

Hybrid models of the Lorenz curve

Tomson Ogowang^{a,*}, U.L. Gouranga Rao^b

^a*Economics Programme, University of Northern British Columbia, 3333 University Way, Prince George,
British Columbia, Canada V2N 4Z9*

^b*Department of Economics, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5*

Received 28 May 1998; accepted 22 December 1999

Abstract

In this paper hybrid Lorenz curves are proposed as a way of circumventing an important drawback of traditional models of the Lorenz curve, namely lack of satisfactory fit over the entire range of a given income distribution. Two categories of hybrid models are identified, namely the additive models and the multiplicative models. Whereas the additive models are obtained by taking convex combinations of the traditional models, the multiplicative models are obtained by taking their weighted products. A comparison of the performances of the hybrid Lorenz curves with those of the constituent Lorenz curves shows that both the additive and multiplicative models perform generally better than the constituent Lorenz curves. © 2000 Elsevier Science S.A. All rights reserved.

Keywords: Additive model; Multiplicative model; Hybrid Lorenz curve; Gini index

JEL classification: C80; D30

1. Introduction

The Lorenz Curve (LC), which plots the cumulative share of total income (the ordinate) against the cumulative proportion of income-receiving units (the abscissa), is a powerful graphical device for analyzing the size distribution of income and wealth. The curve is also used to estimate the Gini index and other measures of inequality and poverty.

A number of parametric models that satisfy the basic properties of a LC have been proposed in the literature. See, for example, Kakwani and Podder (1973, 1976), Rasche et al. (1980), Gupta (1984), Rossi (1985), Arnold (1986), Rao and Tam (1987), Villaseñor and Arnold (1989), Basmann et al. (1990), Ortega et al. (1991), Chotikapanich (1993), Schader and Schmid (1994), Ryu and Slottje

*Corresponding author. Tel.: +1-250-960-6485; fax: +1-250-960-5544.

E-mail address: ogwang@unbc.ca (T. Ogowang).

(1996), Ogwang and Rao (1996), and Sarabia (1997). Recent research (e.g. Rossi, 1985; Basmann et al., 1990; Ryu and Slottje, 1996) has brought to light the fact that certain models of the LC fit certain segments of the observed income distribution better than others. For example, it is not uncommon for a particular model of the LC to fit, say, the first decile of the observed income distribution better but do poorly when fitted to, say, the ninth decile. In view of this problem, it seems reasonable to construct hybrid LCs by combining two or more traditional models of the LC that fit different portions of the observed income distribution well. As will be seen below, such models are obtained either by taking convex combinations of traditional models of the LC (the additive model) or by taking their weighted products (the multiplicative model). A further advantage of hybrid models of the LC is that they have greater flexibility than the constituent models owing to the inclusion of weighting parameters which must also be estimated. The parameters of hybrid LCs, like those of the constituent LCs, can be consistently estimated by nonlinear least squares (NLS).

The format of the rest of the paper is as follows: Section 2 introduces the two types of hybrid models of the LC and provides expressions for the associated Gini indices. A comparison of the performances of the two hybrid LCs with those of the constituent LCs is made in Section 3. The concluding remarks are made in Section 4.

2. Hybrid models of the Lorenz curve

A LC is defined by the ordered points (p, y) where p is the cumulative proportion of the income-receiving units and y is the cumulative share of incomes, if the incomes are arranged in ascending order of magnitude. If $y = f(p)$ is a twice-differentiable function defined in the closed interval $(0, 1)$, then y describes a LC if and only if the following four properties are satisfied: $f(0) = 0$; $f(1) = 1$; $dy/dp \geq 0$ for $0 \leq p \leq 1$; and $d^2y/dp^2 \geq 0$ for $0 \leq p \leq 1$. The properties $f(0) = 0$ and $f(1) = 1$ ensure that the LC passes through the points $(0, 0)$ and $(1, 1)$. The properties $dy/dp \geq 0$ and $d^2y/dp^2 \geq 0$ ensure that the LC has the right curvature, i.e. it should be monotonically increasing and convex towards the p -axis with $0 \leq f(p) \leq p \leq 1$.

