Crawford School of Public Policy

Crawford School working papers

An Alternative Single Parameter Functional Form for Lorenz Curve

Crawford School Working Paper 1712
September 2017

Satya Paul
Amrita University, India

Sriram Shankar
Australian National University

Abstract

This paper proposes a single parameter functional form for the Lorenz curve and compares its performance with the existing single parameter functional forms using Australian income data for 10 years. The proposed parametric functional form performs better than the existing Lorenz functional forms. The Gini based on the proposed functional form is closest to true Gini each year.

Keywords: Gini coefficient; Lorenz curve; Parametric functional form

JEL Classification: C80, D31

Suggested Citation:

Paul, S and Shankar, S. (2017) An Alternative Single Parameter Functional Form for Lorenz Curve, Crawford School working paper 1712, September 2017. Crawford School of Public Policy, The Australian National University.

Address for Correspondence:
Name: Sriram Shankar
Position: Fellow
Address: 2.21 Beryl Rawson Building, ANU Centre for Social Research and Methods, Australian National University
Email: Sriram.Shankar@anu.edu.au

Crawford School of Public Policy
College of Asia and the Pacific
The Australian National University
Canberra ACT 0200 Australia
www.anu.edu.au

An Alternative Single Parameter Functional Form for Lorenz Curve

Satya Paul
Amrita University, India

and

Sriram Shankar[1]
Australian National University

Abstract

This paper proposes a single parameter functional form for the Lorenz curve and compares its performance with the existing single parameter functional forms using Australian income data for 10 years. The proposed parametric functional form performs better than the existing Lorenz functional forms. The Gini based on the proposed functional form is closest to true Gini each year.

Keywords: Gini coefficient; Lorenz curve; Parametric functional form

JEL Classification: C80, D31

[1] Corresponding author: 2.21 Beryl Rawson Building, ANU Centre for Social Research and Methods, Australian National University, email: Sriram.Shankar@anu.edu.au. Contact details for Satya Paul are: Centre for Economics and Governance, Amrita University, Kerala, India. Email: satyapaul@outlook.com.

1. Introduction

The Lorenz curve represents a graphical relationship between the cumulative proportion of population and the cumulative proportion of income and forms the backbone of several inequality measures including the widely used Gini coefficient. *World Institute of Development Economics Research* (WIDER) and *World Bank* publish data on income shares by decile or quintile groups of population for a large number of countries. Based on group data, Lorenz curve can be constructed (i) by interpolation techniques (Gastwirth and Glauberman, 1976), (ii) by assuming a statistical distribution of income and deriving the Lorenz curve (McDonald, 1984), or (iii) by specifying a parametric functional form for the Lorenz curve. The interpolation techniques assume the homogeneity of incomes within sub-groups, thereby leading to a downward bias in Gini estimate. The existing income distribution functions are known to be poorly fitting the Lorenz curve and resulting in inaccurate inequality estimates[2]. The parametric Lorenz functional forms are directly estimated with the group data without assuming homogeneity of incomes within sub-groups and thus are not downwardly biased. Various authors have suggested a variety of parametric functional forms to directly estimate Lorenz curve (Kakwani and Podder, 1973; Aggarwal, 1984; Gupta, 1984; Chotikapanich, 1993; Rohde, 2009). It would be useful and interesting to determine which functional form performs best. This is particularly important when the aim is to construct inequality measures based on Lorenz functional form.

This paper proposes an alternative single parameter Lorenz functional form and compares its performance with the existing single parametric functional forms. The Australian data are used to compare the goodness-of-fit of the widely used parametric Lorenz specifications and the one suggested here

2. Alternative Functional Forms for the Lorenz Curve

If p(x) is the proportion of individuals that receive an income up to x and η is the proportion of total income received by the same units, then the Lorenz curve is defined as

$$\eta = L(p(x)) \tag{1}$$

The regularity conditions for the function L(p) to describe the Lorenz curve are as follows.

$$\text{(i) } L(0) = 0, \text{ (ii) } L(1) = 1, \text{ (iii) } \frac{dL}{dp} \geq 0 \text{ and (iv) } \frac{d^2L}{dp^2} > 0 .$$

[2] Also see Chotikapanich (1993).

