

ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)



Advances in Decision Sciences

Volume 23
Issue 1
March 2019

Michael McAleer
Editor-in-Chief
University Chair Professor
Asia University, Taiwan



Published by Asia University, Taiwan

ADS@ASIAUNIVERSITY

A New Inequality Measure that is Sensitive to Extreme Values and Asymmetries *

Michael McAleer **

Department of Finance, Asia University, Taiwan
and
Discipline of Business Analytics, University of Sydney Business School, Australia
and
Econometric Institute, Erasmus School of Economics
Erasmus University Rotterdam, The Netherlands
and
Department of Economic Analysis and ICAE
Complutense University of Madrid, Spain
and
Institute of Advanced Sciences, Yokohama National University, Japan

Hang K. Ryu

Department of Economics, Chung Ang University, Seoul, Korea

Daniel J. Slottje

Department of Economics, Southern Methodist University, Dallas, USA

Revised: February 2019

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

** Corresponding author: michael.mcaleer@gmail.com

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 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. 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. This paper introduces a new inequality measure to supplement the Gini (not to replace it) that better measures 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. Using Current Population Survey data, we show we can better describe the overall IDF and better detect changes in the tails of the empirical IDF using the two measures concomitantly.

Keywords: Extreme income values, Distributional aspects missed by the Gini coefficient, Lorenz dominance effects, Orthonormal basis expansion, Legendre polynomials.

JEL: D31, D63

1. Introduction

Income inequality research has experienced 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)).

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 described well by the Gini coefficient (or by Theil's entropy measure), but the poorest group's share and precise details about the richest group's share generally are not described well by these measures. In this paper, a second inequality measure is introduced concurrently with the Gini coefficient to describe movements of the extreme values and asymmetries of observed income distributions as they change over time.

In Section 2 and 3 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 4 of the paper discusses an application by fitting the new measure to CPS data. Section 5 concludes the paper.

2. A New Measure of Inequality to Supplement the Gini Coefficient

Given the objective to find a new income inequality measure which is sensitive to extreme values, we propose to describe the income distribution concurrently 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 observed income distributions all having the same underlying 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 component 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.

In order 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.

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 $0 \leq z \leq 1$, 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, $P_0(z), P_1(z), P_2(z), P_3(z), \dots$. Arfken (1985) presents an explanation of the ONB method:

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

An orthonormal sequence satisfies:

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

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

$$a_m = \int P_m(z) \log s_N(z) dz = \int P_m(z) \left[\sum_{n=1}^N a_n P_n(z) \right] dz \quad (3)$$

(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:

$$\int_Z f(z) P_n(z) dz = 0 \quad \text{for } n = 0, 1, 2, \dots \quad (4)$$

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

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

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

Figure 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:

$$\log s_{\text{Gini}}(z) = a_0 + a_1 P_1(z) \quad \text{or} \quad s_{\text{Gini}}(z) = \exp[a_0 + a_1 P_1(z)] \quad (6)$$

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 (6) from the Gini coefficient, consider:

$$a_0 + a_1 P_1(z) = a_0 + a_1 \sqrt{3} (2z - 1) = a_0 - \sqrt{3} a_1 + 2\sqrt{3} a_1 z = A + Bz \quad (7)$$

$$\mu_1 = \int z s(z) dz = \int z \exp[A + Bz] dz = \left[\frac{B}{e^B - 1} \right] \int z \exp[Bz] dz = \frac{1 + \text{Gini}}{2} \quad (8)$$

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 (7). Therefore, the basic model is derived from the given Gini coefficient.

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

Basic model: $\log s_{Gini}(z) = a_0 + a_1 P_1(z) \quad (9)$

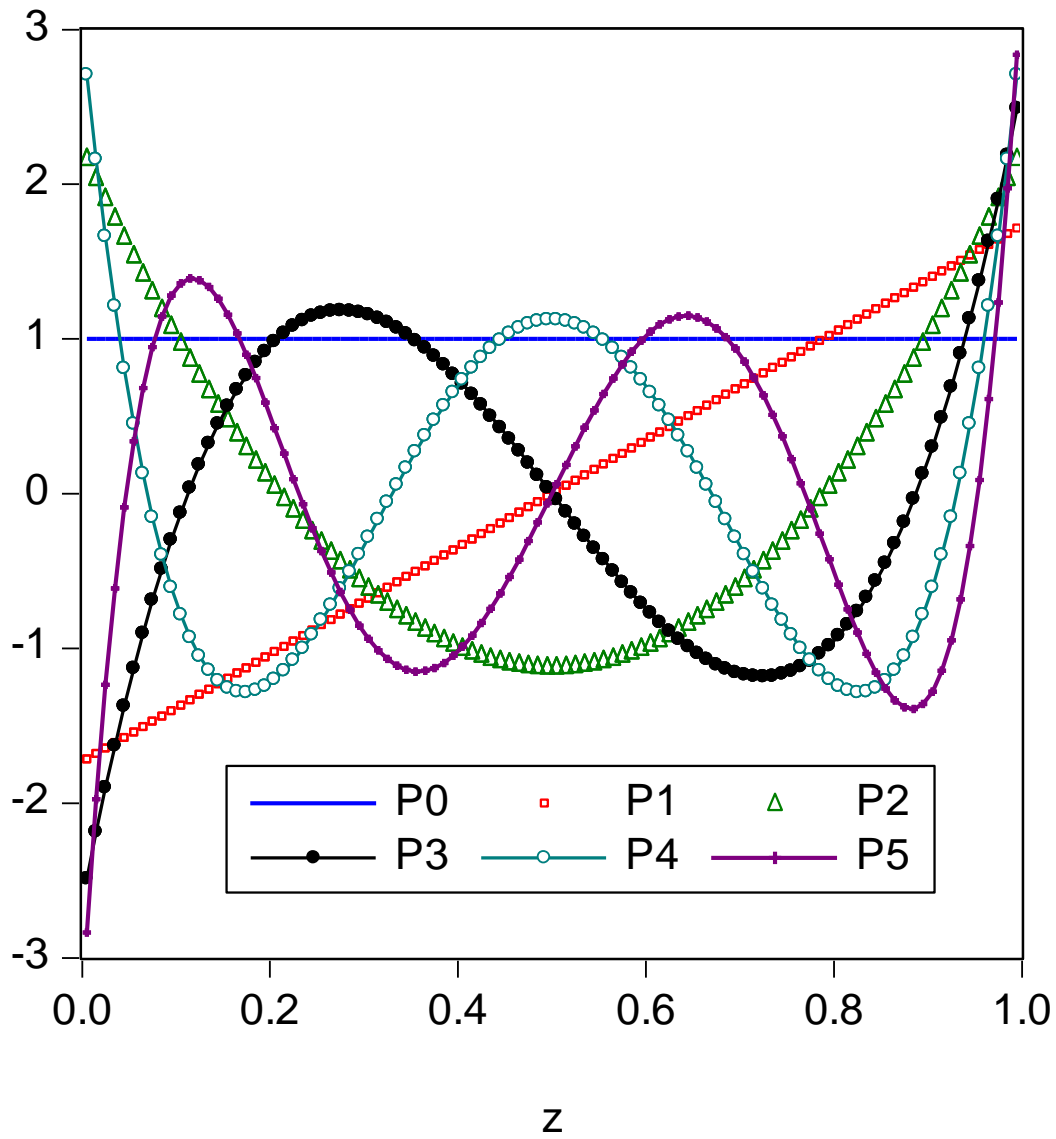
Second order: $\log s_2(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) \quad (10)$

