International Journal of Business and Statistical Analysis
ISSN (2384-4663)
Int. J. Bus. Stat. Ana. 1, No. 1 (July-2014)

Characterization of Income Distribution

T. H. M. Abouelmagd* and I. A. Ahmed

Department of Statistics, Oklahoma State University, USA

Received: 14 Feb. 2014, Revised: 14 May 2014, Accepted: 15 May 2014, Published: 1 July 2014

Abstract: Methods to measure income inequality have for quite some time now been an important subject in statistics and econometric research. Some measures are based on incomplete moments or incomplete conditional moments and they take into consideration the shape of the income distribution but suffer sometimes from low efficiency and or lack or robustness. On the other hand, in recent years a new inferential method called "the probability weighted moments" (PWM) was introduced and studies as a competitor to more traditional inferential methods such as the method of moments or the maximum likelihood method. A class of generalized measures of income inequalities using the PWM are introduced and studied. The new measures are also shown to characterize the income distributions well.

Keywords: Income inequality measures, incomplete moments, probability weighted moments.

1. Introduction

The Lorenz curve is a tool used to represent income distribution; it relates the cumulative proportion of income to cumulative proportion of individuals, but Lorenz curve depends on the income mean. However, several economic questions depend on the shape as well as the mean. Butler and McDonald (1987) used incomplete moments to characterize income inequality and provide the basis for interdistributional Lorenz curves. In defense of their approach, Butler and McDonald argued that these incomplete moments do explain the shape of distributions and also can be used readily to build measures of income inequality. They reintroduced Gini coefficient, Pietra measure, and Lorenz curve in terms of incomplete moments. Gastwirth et al (1989) gave a statistical analysis of a special case of the Butler-McDonald measures. They established its asymptotic normality under minimal conditions.

In this article, we summarize the inequality measures based on incomplete moments and conditional incomplete moments. Also, we summarize some theory for Probability Weighted Moments (PWM) and show how they can be used to estimate of the Inequality Measures. In addition, we propose a general class of inequality measures using PWM which gives fast, straightforward and unbiased estimators. In section 2, the PWM a generalization of the usual moments of probability distribution is introduced and how to use it to obtain unbiased estimators. In section 3, some desirable inequality measures based on incomplete moments are studied such as Butler-McDonald measure (1989) and Ahmad measure (1998). In section 4, we propose a General class of Inequality using PWM and give different special cases including Butler-McDonald and Ahmad measures. In section 5, we introduce the characterization of the income distribution using M^r_{PWM1} and M^s_{PWM2}. In Section 6, we propose the estimation of the new general class of inequality measures. In section 7, we have studied the proposed measures of inequality under Pareto distribution. Section 8, is devoted to conclusion.

2. Probability Weighted Moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with density function $f(x)$, quantile function $x(F) = F^{-1}(x) = Q(F)$, $0 < F < 1$, cumulative distribution function $F(x) = F$, mean $\mu = E(X)$, σ is the standard deviation of the distribution. The probability weighted moments are a generalization of the usual moments of a probability distribution. The probability weighted moments of a random variable X with distribution function $\mathrm{F}(x) = \mathrm{P}(\mathrm{x} \leq x)$ are the quantities

$$M_{p,r,s} = E[X^p\{F(X)\}^r\{1 - F(X)\}^s] \quad (1)$$

E-mail address: tamer.abouelmagd@okstate.edu, iahmad@okstate.edu

Where p, r and s are real numbers. PWM are likely to be most useful when quantile function $x(F)$ can be written in closed form, so we can rewrite

$$M_{p,r,s} = \int_0^1 [x(F)]^p\, F^r\, (1-F)^s dF \qquad (2)$$

It is easily noticeable that $M_{p,r,s}$ is the *p-th* moment of the (r +1)*th* order statistic in a sample of size (r +s +1). Note that using $M_{p,r,s}$ involves information about the value and rank of a random variable in a group of (r +s +1) sample unit; see Bartoluccia et al (1999), Greenwood et al (1979), and Hosking (1990).