Note that (i) and (ii) imply that the Lorenz curve is defined over the domain $0 \le p \le 1$ and (iii) and (iv) suggest that the slope of Lorenz curve is non-negative and monotonically increasing. We propose the following functional form for the Lorenz curve

$$L(p) = p\left[\frac{e^{-\gamma\left(1-e^{p}\right)}-1}{e^{-\gamma(1-e)}-1}\right], \text{ where } \gamma > 0 \tag{2}$$

It is easy to verify that Eq. (2) passes through the coordinate points (0,0) and (1,1) and that the first and second derivatives are greater than zero. That is, L(0) = 0, L(1) =1 and

$$L'(p) = \left[\frac{e^{-\gamma\left(1-e^{p}\right)}-1}{e^{-\gamma(1-e)}-1}\right] + p\gamma\frac{e^{-\gamma\left(1-e^{p}\right)+p}}{e^{-\gamma(1-e)}-1} \ge 0 \tag{3}$$

$$L''(p) = 2\gamma\frac{e^{-\gamma\left(1-e^{p}\right)+p}}{e^{-\gamma(1-e)}-1} + p\gamma\left(\gamma e^{p}+1\right)\frac{e^{-\gamma\left(1-e^{p}\right)+p}}{e^{-\gamma(1-e)}-1} > 0 \text{ for } \gamma > 0, 0 \le p \le 1 \tag{4}$$

This functional form is compared with the existing widely used single parameter functional forms proposed by Kakwani and Podder (1973), Chotikapanich (1993) and Rohde (2009) [3] and a form implied by Pareto distribution.

Pareto: $$L(p) = 1-(1-p)^{\frac{1}{\alpha}}, \alpha > 1 \tag{5}$$

Kakwani-Podder: $$L(p) = pe^{-\delta(1-p)}, \delta > 0 \tag{6}$$

Chotikapanich: $$L(p) = \frac{e^{\kappa p}-1}{e^{\kappa}-1}, \kappa > 0 \tag{7}$$

Rohde: $$L(p) = p\left(\frac{\beta-1}{\beta-p}\right), \beta > 1 \tag{8}$$

The main motivation for fitting a Lorenz curve is to facilitate the estimation of inequality measures such as the Gini coefficient. This widely used index is defined as one minus twice the area under the Lorenz curve. Based on the proposed (new) and existing functional forms, Gini coefficients are expressed as follows:

[3] We do not consider here two other single parameter functional forms given by Aggrawal (1984) and Gupta (1984). As pointed out in Sarabia et al (2010), their functional forms do not add substantive value to the Lorenz curve comparison, as Rohde (2009) is a reparameterization of Aggarwal (1984), and Gupta (1984) that of Kakwani and Podder (1973).

Proposed: $$G = 1 - 2\int_0^1 p\left[\frac{e^{-\gamma\left(1-e^{p}\right)} - 1}{e^{-\gamma(1-e)} - 1}\right]dp \tag{9}$$

Pareto: $$G = 1 - 2\int_0^1 \left[1 - \left(1 - p\right)^{\frac{1}{\alpha}}\right]dp \tag{10}$$

Kakwani-Podder: $$G = 1 - 2\int_0^1 pe^{-\delta(1-p)}dp \tag{11}$$

Chotikapanich: $$\mathrm{G} = \frac{(\kappa-2)\mathrm{e}^{\kappa} + (\kappa+2)}{\kappa(\mathrm{e}^{\kappa}-1)} \tag{12}$$

Rohde: $$G = 2\beta\left[\left(\beta-1\right)\ln\left(\frac{\beta-1}{\beta}\right)+1\right]-1 \tag{13}$$

3. Performance of Alternative Functional Forms of Lorenz Curve

Table 1 presents the Australian income data by decile groups for 2001-2010[4]. Based on these data, the proposed and other four functional forms of Lorenz curve (equations 2 and 5 to 8) are estimated using the non-linear least squares. All the parameter estimates are statistically significant at the 1% level (Table 2). A comparison of these estimated functional forms is done based on two statistics.