Third order: $\log s_3(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) + a_3 P_3(z) \quad (11)$

Fourth order: $\log s_4(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) + a_3 P_3(z) + a_4 P_4(z) \quad (12)$

Fifth order: $\log s_5(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) + a_3 P_3(z) + a_4 P_4(z) + a_5 P_5(z) \quad (13)$

Fig.1 Plots of Legendre Polynomials



The parameters can be found with:

$$a_m = \int P_m(z) \log s_N(z) dz \quad (3)$$

The parameter values calculated by (3) do not depend on the length of the series. For example, the a_2 parameters of (10), (11), (12), and (13) 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 (1), so that the parameters were found with (3) using the orthogonality of the Legendre functions.

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

$$\int [\log s_N(z)]^2 dz = \int \left[\sum_{n=1}^N a_n P_n(z) \right]^2 dz = a_0^2 + a_1^2 + a_2^2 + \dots + a_N^2 \quad (14)$$

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

$$\begin{aligned} a_0^2 &= 27.921, & a_1^2 &= 1.190, & a_2^2 &= 0.0376, & a_3^2 &= 0.1340, \\ a_4^2 &= 0.0146, & a_5^2 &= 0.0740 \end{aligned} \quad (15)$$

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:

$$\text{Typical model:} \quad \log sh_N(z) = a_0 + a_1 P_1(z) + a_N P_N(z) \quad (16)$$

$$\text{Basic model:} \quad \log s_{Gini}(z) = a_0 + a_1 P_1(z) \quad (9)$$

$$\text{Second order model:} \quad \log s_2(z) = a_0 + a_1 P_1(z) + a_2 P_2(z) \quad (10)$$

$$\text{Third order model:} \quad \log sh_3(z) = a_0 + a_1 P_1(z) + a_3 P_3(z) \quad (17)$$

$$\text{Fourth order model:} \quad \log sh_4(z) = a_0 + a_1 P_1(z) + a_4 P_4(z) \quad (18)$$

$$\text{Fifth order model:} \quad \log sh_5(z) = a_0 + a_1 P_1(z) + a_5 P_5(z) \quad (19)$$

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 :

$$\frac{\partial \log sh_3(z)}{\partial z} = \frac{\partial a_0 + a_1 P_1(z) + a_3 P_3(z)}{\partial z} = 2\sqrt{3} a_1 + \sqrt{7} a_3 (60z^2 - 60z + 12) > 0 \quad (20)$$

Later, we use the CPS data for 2000~2016, the slope is positive with minimum values between 1.66 and 2.14.

If a monotonicity test is passed for (20), then the third-order parameter a_3 can be used as the second inequality measure. A similar monotonicity test can be performed for (19):

$$\frac{\partial \log sh_5(z)}{\partial z} = \frac{\partial a_0 + a_1 P_1(z) + a_5 P_5(z)}{\partial z} > 0 \quad (21)$$

Later in Figure 5, we find the slope is not positive for the data we used unless one imposes further restrictions.

In applying a_1 as an inequality measure and using a_3 as a second supplementary measure, we must first confirm that the new proffered measures satisfy desirable properties that any inequality measure must possess. Since a_1 can be converted to the Gini coefficient, all well-known properties are satisfied. For the measure a_3 , anonymity (any permutation of income y), scale independence (every person's income doubled), population independence, and egalitarian zero (when everyone has the same income) are satisfied.

The condition of being bounded above by maximum inequality and below by zero is not necessary for the supplementary measure. The Pigou-Dalton transfer principle states that if a unit of 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 the Appendix to this paper, we demonstrate that this property holds.