If $y_1 = f_1(p)$ and $y_2 = f_2(p)$ describe a LC (i.e. satisfy $f_i(0) = 0$; $f_i(1) = 1$; $df_i(p)/dp \geq 0$ and $d^2f_i(p)/dp^2 \geq 0$ for $i = 1, 2$), then a convex combination $y = f(p) = \delta f_1(p) + (1 - \delta)f_2(p)$ of the two models, where $0 \leq \delta \leq 1$, also describes a LC.¹ To see this, we note that $f(0) = \delta f_1(0) + (1 - \delta)f_2(0) = 0$; $f(1) = \delta f_1(1) + (1 - \delta)f_2(1) = 1$; $dy/dp = \delta df_1(p)/dp + (1 - \delta)df_2(p)/dp \geq 0$; and $d^2y/dp^2 = \delta d^2f_1(p)/dp^2 + (1 - \delta)d^2f_2(p)/dp^2 \geq 0$. Accordingly, a convex combination of any two LCs satisfies all the four properties of a LC. Hereafter, we shall refer to the model $y = f(p) = \delta f_1(p) + (1 - \delta)f_2(p)$, where $0 \leq \delta \leq 1$, as the additive model of the LC. Estimation of the additive model entails estimation of the parameters of the two constituent LCs and the weighting parameter δ .

If $y_1 = f_1(p)$ and $y_2 = f_2(p)$ describe a LC, then their weighted product $y = f(p) = f_1(p)^\gamma f_2(p)^\lambda$ describes a LC certainly when $\gamma \geq 1$ and $\lambda \geq 1$ but also for other positive values of γ and λ depending on the properties of the $f_i(p)$ s. To see this, we note that $f(0) = f_1(0)^\gamma f_2(0)^\lambda = 0$; $f(1) = f_1(1)^\gamma f_2(1)^\lambda = 1$; $dy/dp = \lambda f_1(p)^\gamma f_2(p)^{\lambda-1} df_2(p)/dp + \gamma f_2(p)^\lambda f_1(p)^{\gamma-1} df_1(p)/dp \geq 0$; and $d^2y/dp^2 =$

¹In principle a hybrid LC could combine two or more LCs. However, in this paper, we consider hybrid LCs which are obtained by combining only two LCs.

$$\lambda f_1(p)^\gamma f_2(p)^{\lambda-1} d^2 f_2(p) / dp^2 + \gamma f_2(p)^\lambda f_1(p)^{\gamma-1} d^2 f_1(p) / dp^2 + 2\lambda \gamma f_1(p)^{\gamma-1} f_2(p)^{\lambda-1} df_1(p) / dp \cdot df_2(p) / dp + \lambda(\lambda-1) f_1(p)^\gamma f_2(p)^{\lambda-2} [df_2(p)/dp]^2 + \gamma(\gamma-1) f_2(p)^\lambda f_1(p)^{\gamma-2} [df_1(p)/dp]^2.$$

It is easy to verify that if $\gamma \geq 1$ and $\lambda \geq 1$ then $d^2 y / dp^2 \geq 0$ in which case the weighted product of the two LCs satisfies all the four properties of a LC. For other values of γ and λ , it is also possible that $d^2 y / dp^2 \geq 0$ depending on the properties of $f_1(p)$ and $f_2(p)$.² Hereafter, we shall refer to the model $y = f(p) = f_1(p)^\gamma f_2(p)^\lambda$ as the multiplicative model of the LC provided that the values of γ and λ are such that all the four properties are satisfied. Estimation of the multiplicative model entails estimation of the parameters of the two constituent LCs and the two weighting parameters γ and λ .