(i) Information Inaccuracy Measure (I) = $\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{q}_{\mathrm{i}} \ln(\mathrm{q}_{\mathrm{i}} / \hat{\mathrm{q}}_{\mathrm{i}})$

where q_{i} and $\hat{\mathrm{q}}_{\mathrm{i}}$ denote actual and predicted income shares. The estimated function with smaller value of *I* is better than those with larger values.

(ii) Mean Square Error (MSE) = $\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left[\eta_{\mathrm{i}} - \mathrm{L}\left(\mathrm{p}_{\mathrm{i}}, \hat{\theta}\right)\right]^2$.

It is always non-negative, and values closer to zero are better. Both the statistics are measures of goodness-of-fit.

[4] These group data were constructed using the individual income data from the first ten waves (2001-2010) of Household, Income and Labour Dynamics Australia (HILDA) Survey.

In terms of measure *I*, the proposed functional form performs best followed by Kakwani-Podder, Chotikapanich, Rohde and Pareto respectively in each year (Table 3). In terms of MSE, the proposed functional form performs best in seven years, 2001 to 2006 and 2009, whereas Kakwani-Podder functional form performs best in the other 3 years (Table 4). In 2007 and 2010, the Chotikapanich functional form performs the second best. The worst performance is shown again by the Pareto functional form.

Table 1: Actual Income Shares by Decide Groups, 2001-2010

Decide Groups	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
1	0.0009	0.0009	0.0010	0.0012	0.0013	0.0012	0.0009	0.0011	0.0011	0.0009
2	0.0146	0.0145	0.0135	0.0166	0.0186	0.0193	0.0161	0.0168	0.0164	0.0170
3	0.0364	0.0364	0.0365	0.0386	0.0401	0.0400	0.0389	0.0389	0.0395	0.0387
4	0.0494	0.0505	0.0526	0.0544	0.0547	0.0547	0.0526	0.0528	0.0545	0.0533
5	0.0705	0.0713	0.0733	0.0736	0.0738	0.0746	0.0727	0.0733	0.0765	0.0714
6	0.0944	0.0947	0.0952	0.0956	0.0951	0.0946	0.0932	0.0949	0.0960	0.0943
7	0.1176	0.1177	0.1179	0.1178	0.1155	0.1149	0.1126	0.1147	0.1174	0.1138
8	0.1443	0.1440	0.1450	0.1432	0.1409	0.1393	0.1366	0.1389	0.1406	0.1384
9	0.1804	0.1804	0.1793	0.1782	0.1759	0.1744	0.1717	0.1737	0.1738	0.1729
10	0.2915	0.2896	0.2857	0.2808	0.2841	0.2870	0.3047	0.2949	0.2842	0.2993

Source: These decide groups are constructed using the data on individual income from first 10 waves of HILDA Survey.

Table 2: Estimates of Lorenz Parameters (2001-2010)

	Proposed		Rohde		Chotikapanich		Pareto		Kakwani-Podder	
Year	γ	SE	β	SE	κ	SE	α	SE	δ	SE
2001	0.259	0.040	1.327	0.026	3.096	0.080	2.494	0.238	2.137	0.059
2002	0.244	0.040	1.331	0.027	3.074	0.081	2.480	0.236	2.119	0.059
2003	0.213	0.040	1.338	0.028	3.033	0.085	2.452	0.234	2.084	0.062
2004	0.154	0.042	1.353	0.028	2.947	0.080	2.398	0.220	2.010	0.058
2005	0.142	0.049	1.356	0.027	2.925	0.077	2.390	0.210	1.992	0.057
2006	0.144	0.053	1.355	0.026	2.927	0.078	2.393	0.207	1.993	0.059
2007	0.248	0.062	1.329	0.023	3.069	0.086	2.492	0.217	2.114	0.071
2008	0.203	0.054	1.340	0.025	3.010	0.082	2.448	0.217	2.063	0.064
2009	0.139	0.049	1.356	0.028	2.925	0.084	2.389	0.215	1.992	0.063
2010	0.226	0.057	1.335	0.024	3.040	0.082	2.469	0.217	2.088	0.066