The quantities $M_{p,0,0}$ are the usual non-central moments. When r and s are integers, $F^r(1-F)^s$ may be expressed as a linear combination of either powers of For powers of $(1-F)$, so it is natural to summarize a distribution either by the moments $M_{1,r,0}$ or $M_{1,0,s}$, where

$$\beta_r = M_{1,r,0} = E[X\{F(X)\}^r], \quad r = 0,1,2,\ldots \qquad (3)$$

and,

$$\beta_s = M_{1,0,s} = E[X\{1-F(X)\}^s], \quad s = 0,1,2,\ldots \qquad (4)$$

where the expected value of order statistics is

$$E(X_{r:n}) = \frac{n!}{(r-1)!(n-r)!} \int_0^1 x(F)\, F^{r-1}\, (1-F)^{n-r} dF \qquad (5)$$

Therefore the probability weighted moments can be written in terms of expected value of order statistics as

$$\beta_r = \int_0^1 x(F) F^r dF = \frac{E(X_{r+1:r+1})}{r+1}, \qquad r = 0,1,\ldots \qquad (6)$$

and

$$\beta_s = \int_0^1 x(F)(1-F)^s dF = \frac{E(X_{1:s+1})}{s+1}, \qquad s = 0,1,\ldots \qquad (7)$$

Note that β_r and β_s are used in many applications such as linear moments, estimation of the generalized extreme value distribution ,analysis of hydrological extremes, and estimating the three-parameter Weibull distribution; see for example, Hosking et.al (1985), Moisello (2007), and Wang (1990).

3. INCOMPLETE MOMENT MEASURES OF INCOME INEQUALITY

Butl Butler--McDonald measure

Let X denote the (random) income in an economic system (population) having distribution function (df) $F(x)$, probability density function (pdf) $f(x)$, y_r, y_s pre-chosen value and rth non-central moment

$$\mu_X^{(r)} = E(X^r) = \int_0^\infty x^r dF(x) \qquad (8)$$

The Butler-McDonald measure of income inequality is given by

$$M_{BM}^r = \frac{\int_0^{y_r} x^r dF(x)}{E(X^r)} = \frac{I(y;r)}{\mu_X^{(r)}} = \phi(y_r; r) \qquad (9)$$

The most important cases are when $r = 0$ or 1 because the first two incomplete moments, provide especially useful information in many applications about the shape of the distribution. In terms of the income distribution setting $r = 0$ yields the fraction of the population with income less than *x*, whereas r=1 yields the proportion of total income accounted for by those with income less than *x*, see; Butler and McDonald (1989), Gastwirth (1971) and Gastwirth et.al. (1989). Note that $\phi(y; r)$is *df's* and for some distributions, see; Butler and McDonald of (1989) it exhibits same properties and form as the original *df*. Gastwrith et.al. (1989) established the asymptotic normality of an estimate of $M_{BM}^{(0)}$ and a variation of it when the means are replaced by the medians. Gastwrith et al also compared the asymptotic relative efficiency of their estimate of $M_{BM}^{(0)}$ to the Wilcoxon and *t* statistics for location and scale alternatives.

Ahmad measure

Ahmad introduced a measure of income inequality based on conditional incomplete moments as

$$M_A^{(r)} = E(X^r \mid X^r \le y) = \frac{\int_0^{y_r} x^r dF(x)}{F(y)} = \psi(y_r; r) \qquad (10)$$

Note here that for $r = 0$, $M_A^{(0)}$ =0 . Ahmad had provided the asymptotic efficiency for the location and scale cases and focused on the comparison on differences between moments of the portions of populations that are affected with the inequality. In terms of the income distribution $M_A^{(r)}$ can be explained as the fraction of poor people in the population who have income less than y_r income of the population.

4. A UNIFIED GENERAL CLASS OF PWM INCOME INEQUALITY MEASURES

A proposed class of inequality measures which is a generalization of Butler-McDonald measure is

$$M_{PWM}^{(h,r,s)} = \frac{\int_0^{y_*}[x^h\{F(x)\}^r\{1-F(x)\}^s]dF(x)}{\int_0^1[x^h\{F(x)\}^r\{1-F(x)\}^s]dF(x)} = \phi(y_*; h, r, s) \qquad (11)$$

This can be called "normalized weighted incomplete moment" and the Butler-McDonald measure is a special case of $M_{PWM}^{(h,r,s)}$when $r = s = 0$.