3. Lorenz Dominance Effects 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 effects. There are many Lorenz curves which can generate the same Gini coefficient. If we expand the Lorenz curve with a Legendre polynomial series, the zeroth 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 improve 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:

$$L_N(z) = \sum_{n=1}^N b_n P_n(z) \tag{22}$$

The parameters can be found from the following relation:

$$b_m = \int P_m(z) L_N(z) dz = \int P_m(z) \left[\sum_{n=1}^N b_n P_n(z) \right] dz \quad (23)$$

The Gini coefficient determines the zero-th order parameter:

$$\frac{1 - \text{Gini}}{2} = \int_0^1 L(z) dz \approx \int_0^1 L_N(z) dz = b_0 \quad (24)$$

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:

$$L_N(z=0) = 0 \quad \text{and} \quad L_N(z=1) = 1 \quad (25)$$

Now using:

$$P_n(z=0) = (-1)^n \sqrt{2n+1} \quad \text{and} \quad P_n(z=1) = \sqrt{2n+1} \quad (26)$$

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

$$L_2(z) = b_0 P_0(z) + b_1 P_1(z) + b_2 P_2(z) \quad (27)$$

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:

$$L_2(z) = \left(\frac{1 - \text{Gini}}{2} \right) + \frac{1}{2\sqrt{3}} P_1(z) + \frac{\text{Gini}}{2\sqrt{5}} P_2(z) = 3\text{Gini} z^2 + (1 - 3\text{Gini}) z \quad (28)$$

This function becomes a nonnegative convex function if $\text{Gini} < 1/3$ 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 $1/3$, (28) 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:

$$L_3(z) = b_0 P_0(z) + b_1 P_1(z) + b_2 P_2(z) + b_3 P_3(z) \quad (29)$$

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

$$L_3(z) = \left(\frac{1 - \text{Gini}}{2} \right) + b_1 P_1(z) + \frac{\text{Gini}}{2\sqrt{5}} P_2(z) + \frac{(1 - 2\sqrt{3} b_1)}{2\sqrt{7}} P_3(z) \quad (30)$$

if $B = (1 - 2\sqrt{3} b_1) / 2$, rewrite (30) as:

$$L_3(z) = (1 - 3\text{Gini} + 5B)z + 3(\text{Gini} - 5B)z^2 + 10Bz^3 \quad (31)$$

Sufficient conditions to make (31) a positive convex function are:

$$B \geq 0, \quad \text{Gini} \geq 5B, \quad \text{and} \quad 1 - 3\text{Gini} + 5B \geq 0 \quad (32)$$

These conditions can be simplified as:

$$0 \leq 5B < \text{Gini} \leq \frac{1+5B}{3} \quad (33)$$

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 (31) may not be a nonnegative convex function. If the Gini coefficient is 0.5 and $B = 0.1$, then $L(z) = z^3$.

4. 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 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 2000-2016.¹ The distribution of the data for each year can be summarized by the Gini index. Now using the logarithmic share function given in (17), we can calculate a secondary measure to supplement the Gini index.

In Figure 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 (9) is a basic model, and it performs poorly for the very richest income group. Even-order polynomials of the second-order in (10) and fourth-order in (12) 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 (11) seems to perform well, but the fifth-order model in (13) produced minor fluctuations in the middle range of the IDF.

In Figure 3, the Gini-based model produced a straight line and could not approximate the logarithm of the share function for the very poor and very rich groups properly. In comparison, if the third-order term is added, (17) showed an improved result for the poorest and very richest group. In the middle ranges, slight improvements were observed.

¹ We are most grateful to Martha Starr for providing the data.

In Figure 4, the performance of the third-order model of (17) 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 (19) is shown. Here, there is a small fluctuation around $z = 0.7$, but it produced a better performance for the richest group.

In Figure 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 (33). 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 (30). 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 (28).

In Figure 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

Figure 8 shows the third order parameter (a_3) of an ONB expansion of the log share in (17). 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.2 Conv of the Legendre Polynomial to obs shares