Table 3 Information Inaccuracy Measure

Year	Proposed	Rohde	Chotikapanich	Pareto	Kakwani-Podder
2001	0.0073	0.0373	0.0205	0.1070	0.0172
2002	0.0073	0.0378	0.0209	0.1072	0.0175
2003	0.0074	0.0394	0.0221	0.1087	0.0185
2004	0.0073	0.0367	0.0211	0.1022	0.0178
2005	0.0085	0.0344	0.0209	0.0955	0.0180
2006	0.0096	0.0342	0.0215	0.0932	0.0187
2007	0.0125	0.0355	0.0236	0.0927	0.0211
2008	0.0102	0.0354	0.0221	0.0954	0.0193
2009	0.0092	0.0378	0.0232	0.0994	0.0200
2010	0.0112	0.0352	0.0226	0.0942	0.0200

Table 4 Mean Squared Errors (MSE)

Year	Proposed	Rohde	Chotikapanich	Pareto	Kakwani-Podder
2001	0.00012	0.00084	0.00023	0.00409	0.00017
2002	0.00012	0.00085	0.00023	0.00409	0.00017
2003	0.00012	0.00090	0.00026	0.00416	0.00019
2004	0.00014	0.00081	0.00024	0.00390	0.00018
2005	0.00019	0.00070	0.00022	0.00360	0.00018
2006	0.00022	0.00067	0.00023	0.00349	0.00019
2007	0.00030	0.00065	0.00026	0.00340	0.00025
2008	0.00023	0.00070	0.00025	0.00357	0.00021
2009	0.00018	0.00080	0.00027	0.00376	0.00021
2010	0.00025	0.00067	0.00024	0.00349	0.00022

The true Ginis and the estimated Ginis based on alternative functional forms of Lorenz curve are statistically significant at the 1 % level. (Table 5). For each year, the Gini based on the proposed Lorenz curve specification is closest to true Gini. The Gini based on the Kakwani-Podder functional form is second closest to true Gini. The Ginis based on the Chotikapanich, Rohde and Pareto Lorenz curve functional forms rank respectively third, fourth and fifth in terms of their closeness to true Gini.

Table 5: Estimates of Gini (2001-2010)

Year	TRUE	SE	Proposed	SE	Rohde	SE	Chotika-panich	SE	Pareto	SE	Kakwani-Podder	SE
2001	0.4633	0.0024	0.4576	0.0061	0.4384	0.0084	0.4487	0.0034	0.4276	0.0390	0.4504	0.0075
2002	0.4613	0.0025	0.4553	0.0061	0.4361	0.0084	0.4464	0.0035	0.4253	0.0390	0.4481	0.0077
2003	0.4572	0.0026	0.4506	0.0060	0.4316	0.0087	0.4418	0.0037	0.4206	0.0393	0.4435	0.0081
2004	0.4464	0.0026	0.4416	0.0064	0.4226	0.0082	0.4321	0.0035	0.4115	0.0381	0.4337	0.0078
2005	0.4442	0.0025	0.4397	0.0075	0.4208	0.0077	0.4297	0.0033	0.4100	0.0366	0.4313	0.0078
2006	0.4446	0.0025	0.4400	0.0081	0.4212	0.0075	0.4298	0.0034	0.4106	0.0360	0.4314	0.0081
2007	0.4619	0.0025	0.4559	0.0094	0.4371	0.0074	0.4458	0.0037	0.4272	0.0356	0.4474	0.0091
2008	0.4547	0.0026	0.4491	0.0083	0.4303	0.0077	0.4392	0.0036	0.4199	0.0365	0.4408	0.0084
2009	0.4455	0.0025	0.4393	0.0074	0.4206	0.0082	0.4297	0.0037	0.4099	0.0374	0.4313	0.0085
2010	0.4579	0.0025	0.4525	0.0087	0.4337	0.0075	0.4425	0.0036	0.4235	0.0360	0.4441	0.0086

Note: True Ginis are calculated using the data on individual income from the first 10 waves of HILDA Survey. Standard errors for true Ginis are calculated based on a method provided in Giles (2004). The expressions for the variances for Ginis corresponding to the alternative parametric functional forms of Lorenz curve are presented in Appendix A.

