AN ALTERNATIVE FUNCTIONAL FORM FOR ESTIMATING THE LORENZ CURVE

Nicholas Rohde, School of Economics Discussion Paper No. 383, December 2008, School of Economics, The University of Queensland. Australia.

Abstract

We propose a simple single parameter functional form for the Lorenz curve. The underlying probability density function and cumulative density functions for the Lorenz curve are derived and are shown to have some useful properties. The proposed functional form is fitted to existing data sets and is shown to provide a better fit than existing single parameter Lorenz curves for the given data.

EPrint Type: Departmental Technical Report

Keywords: Gini coefficient; parametric form; Lorenz curve

Subjects:

ID Code: JEL Classification: C80, D31

Deposited By:

Nicholas Rohde
University of Queensland
School of Economics
Brisbane, QLD 4072
n.rohde@.uq.edu.au

AN ALTERNATIVE FUNCTIONAL FORM FOR ESTIMATING THE LORENZ CURVE

Nicholas Rohde
School of Economics
The University of Queensland
Brisbane, Australia
July 2008

Email: n.rohde@uq.edu.au
Telephone: +61 042 229 7225

We propose a simple single parameter functional form for the Lorenz curve. The underlying probability density function and cumulative density functions for the Lorenz curve are derived and are shown to have some useful properties. The proposed functional form is fitted to existing data sets and is shown to provide a better fit than existing single parameter Lorenz curves for the given data.

Author key words: Gini coefficient; Parametric form; Lorenz curve

JEL classification codes: C80, D31

1. INTRODUCTION

The Lorenz curve is an intuitive method for representing the distribution of income. Created by plotting cumulative income shares against cumulative population shares, the Lorenz curve forms the backbone of several inequality measures including the popular Gini coefficient. Lorenz curves may be constructed from grouped data using interpolation techniques (Gastwirth, 1976) or may be presumed to follow a particular parametric form and fit to tabulated data (see Kakwani and Podder (1976), Rasche et al. (1980) Ortega et al. (1991)). Parametric forms such as these are advantaged over Lorenz curves constructed directly from grouped data as they do not assume homogeneity of incomes within subgroups and thus are not downwardly biased (Lerman and Yitzaki, 1989). These techniques however face the disadvantage of imposing a rigid and in some cases unrealistic distribution upon the data and may result in poorly fitting Lorenz curves and inaccurate inequality estimates.

In this paper we propose an alternative single parameter functional form for the Lorenz curve and derive the implicit probability density function (PDF) and cumulative distribution function (CDF). We argue that these functions take appropriate shapes for modeling the distribution of income and demonstrate this by showing that the proposed method provides a better fit than other comparable Lorenz curves for our data set.

2. FUNDAMENTALS OF THE LORENZ CURVE

A Lorenz curve may be defined as

$$\eta = f(\pi) \tag{1}$$

where π is the cumulative population share of persons earning income equal to or below income level *x*.
η is the cumulative income share of population subgroup π.

A Lorenz curve must have the following properties:

$$\frac{d\eta}{d\pi} > 0, \qquad \frac{d^2\eta}{d\pi^2} > 0, \qquad \eta(0) = 0, \qquad \eta(1) = 1$$

and is defined on the domain $0 \le \pi \le 1$

The most popular single parameter Lorenz curves are the forms proposed by Kakwani and Podder (1973), Gupta (1984) Chotikapanich (1993) and a form implied by the Pareto distribution. These are:

Kakwani-Podder: $$\eta(\pi) = \pi e^{-\delta(1-\pi)}, \quad \delta > 0 \tag{2}$$

Gupta: $$\eta(\pi) = \pi a^{\pi-1}, \quad a > 1 \tag{3}$$

Chotikapanich: $$\eta(\pi) = \frac{e^{k\pi} - 1}{e^k - 1}, \quad k > 0 \tag{4}$$

Pareto: $$\eta(\pi) = 1 - (1-\pi)^{\frac{1}{\gamma}}, \quad \gamma > 1 \tag{5}$$

Here we propose the functional form

$$\eta(\pi) = \pi\left(\frac{\beta - 1}{\beta - \pi}\right), \quad \beta > 1 \tag{6}$$

It is simple to verify that (6) passes through the coordinate points (0,0) and (1,1) and that the first and second derivatives are greater than zero. The derivatives are:

$$\eta'(\pi) = \pi\left(\frac{\beta - 1}{(\beta - \pi)^2}\right) + \frac{\beta - 1}{\beta - \pi} \qquad >0 \text{ for } \beta > 1, \quad 0 \le \pi \le 1 \tag{7}$$

$$\eta''(\pi) = 2\pi\left(\frac{\beta - 1}{(\beta - \pi)^3}\right) + 2\frac{(\beta - 1)}{(\beta - \pi)^2} \quad >0 \text{ for } \beta > 1, \quad 0 \le \pi \le 1 \tag{8}$$

Using Kakwani's (1980) result $x(F) = \dfrac{\eta'(\pi)}{\mu}$ we can derive the implicit PDF and CDF for this Lorenz curve.