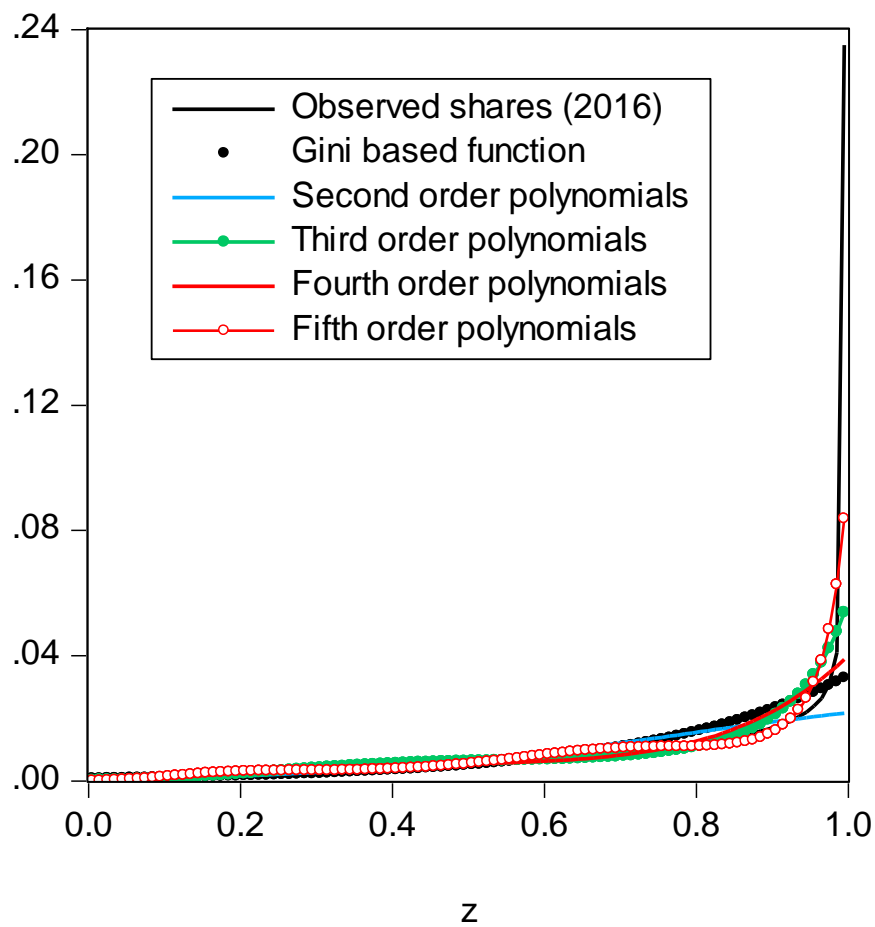


Fig.3 The Gini model and third order model for log share

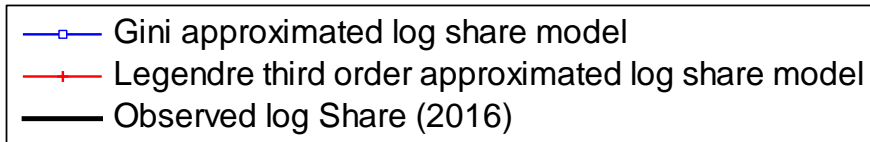
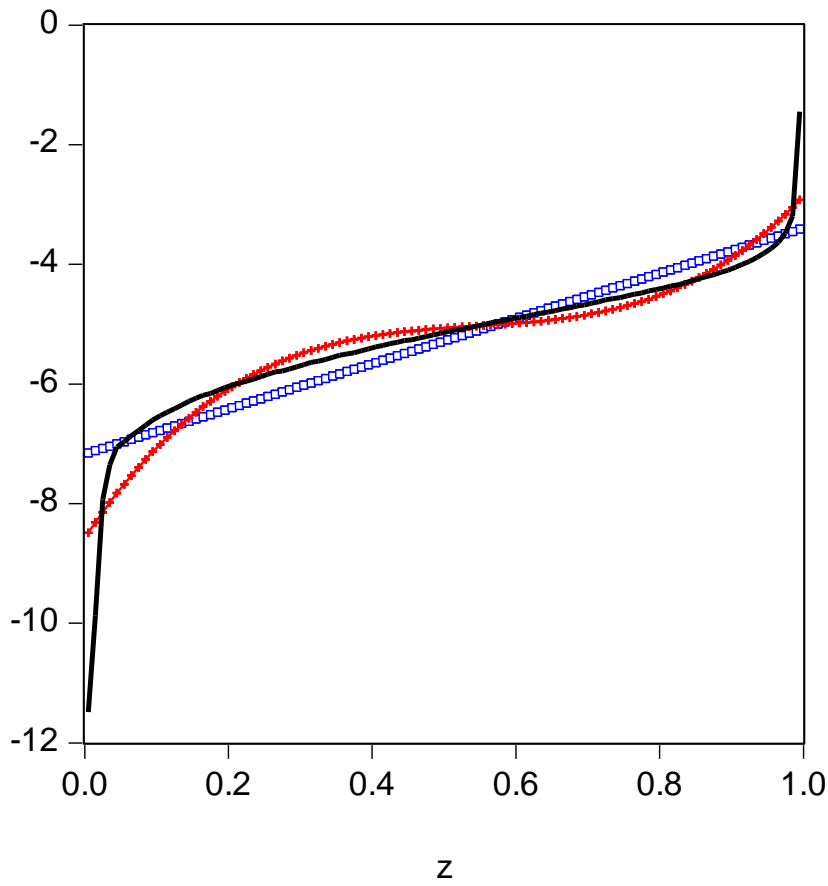