4. Concluding Remarks

The Australian data show the superiority of the proposed Lorenz curve functional form over other functional forms. In terms of information inaccuracy measure, the proposed form outperforms all the four functional forms in all the 10 years, However, in terms of MSE, the proposed Lorenz functional form performs best in 7 years whereas the Kakwani-Podder functional form performs best in the other 3. The Gini based on our Lorenz curve specification is closest to true Gini each year.

Appendix A

The standard error for Gini coefficient is calculated using the asymptotic approximation: $\mathrm{var}(\hat{G}) = \left(\frac{dG}{d\theta}\right)^2 V_\theta$, where V_θ is the asymptotic variance of the single parameter θ estimated using non-linear least squares for the various functional forms of Lorenz curve.

Chotikapanich: $\mathrm{Var}\left(\hat{G}\right) = \left[\frac{2\left(e^{\hat{\kappa}}\left(e^2 - \hat{\kappa}^2 - 2\right)+1\right)}{\left(\hat{\kappa}\left(e^{\hat{\kappa}}-1\right)\right)^2}\right]^2 \mathrm{Var}\left(\hat{\kappa}\right)$.

Rohde: $\mathrm{Var}\left(\hat{G}\right) = \left(\frac{dG}{d\beta}\right)^2 \mathrm{Var}\left(\hat{\beta}\right)$, where $\frac{dG}{d\beta} = 2\left[(2\beta-1)\ln\left(\frac{\beta-1}{\beta}\right)+2\right]$.

Kakwani-Podder: $\mathrm{Var}\left(\hat{G}\right) = \left(\frac{dG}{d\delta}\right)^2 \mathrm{Var}\left(\hat{\delta}\right)$, where $\frac{dG}{d\delta} = 2\int_0^1 p(1-p)e^{-\delta(1-p)}dp$.

Pareto: $\mathrm{Var}\left(\hat{G}\right) = \left(\frac{dG}{d\alpha}\right)^2 \mathrm{Var}\left(\hat{\alpha}\right)$, where $\frac{dG}{d\alpha} = -2\int_0^1 \frac{1}{\alpha^2}\ln(1-p)(1-p)^{\frac{1}{\alpha}}dp$.

Proposed: $\mathrm{Var}\left(\hat{G}\right) = \left(\frac{dG}{d\gamma}\right)^2 \mathrm{Var}\left(\hat{\gamma}\right)$.

where

$$\frac{dG}{d\gamma} = 2\int_0^1\left(-\frac{1-e^p}{1-e^{\gamma\left(1-e^p\right)}} + \frac{1-e}{1-e^{\gamma(1-e)}}\right)\left(p\left[\frac{e^{-\gamma\left(1-e^p\right)}-1}{e^{-\gamma(1-e)}-1}\right]\right)dp.$$

Acknowledgements

This article uses unit record data from the HILDA Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this article, however, are those of the authors and should not be attributed to either

FaHCSIA or the Melbourne Institute.

References

Aggarwal, V., 1984. On optimum aggregation of income distribution data. *Sankhya B,* 46, 343–355.

Chotikapanich, D., 1993. A comparison of alternative functional forms for the Lorenz curve. *Economics Letters*, 41, 21–29.

Gastwirth, J., Glauberman, M., 1976. The interpolation of the Lorenz curve and Gini index from grouped data. *Econometrica,* 44, 479–483.

Giles, D. E. A., 2004. Calculating a standard error for the Gini coefficient: some further results. *Oxford Bulletin of Economics and Statistics*, 66, 425–433.

Gupta, M.R., 1984. Functional form for estimating the Lorenz curve. *Econometrica*, 52, 1313–1314.

Kakwani, N.C., Podder, N., 1973. On the estimation of Lorenz curves from grouped observations. *International Economic Review*, 14, 278–292.

McDonald, J. B., 1984. Some generalized functions for the size distribution of income. *Econometrica*, 52, 647–663.

Rohde, N., 2009. An alternative functional form for estimating the Lorenz curve. *Economics Letters,* 105, 61–63.

Sarabia, J.M., Prieto, F., Sarabia, M., 2010. Revisiting a functional form for the Lorenz curve. *Economics Letters*, 107, 249-252.