Measure based on β_r

The proposed first linear version of measure of inequality based on PWM is

$$M_{PWM1}^{r} = \frac{\int_0^{y_r} xF^r dF(x)}{\int_0^1 xF^r dF(x)} = \frac{\int_0^{y_r} xF^r dF(x)}{E(XF^r)} = \frac{I_L(y_r;r)}{\beta_r} = \frac{(r+1)I_L(y_r;r)}{E(X_{r+1:r+1})} = \phi_L(y_r;r)$$

(12)

At $r = 0$ M_{PWM1}^{0} can be explained as the proportion of total income accounted for by those with income less than y_0. More general M_{PWM1}^{r} can be explained as the proportion of total weighted income accounted for by those with weighted income less than y_r. This version will give more weight for income in the largest part of the distribution.

Moreover the type of Ahmed measure based on PWM can be defined as

$$A_{PWM1}^{r} = \frac{\int_0^{y_r} xF^r dF(x)}{F(y_r)} = \psi_L(y_r;r) \quad (13)$$

This can be explained as the fraction of total weighted income accounted for poor people with weighted income less than y_r

Measure based on β_s

The proposed second linear version of measure of inequality based on PWM is

$$M_{PWM2}^{s} = \frac{\int_0^{y_s} x(1-F)^s dF(x)}{\int_0^1 x(1-F)^s dF(x)} = \frac{\int_0^{y_s} x(1-F)^r dF(x)}{E(X(1-F)^r)} = \frac{I_L(y_s;s)}{\beta_s} = \frac{(s+1)I_L(y_s;s)}{E(X_{1:s+1})} = \phi_L(y_s;s) \quad (14)$$

Moreover at $s = 0$ M_{PWM2}^{0} can be explained as the proportion of total income accounted for by those with income less than y_0. More general M_{PWM2}^{s} can be explained as the proportion of total weighted income accounted for by those with weighted income less than y_s. This version will give more weight for income in the lowest part of the distribution.

Moreover the type of Ahmed measure based on PWM can be defined as

$$A_{PWM2}^{s} = \frac{\int_0^{y_s} x(1-F)^s dF(x)}{F(y_s)} = \psi_L(y_s;s) \quad (15)$$

This can be explained as the fraction of total weighted income accounted for poor people with weighted income less than y_r.

5. CHARACTERIZATION OF THE INCOME DISTRIBUTION BASED ON M_{PWM1}^{r} AND M_{PWM2}^{s}

It can characterize the income distribution as follows.

Theorem 1: if $E|X| < \infty$, then the income distribution F is uniquely determined by

$$\left\{\frac{(s+1)I_L(y_s;s)}{M_{PWM2}^{s}};\quad s = 0,1,2,\ldots.\right\} \quad (16)$$

Or equivalently, by

$$\left\{\frac{(r+1)I_L(y_r;r)}{M_{PWM1}^{r}};\quad r = 0,1,2,\ldots.\right\} \quad (17)$$

Proof: From Chan (1967), Konheim (1971), and Haung and Lin (1988) who showed that $F(x)$ is uniquely determined by $\{E(X_{1:n});\ n = 1,2,3,\ldots\}$, therefore $\frac{(s+1)I_L(y;s)}{M_{PWM2}^{s}}$ will uniquely determine the income distribution for $s = 0,1,2,\ldots$ where it is equal to $E(X_{1:s+1})$. Similarly $\frac{(r+1)I_L(y;r)}{M_{PWM1}^{r}}$ will uniquely determine the income distribution for $r = 0,1,2,\ldots$where it is equal to $E(X_{r+1:r+1})$.

It is not necessary to have the full sequence in order to characterize the income distribution as follows.

Theorem 2: For any fixed positive integers a, b and $a, b > 0$ the subsequence

$$\left\{\frac{(s+1)I_L(y;s)}{M_{PWM2}^{s}};\quad s = (a-1), (a-1)+b, (a-1)+2b,\ldots.\right\}$$

(18)

Or equivalently, by

$$\left\{\frac{(r+1)I_L(y;r)}{M_{PWM1}^{r}};\quad r = (a-1), (a-1)+b, (a-1)+2b,\ldots.\right\}$$

(19)

is sufficient to characterize income distribution.

Proof: It can be proven directly from Haung (1975), Efromovich (1999), and Wand and Jones (1995) who showed that for any fixed positive integers a, b and $k(\leq a)$ the subsequence $\{E(X_{k:n}),\ n = a, a+b, a+2b,\ldots.\}$ is sufficient to characterize F by replacing n by $s+1$.