As noted above, an important application of LCs is in the estimation of the Gini index, defined as twice the area between the Lorenz curve and the perfect equality line. For a LC model $y = f(p)$, the associated Gini index is given by $G = 2 \int [p - f(p)] dp$. All the integrals in this paper run from 0 to 1. It can be shown that the Gini index associated with the additive hybrid LC model $y = f(p) = \delta f_1(p) + (1 - \delta) f_2(p)$, where $0 \leq \delta \leq 1$, is given by $G = \delta G_1 + (1 - \delta) G_2$, where G_i is the Gini index associated with the LC $y_i = f_i(p)$, $i = 1, 2$. Clearly, the Gini index for the additive hybrid LC is also the same convex combination of the Gini indices of the two constituent LCs. Hence, it lies in between the Gini indices for the two constituent LCs. It can also be shown that the Gini index associated with the multiplicative hybrid LC model $y = f(p) = f_1(p)^\gamma f_2(p)^\lambda$ where $\gamma \geq 1$ and $\lambda \geq 1$, is given by $G = G_2 + 2 \int f_2(p) [1 - f_1(p)^\gamma f_2(p)^{\lambda-1}] dp$, where G_2 is the Gini index associated with the LC $y_2 = f_2(p)$. If $\gamma \geq 1$ and $\lambda \geq 1$, the Gini index for the multiplicative hybrid LC is greater than the Gini index for either of the two constituent LCs. For other values of γ and λ , the Gini index for the associated multiplicative hybrid LC may be greater than, equal to, or even less than the Gini index for either of the two constituent LCs.

3. Comparison of the hybrid LCs with the constituent LCs

For purposes of comparing the performances of the additive and multiplicative hybrid LC models relative to those of the constituent LC models, we used the 1977 US income data reported by Basmann et al. (1990, pp. 87–89). These data, that are divided into 100 income classes corresponding to 99 percentiles, have the advantage of having sufficient observed points on the LC for estimating the parameters of the presumed LC model by NLS.

Owing to space constraints, we have only presented the results for two hybrid LCs, namely the Ortega et al. (1991)–Chotikapanich (1993) model and the Rao-Tam(1987)–Chotikapanich (1993) model. Other LC models we experimented with yielded similar results.³ The parameters of all the models were estimated by NLS and their standard errors computed from heteroscedasticity consistent covariance matrix.

The results for the Ortega–Chotikapanich model are reported in Table 1. Three features of the Table are striking. First, the Ortega et al.–Chotikapanich additive hybrid LC model performs distinctly better than the model proposed by either Chotikapanich or by Ortega et al. by yielding

²In such cases, it is necessary to examine the signs of the first and second derivatives at every point in the sample for violation of monotonicity and convexity conditions. See, for example, Basmann et al. (1990) and Ogwang and Rao (1996).

³Detailed results will be made available by the authors on request.

Table 1

Results for the Ortega et al. –Chotikapanich hybrid LC^a

Model	Estimated parameters						SSE ^f ($\times 10^{-3}$)	MAVE ^g	G ^h
	<i>a</i>	<i>b</i>	<i>k</i>	δ	γ	λ			
Ortega et al. only ^b	0.6532 (0.0032)	0.6953 (0.0015)	–	–			0.2940	0.0066	0.370
Chotikapanich only ^c	–	–	2.3601 (0.0131)	–			12.3967	0.0280	0.361
Hybrid (additive) ^d	0.6315 (0.0159)	0.6921 (0.0056)	2.5364 (0.3519)	0.8713 (0.0097)			0.0492	0.0044	0.368
Hybrid (multiplicative) ^e	3.0539 (1.1213)	3.0309 (0.0987)	1.0624 (0.0950)		0.4756 (0.1576)	0.4411 (0.0419)	7.7980	0.0510	0.369