The probability density function in terms of income (x) with average income μ is:

$$f(x;\alpha) = \frac{\alpha}{2x^2\sqrt{\frac{\alpha}{x}}} \qquad \text{for} \qquad \frac{\alpha}{\beta^2} \le x \le \frac{\alpha}{(\beta - 1)^2} \tag{9}$$

$$\text{where} \qquad \alpha = \beta(\beta - 1)\mu \tag{10}$$

The cumulative distribution function is:

$$F(x) = \beta - \sqrt{\frac{\alpha}{x}} \qquad \text{for} \qquad \frac{\alpha}{\beta^2} \le x \le \frac{\alpha}{(\beta - 1)^2} \tag{11}$$

The PDF of income defined in (9) has some unusual properties. Ostensibly this function of income is manipulated by the single parameter α, which depends on mean income level μ and Lorenz curve parameter β. Such a parameterization appears flawed as different combinations of β and μ can yield the same curve for equation (9). This does not imply that density functions with the same value for parameter α will be identical however, as the PDF is only defined on the domain $(x_{\min} \le x \le x_{\max})$ where the lower and upper bounds depend on both β and μ. As such each combination of β and μ defines a unique PDF for x, which is typically downward sloping over the domain in a manner similar to an exponential decay function.

Lorenz curves such as proposed in (6) that imply probability distribution functions that only exist on a subset of x are not uncommon, with the Chotikapanich, Gupta and Kakwani-Podder functional forms also exhibiting positive lower bounds and finite upper bounds. Such restrictions on x need not be unrealistic as institutional structures such as social welfare systems, minimum wages and high marginal tax rates may effectively constrain incomes to lie within certain bounds.

A further interesting property of the domain of equation (9) is that the distribution mean may be calculated directly from the upper and lower limits. The bounds on x are:

$$x_{\min} = \frac{\alpha}{\beta^2} \qquad\qquad x_{\max} = \frac{\alpha}{(\beta - 1)^2}$$

Solving these expressions with equation (10) gives the result

$$\mu = \sqrt{x_{\min} x_{\max}} \quad (12)$$

-that the mean of the implicit PDF is equal to the geometric average of the highest and lowest incomes available under the distribution.

3. A COMPARISSON OF SINGLE PARAMETER FUNCTIONAL FORMS

In this section we estimate the single parameter Lorenz curves given in Section (2) using decile data from Chotikapanich et al. (2005). The data covers incomes statistics eight Asian countries in 1988 and 1993 and is referred to in an abbreviated form in Table 1. The abbreviations are: Hong Kong-HK, Japan-JP, Korea-KR, Malaysia-ML, Philippines-PH, Singapore-SG, Taiwan-TW and Thailand-TL. Each Lorenz curve is fitted to every data set and we follow Sarabia et al. (2001) by measuring the goodness of fit with the Mean Squared Error. We calculate this as

$$MSE = \sum_{i=1}^{n} \frac{(\eta_i - \eta(\pi_i))^2}{n} \quad (13)$$

where η_i is the cumulative income share of population group *i* calculated from raw data and $\eta(\pi_i)$ is the fitted value of the Lorenz curve at π_i. We exclude the Gupta Lorenz curve from this analysis as it can be shown to be functionally equivalent to the Kakwani-Podder specification and thus gives identical goodness of fit statistics[1]. The results are presented in Table 1, with parameter estimates and goodness of fit statistics given for each Lorenz curve.