Fig.4 Approximated share function with third order model

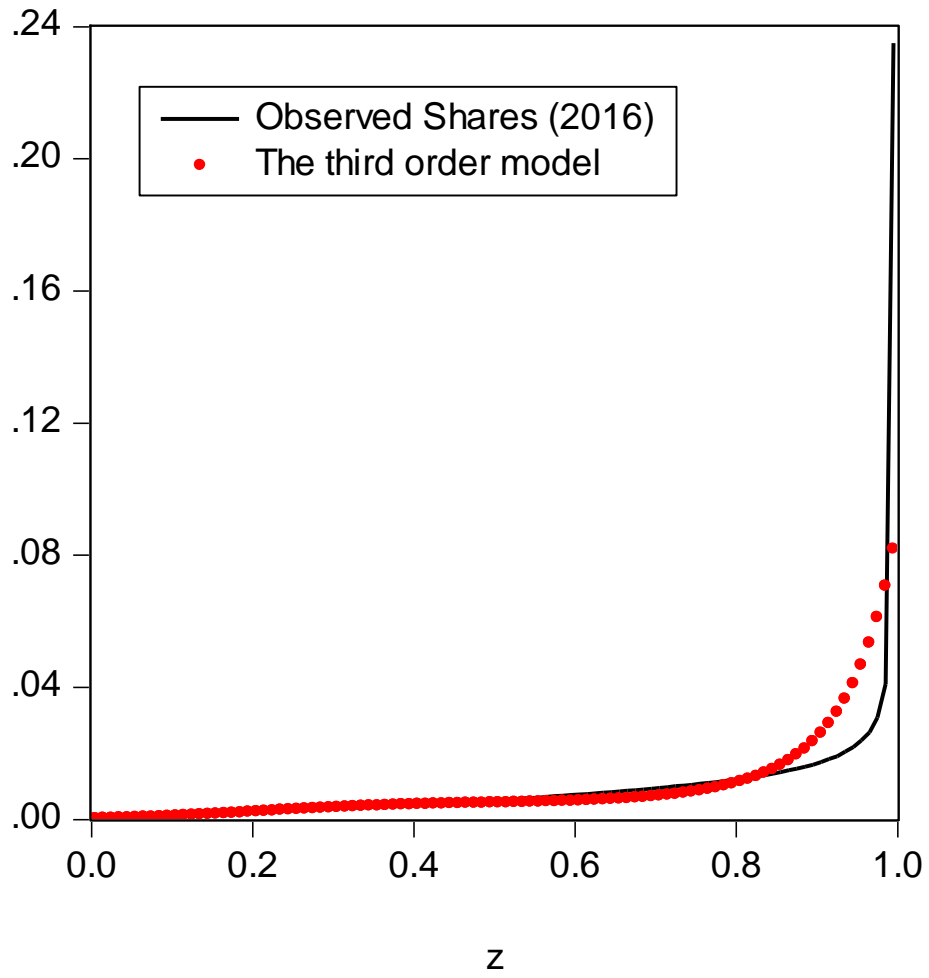


Fig.5 Approximate share function with fifth order model

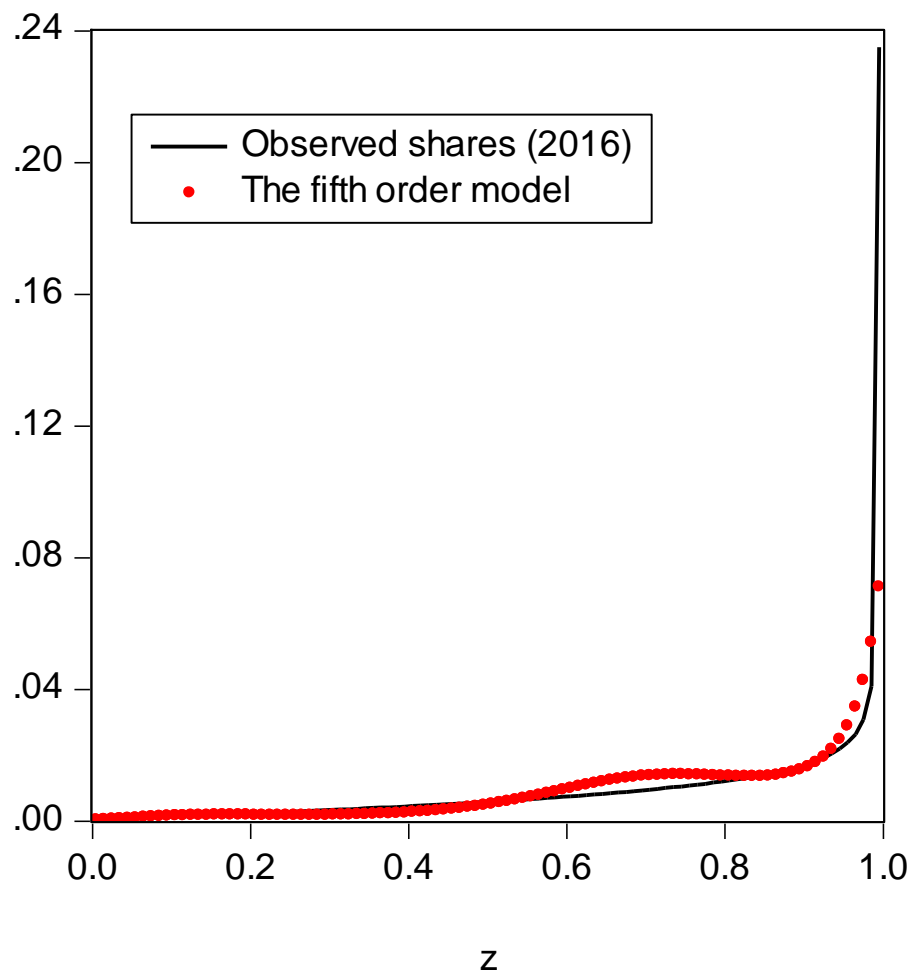


Fig.6 approximate the Lorenz curve for 2000

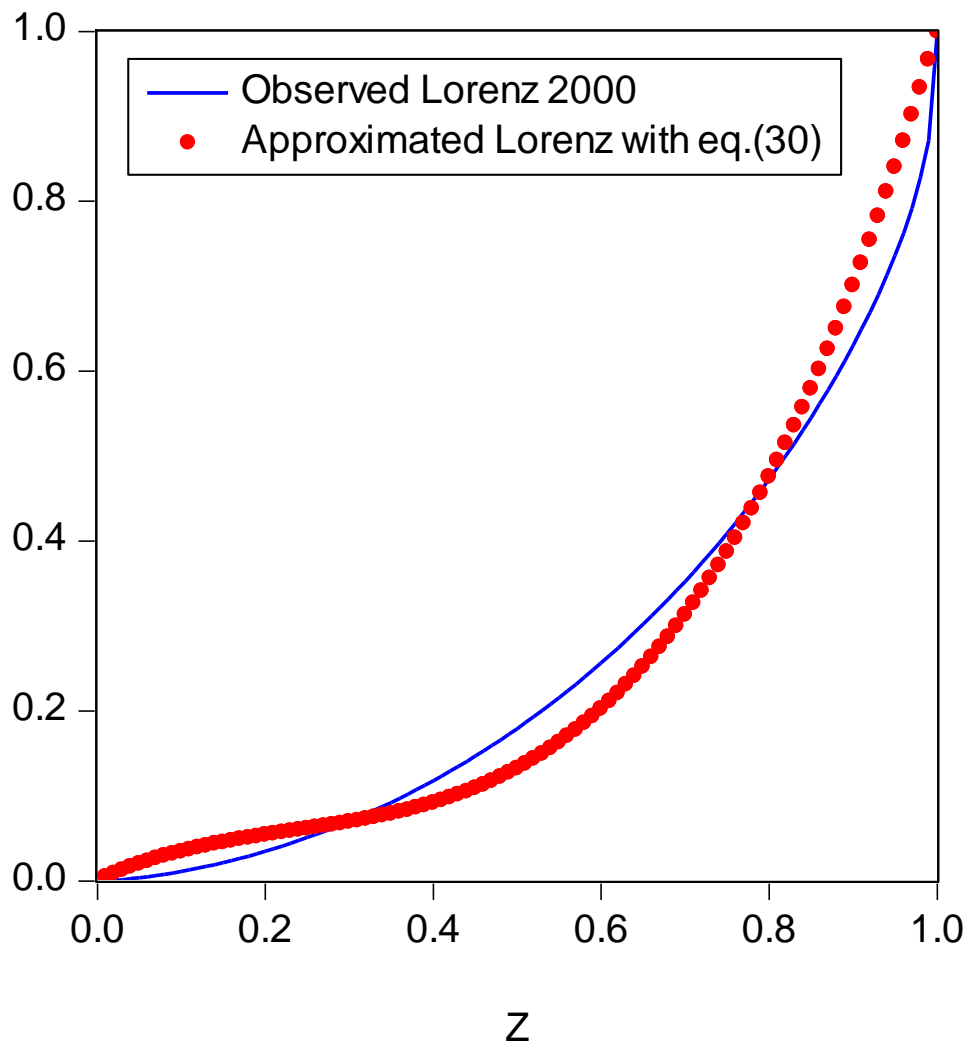