6. ESTIMATION

Using the sample income data $x_1, x_2, \ldots., x_n$ and the empirical distribution function the above measures can be estimated as follows.

$$\widehat{M}_{PWM1}^{r} = \frac{\sum_{i=1}^{n} I(w_{ri}x_i < y_r) w_{ri}x_i}{\sum_{i=1}^{n} w_{ri}x_i}$$

(20)

and

$$\widehat{M}_{PWM2}^{s} = \frac{\sum_{i=1}^{n} I(w_{si}x_i < y_s) w_{si}x_i}{\sum_{i=1}^{n} w_{si}x_i}$$

(21)

where

$$w_{ri} = \left(\frac{i - 0.5}{n}\right)^r \quad \text{and} \quad w_{si} = \left(1 - \frac{(i - 0.5)}{n}\right)^s$$

Moreover,

$$\hat{A}^r_{PWM1} = \frac{\sum_{i=1}^{n} I(w_{ri}x_i < y_r) w_{ri}x_i}{\hat{F}(y_r)} \quad (22)$$

and

$$\hat{A}^s_{PWM1} = \frac{\sum_{i=1}^{n} I(w_{si}x_i < y_s) w_{si}x_i}{\hat{F}(y_s)} \quad (23).$$

Where $\hat{F}$ is the empirical distribution function.

REFERENCES

[1] Ahmad, I.A. 1998. "Measuring inequality based on incomplete moments," Unpublished

[2] Bartoluccia, A.A., Singh, K.P., Bartolucci, A.D., and Bae, S. 1999. "Applying medical survival data to estimate the three-parameter Weibull distribution by the method of probability-weighted moments," Mathematics and Computers in Simulation. 48, Issues 4–6, pp. 385–392.

[3] Butler, R. J. and Mcdonald, J.B. 1989. "Using Incomplete Moments to Measure Inequality". J. of Economctrics. 18, pp. 109-119.

[4] Butler, R.J. And Mcdonald , J.B. 1987. " Interdistributional Income Inequality," J. of Business and Economic Statistics, 5, pp. 13-18.

[5] Chan,L.K. 1967. "On a characterization of distributions by expected values of extreme order statistics," Amer.Math.Monthly, 7, pp 950-951.

[6] Efromovich, S.Y. 1999. "Nonparametric Curve Estimation: Models, Theory and Applications," Springer, NewYork.

[7] Gastwirth, J.L. Nayak, T.K. and Wang, J.L. 1989. "Statistical Properties of Measures of between group Income Differentials," J. Econometrics, 42, pp 5-19.

[8] Gastwirth, J.L. 1971. "A general definition of the Lorenz curve," Econometrica,39, pp 1037-1039.

[9] Greenwood, J,A., Landwehr,J.M., Matalas, N.C., and Wallis, J.R. 1979. "Probability Weighted Moments :Definition and Relation to Parameters of Several Distributions Expressable in form," Water Resources Research, 15, pp1049-1054.

[10] Haung,J.S. 1975. "Characterization of distribution by the expected value of order statistics," Annals of the Institute of Statistical Mathematics, 27, Issue 1, pp 87-93.

[11] Haung,J.S. and Lin,G.D. 1988. "Characterization of distribution via two expected values of order statistics," Metrik, 35, pp 329-338.

[12] Hosking.J.R.M. 1990. "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," Royal Statistical Society, 52, pp 105-124.

[13] Hosking.J.R.M., Wallis,J.R., and Wood, E.F. 1985. "Estimation of the Generalized Extreme value Distribution by the Method of probability Weighted Moments," Technometrics, 27, N 3.

[14] Konheim,A.G. 1971. "A note on order statistics," Amer.Math.Monthly,78,pp 524.

[15] Moisello,U. 2007. "On the use of partial probability weighted moments in the analysis of hydrological extremes," John Wiley & Sons, 21, Issue 10, pp 1265–1279.

[16] Wand, M.P and Jones,M.C. 1995. "Kernel smoothing," Chapman and Hall, London.

[17] Wang,Q.J. 1990. "Estimation of the GEV distribution from censored samples by method of partial probability weighted moments," Journal of Hydrology, 120, pp 103-11.