Table 1. A comparison of functional forms for the Lorenz curve

Data set	**Kakwani-Podder**		**Chotikapanich**		**Pareto**		**Proposed**	
	δ	**MSE**	***k***	**MSE**	**γ**	**MSE**	**β**	**MSE**
HK88	2.312	0.00196	3.302	0.00165	2.719	0.00068	1.282	**0.00044**
HK93	2.703	0.00270	3.737	0.00233	3.039	**0.00058**	1.230	0.00062
JP88	1.031	0.00027	1.700	0.00021	1.721	0.00060	1.774	**0.00009**
JP93	1.064	0.00036	1.747	0.00029	1.749	0.00052	1.743	**0.00012**
KR88	1.378	0.00068	2.171	0.00056	1.987	0.00067	1.542	**0.00023**
KR93	1.219	0.00025	1.957	0.00020	1.852	0.00101	1.637	**0.00011**
ML88	2.206	0.00138	3.180	0.00112	2.620	0.00096	1.301	**0.00024**
ML93	2.334	0.00148	3.326	0.00121	2.720	0.00101	1.280	**0.00025**
PH88	1.216	0.00057	1.956	0.00046	1.865	0.00047	1.631	**0.00015**
PH93	1.974	0.00115	2.910	0.00092	2.438	0.00090	1.347	**0.00020**
SG88	1.117	0.00033	1.820	0.00025	1.785	0.00060	1.702	**0.00008**
SG93	1.939	0.00098	2.868	0.00077	2.407	0.00104	1.356	**0.00017**
TW88	1.136	0.00045	1.847	0.00036	1.804	0.00051	1.686	**0.00015**

[1] Setting $\delta = \ln \alpha$ allows the Kakwani-Podder specification given in equation (2) to be written in terms of the Gupta specification in equation (3).

TW93	1.149	0.00040	1.864	0.00032	1.811	0.00058	1.678	**0.00012**
TL88	2.023	0.00150	2.968	0.00123	2.486	0.00068	1.334	**0.00031**
TL93	2.298	0.00176	3.286	0.00145	2.700	0.00075	1.285	**0.00032**

The results demonstrate the capability of the proposed functional form to closely model a variety of different data sets. The new Lorenz curve provides the best fit (as measured by MSE) of the four single parameter forms in 15 of the 16 considered cases (indicated on the table in bold) with the Pareto Lorenz curve being slightly superior for Hong Kong data in 1993. The Chotikapanich and Pareto specifications appear roughly equivalent at fitting the given data while the Kakwani-Podder functional form was the poorest performer.

4. CONCLUSION

The proposed functional form appears to be a worthy addition to the existing class of single parameter Lorenz curves. The new specification is shown to meet the required regularity conditions for a Lorenz curve and demonstrates a strong capacity for modeling income data. The ability for this Lorenz curve to effectively model data is likely to be due a similarity of shapes behind the underlying PDF and typical income distributions.

REFERENCES

Chotikapanich, D, 1993 "A Comparison of Alternative Functional Forms for the Lorenz Curve" *Economics Letters* 41, pp. 21-29.

Chotikapanich, D, Griffiths, B, and Rao, D.S, 2007 "Estimating and Combining National Income Distributions Using Limited Data" *Journal of Business and Economic Statistics* 25, pp. 97-109.

Gastwirth, J and Glauberman, M, 1976 "The Interpolation of the Lorenz Curve and Gini Index from Grouped Data" *Econometrica* 44, pp. 479-483.

Gupta, M.R, 1984 "Functional Form for Estimating the Lorenz Curve" *Econometrica* 52, pp. 1313-1314.

Kakwani, N.C, 1980 "Income Inequality and Poverty: Methods of Estimation and Policy Applications" Oxford University Press, 1980.

Kakwani, N.C, and Podder, N, 1973 "On the Estimation of Lorenz Curves from Grouped Observations" *International Economic Review* 14, pp. 278-292.

Kakwani, N.C, and Podder, N, 1976 “Efficient estimation of the Loren curve and associated inequality measures from grouped observations” *Econometrica* 44 pp.137–148.

Lerman, R and Yitzhaki, S, 1989 “Improving the accuracy of estimates of Gini coefficients” *Journal of Econometrics* 42, pp. 43-47.

Ortega, P. Martín, A. Fernández, A. Ladoux, M. and García, A, 1991 “ A new functional form for estimating Lorenz curves” *Review of Income and Wealth* 37, pp. 447-452.

Rasche, R. Gaffney, J. Koo, AYC and Obst, N, 1980 “Functional Forms for Estimating the Lorenz Curve” *Econometrica* 48, pp.1061-1062.

Sarabia, J. M, Castillo, E. and Slottje, D.J, 1999 “An ordered family of Lorenz curves” *Journal of Econometrics* 91, pp. 43–60.