Fig.7 comparison of Gini and Rich 5% movements

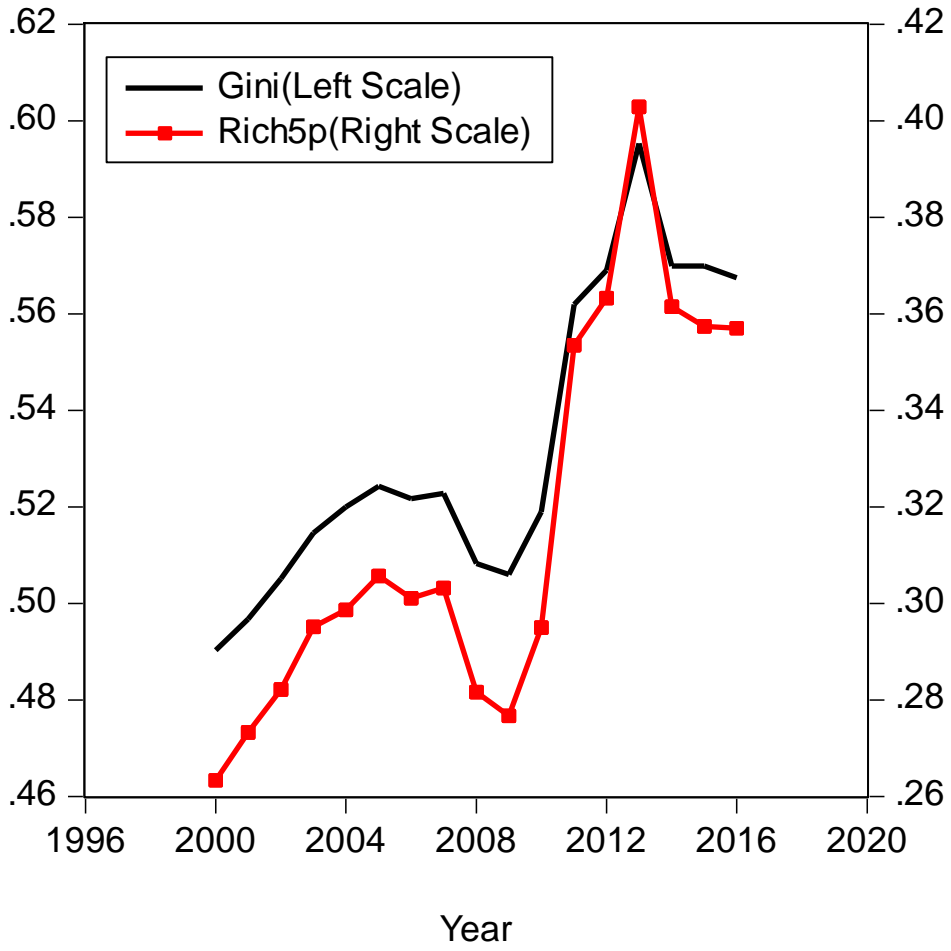


Fig.8 Comparison of a3(ONB), rich 5P, and poor 5P

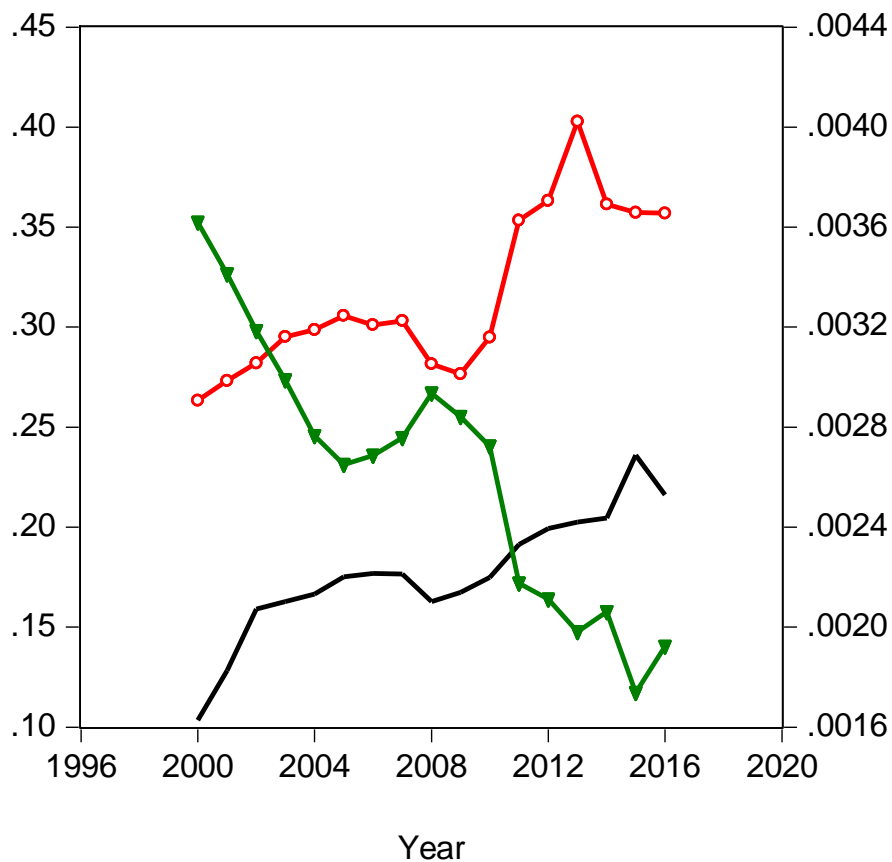


Fig.9 Compare Gini, Theil, a3(ONB), Rich 5P, and Poor 5P

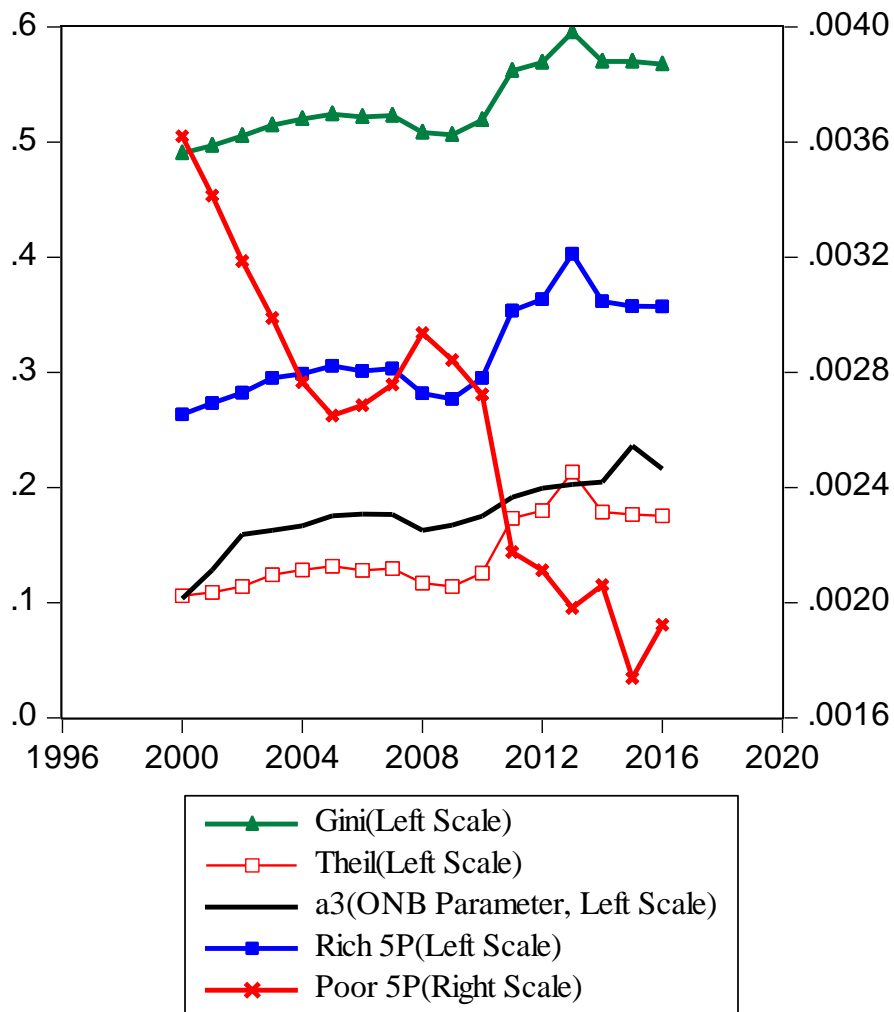


Figure.9 shows the usefulness of the Gini coefficient, Theil’s entropy measure, and the third order parameter (a_3) in describing the movements of the poorest 5P and the richest 5P.

The Gini coefficient and Theil’s measure are more or less the same in that they are both 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.

