *An introduction to Statistical Modeling of Extreme Values*, Springer-Verlag.

Cowell, F. (2011), *Measuring Inequality*, 3rd Edition, Oxford: Oxford U. Press.

Cowell, F. and E. Flachaire (2002), "Sensitivity of Inequality Measures to Extreme Values," LSE STICERD Paper No. DARP 60.

Cowell, F. and E. Flachaire (2007), "Income Distributions and Inequality Measurement: the Problem of Extreme Values," *Journal of Econometrics*, 141: 1044-1072.

Maasoumi, E. (1986), "The Measurement and Decomposition of Multidimensional Inequality," *Econometrica,* 54: 991-998.

Maasoumi, E. (1989), "Continuously Distributed Attributes and Measures of Multivariate Inequality," *Journal of Econometrics,* 42: 131-144.

McDonald, J.B. (1984), "Some Generalized Functions for the Size Distributions of Income," Econometrica, 52: 647 – 663.

McDonald, J., J. Sorenson and P. Turley (2013), "Skewness and Kurtosis Properties of Income Distribution Models," *Review of Income and Wealth*, 59: 360 – 374.

Milne, W. (1949), *Numerical Calculus*, Princeton University Press, Princeton.

Pareto, V. (1876), *Cours d'Économie Politique Professé a l'Université de Lausanne*.

Piketty, T. (1995), "Social Mobility and Redistributive Politics", *Quarterly Journal of Economics*, 110: 551-584.

Piketty, T. (2014), *Capital in the Twenty-First Century*, , Harvard University Press, Cambridge .

Ryu, H. (1993), "Maximum entropy estimation of density and regression functions", *Journal*

of Econometrics, 56: 397-440.

Ryu, H. (2013), “A bottom poor sensitive Gini coefficient and maximum entropy estimation of income distributions, *Economics Letters,* 118: 370-374

Ryu, H. and D. Slottje, (1996), "Two Flexible Functional Form Approaches for Approximating the Lorenz Curve", *Journal of Econometrics,* 72: 251-274.

Ryu, H. and D. Slottje, (1998), *Measuring Trends in U.S. Income Inequality, Theory and Applications*, Springer, New York.

Ryu, H. and D. Slottje (2017), “Maximum Entropy Estimation of Income Distributions from Basmann’s WGM Class,” *Journal of Econometrics*, 199 (2): 221-231.

Slottje, D. (1987), “Relative Price Changes and Inequality in the Size Distribution of Various Components of Income,” *Journal of Business and Economic Statistics*, 5: 19-26.

Yitzhaki, S. (2013), More than a dozen ways of spelling Gini, ch-2 in *The Gini Methodology*, Springer, 11-13.

Zellner, A. and R. Highfield, (1988), “Calculation of maximum entropy distributions and approximation of marginal posterior distributions,” *Journal of Econometrics,* 37: 195-209.

Appendix: Pigou-Dalton Principle (PDP) for model (26)

The logarithm of the share function can be expanded in the Legendre series:

$$log\ s_N(z) = a_0P_0 + a_1P_1 + a_2P_2 + a_3P_3 + \cdots + a_NP_N \tag{4}$$

Suppose we want to summarize income inequality with only a Gini coefficient. This corresponds to taking a basic Gini model (12) because higher-order Legendre polynomials do not influence the choice of a_0 and a_1:

Basic model: $$log\ s_{Gini}(z) = a_0 + a_1P_1(z) \tag{12}$$

The Gini coefficient can be determined from a_1 and vice-versa, as discussed in (11). Even if we include higher-order terms of (4), a_1 will be the same in (4) and (12).

Now to prove the PDP condition holds for our new measure, suppose $i < j$ and $s(z_i) < s(z_j)$. After a transfer of small income share (Δ) from the j[th] person to the i[th] person, new income shares of these two people become $s(z_i) + \Delta$ and $s(z_j) - \Delta$. This means the slope of $\log s(z)$ is now lower. Thus a_1 and the Gini coefficient are lower, and $\int [log\ s_N(z)]^2 dz$ has decreased. If $\int [log\ s_{Gini}(z)]^2 dz$ is a good approximation of $\int [log\ s_N(z)]^2 dz$, $a_0^2 + a_1^2$ will decrease because we have:

$$\int [log\ s_{Gini}(z)]^2 dz = a_0^2 + a_1^2 \tag{A1}$$

In the standard discussion, income transfers from a rich person to a poor person is described with a lower value of the Gini coefficient, but here the same effect is represented with lower values of $\int[log\ s_{Gini}(z)]^2 dz$ and $a_0^2 + a_1^2$.

Similarly, if the logarithm of the share function is approximated with the first-order and third-order Legendre polynomials, then the logarithm of the share function is summarized with the ONB parameters a_1 and a_3.

For the Third-order model:

$$log\ sh_3(z) = a_0 + a_1 P_1(z) + a_3 P_3(z) \tag{26}$$

The parameters a_1 of (12) and (26) are the same, and can be derived from the given Gini coefficient. If the income share transfer decreases $\int[log\ s_N(z)]^2 dz$, and if $\int[log\ sh_3(z)]^2 dz$ is a good approximation of $\int[log\ s_N(z)]^2 dz$, then the income share transfer lowers $a_0^2 + a_1^2 + a_3^2$:

$$\int[log\ sh_3(z)]^2 dz = a_0^2 + a_1^2 + a_3^2 \tag{A2}$$

Therefore, the PDP will have a decrease of $a_0^2 + a_1^2 + a_3^2$ which completes the proof.