In order 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:

$$P5 = 0.01124_{(0.0007662)} - 0.01616_{(0.001435)} \text{Gini} + u_1, \quad R^2 = 0.8943 \quad (34)$$

$$P5 = 0.004793_{(0.0002582)} - 0.01523_{(0.001768)} \text{Theil} + u_2, \quad R^2 = 0.8319 \quad (35)$$

$$P5 = 0.008385_{(0.0004459)} - 0.007416_{(0.001144)} \text{Gini} - 0.01025_{(0.001157)} a_3 + u_3, \quad R^2 = 0.9840 \quad (36)$$

$$R5 = -0.3677_{(0.01404)} + 1.2824_{(0.02630)} \text{Gini} + u_4, \quad R^2 = 0.9937 \quad (37)$$

$$R5 = 0.1371_{(0.003108)} + 1.2544_{(0.02128)} \text{Theil} + u_5, \quad R^2 = 0.9957 \quad (38)$$

$$R5 = -0.4111_{(0.01355)} + 1.4155_{(0.03475)} \text{Gini} - 0.1559_{(0.03514)} a_3 + u_6, \quad R^2 = 0.9974 \quad (39)$$

Equations (34) and (36) 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.

5. 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.

Appendix

New interpretation of the Pigou-Dalton Principle (PDP) for model (17)

If income is transferred from a rich person to a poor person while still preserving the order of income ranks, then the inequality measure should not increase. For supplementary measure a_3 , a different interpretation is presented as follows.

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

$$\log s_N(z) = a_0 P_0 + a_1 P_1 + a_2 P_2 + a_3 P_3 + \dots + a_N P_N \quad (1)$$

$$\int [\log s_N(z)]^2 dz = \int \left[\sum_{n=1}^N a_n P_n(z) \right]^2 dz = a_0^2 + a_1^2 + a_2^2 + \dots + a_N^2 \quad (14)$$

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 $s(z_j) - \Delta$. Eq.(14) can be rewritten before and after the transition, and we want to show A is larger than B.

$$\begin{aligned} A &= \Delta z \left(\log [s(z_j)] \right)^2 + \Delta z \left(\log [s(z_i)] \right)^2 + \text{Remaining terms excluding } s(z_j) \text{ and } s(z_i) \\ &= a_0^2 + a_1^2 + a_2^2 + \dots + a_N^2 \end{aligned}$$

$$\begin{aligned} B &= \Delta z \left(\log [s(z_j) - \Delta] \right)^2 + \Delta z \left(\log [s(z_i) + \Delta] \right)^2 + \text{Remaining terms excluding } s(z_j) \text{ and } s(z_i) \\ &= b_0^2 + b_1^2 + b_2^2 + \dots + b_N^2 \end{aligned}$$

If we can show $A > B$ after the income transfer, then the sum of squared parameters has decreased. PDP originally considered a decrease of the chosen inequality measure after the income transfer, but under the orthonormal basis expansion, the sum of squared parameters can be considered as a measure of inequality. This is our new interpretation of PDP. When inequality worsens, both the left hand side and right side of (14) will increase, but one cannot tell about the direction of changes of the individual parameters nor about the supplementary inequality measure a_3 .

Let us show $A > B$. Expansion of square terms in A and B:

$$B - A = 2\log(s(z_j)) \cdot \left(-\frac{\Delta}{s(z_j)}\right) + 2\log(s(z_i)) \cdot \left(\frac{\Delta}{s(z_i)}\right) + \text{remaining small terms} < 0$$

because $s(z_i) < s(z_j)$, $\log(s(z_i)) < 0$, $\log(s(z_j)) < 0$, and $|\log(s(z_j))| < |\log(s(z_i))|$.

References

- Arfken, G.: Mathematical methods for physicists, third edition, Academic Press, Inc. San Diego (1985).
- Boushey, H., DeLong, J., Steinbaum, M.: *After Piketty*, Harvard University, Cambridge, MA (2017).
- Choo, H., Ryu, H.: Gini coefficient, Lorenz curves, and Lorenz dominance effect: An application to Korean income distribution data, *Journal of Economic Development* 19, No.2, 47-65 (1994).
- Coles, S.: *An introduction to Statistical Modeling of Extreme Values*, Springer-Verlag (2001).
- Cowell, F.: *Measuring Inequality*, 3rd Edition, Oxford: Oxford U. Press (2011).
- Cowell, F., Flachaire, E.: Sensitivity of Inequality Measures to Extreme Values, LSE STICERD Paper No. DARP 60 (2002).
- Cowell, F., Flachaire E.: Income Distributions and Inequality Measurement: the Problem of Extreme Values, *Journal of Econometrics*, 141: 1044-1072 (2007).
- McDonald, J.: Some Generalized Functions for the Size Distributions of Income, *Econometrica*, 52: 647 – 663 (1984).
- McDonald, J., Sorenson, J., Turley, P.: Skewness and Kurtosis Properties of Income Distribution Models, *Review of Income and Wealth*, 59: 360 – 374 (2013).
- Milne, W.: *Numerical Calculus*, Princeton University Press, Princeton (1949.)
- Pareto, V.: *Cours d'Économie Politique Professé a l'Université de Lausanne* (1896).
- Piketty, T.: Social Mobility and Redistributive Politics, *Quarterly Journal of Economics*, 110: 551-584 (1995).
- Piketty, T.: *Capital in the Twenty-First Century*, , Harvard University Press, Cambridge (2014)
- Ryu, H.: Maximum entropy estimation of density and regression functions, *Journal of Econometrics*, 56: 397-440 (1993).
- Ryu, H.: A bottom poor sensitive Gini coefficient and maximum entropy estimation of income distributions, *Economics Letters*, 118: 370-374 (2013).
- Ryu, H., Slottje, D.: Two Flexible Functional Form Approaches for Approximating the Lorenz Curve, *Journal of Econometrics*, 72: 251-274 (1996).
- Ryu, H., Slottje, D.: Maximum Entropy Estimation of Income Distributions from Basmann's WGM Class, *Journal of Econometrics*, 199 (2): 221-231 (2017).

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

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