^a Standard errors in parentheses.^b Based on the model $y = p^a(1 - (1 - p)^b)$. The corresponding expression for the Gini index is $[(a - 1)/(a + 1)] + 2B(1 + a, 1 + b)$, where $B(1 + a, 1 + b)$ denotes the beta function with parameters $(1 + a, 1 + b)$.^c Based on the model $y = [(\exp(kp) - 1)/(\exp(k) - 1)]$. The corresponding expression for the Gini index is $[(k - 2)\exp(k) + (k + 2)]/[k(\exp(k) - 1)]$.^d Based on the model $y = \delta[p^a(1 - (1 - p)^b)] + (1 - \delta)[(\exp(kp) - 1)/(\exp(k) - 1)]$.^e Based on the model $y = \{p^a(1 - (1 - p)^b)\}^\gamma \{(\exp(kp) - 1)/(\exp(k) - 1)\}^\lambda$.^f SSE denotes the sum of squares of model estimation errors.^g MAVE denotes the maximum absolute value of model estimation errors.^h G denotes the Gini index.

smaller sums of squares of model estimation errors (SSE) and smaller maximum absolute values of model estimation errors (MAVE), where the errors are defined as the difference between the actual income shares and the estimated income shares. In this case the additive hybrid LC performs distinctly better than the individual component LCs. Second, the multiplicative hybrid LC performs slightly worse than the two constituent LCs on the basis of MAVE and slightly better than one of the two constituent models on the basis of SSE. Combining the two observations, we conclude that the additive model performs better than its multiplicative counterpart. Third, the Gini indices associated with the individual constituent models and those based on the two versions of the hybrid model are remarkably similar. Basmann et al. (1990) also used the same data to estimate five models of the LC and found that the corresponding estimates of the Gini index, obtained using numerical integration techniques, ranged from 0.36 to 0.39. Ogwang and Rao (1996) estimated a LC, defined as an arc of an optimal circle, using the same data and obtained an estimate of the Gini index of 0.39. The estimates of the Gini index for the same data, obtained by Ryu and Slottje (1996, p. 266) using Bernstein polynomial approximations and exponential polynomial approximations of the LC are all between 0.36 and 0.37, which are remarkably close to our estimates.⁴

The results for the Rao-Tam–Chotikapanich hybrid LCs, reported in Table 2, indicate that both the additive and multiplicative versions of the Rao-Tam Chotikapanich hybrid LC perform better than either of the constituent LCs in terms of smaller SSEs and MAVEs. Another striking feature of Table 2 is that the additive hybrid LC performs better than its multiplicative counterpart. As in the previous

⁴We also checked for monotonicity and convexity of all models by examining the signs of first and second derivatives at every point in the sample and did not encounter any violations.

Table 2

Results for the Rao–Tam–Chotikapanich hybrid LC^a

Model	Estimated parameters						SSE ^f ($\times 10^{-3}$)	MAVE ^g	G^h
	a	b	k	δ	γ	λ			
Rao–Tam only ^b	0.9323 (0.0454)	5.0461 (0.3319)	–	–			13.9458	0.0301	0.361
Chotikapanich only ^c	–	–	2.3601 (0.0131)	–			12.3967	0.0280	0.361
Hybrid (additive) ^d	1.8258 (0.0901)	0.7543 (0.2004)	7.5885 (1.3962)	0.7379 (0.0684)			1.6498	0.0140	0.368
Hybrid (multiplicative) ^e	3.3199 (0.9824)	3.3746 (1.0004)	4.3765 (1.0013)		0.2010 (0.0290)	0.3729 (0.0346)	12.2525	0.0278	0.362

^a Standard errors in parentheses.^b Based on the model $y = p^a b^{p-1}$. The corresponding expression for the Gini index is $1 - [2\exp(\ln b)/(1+a)] {}_1F_1(1+a, 2+a, \ln b)$, where ${}_1F_1$ denotes the confluent hypergeometric function with parameters $(1+a)$, $(2+a)$ and $\ln b$. Note that ${}_1F_1(1+a, 2+a, \ln b) = (1+a) \sum_{k=0}^{\infty} [1/(1+a+k)] [(\ln b)^k / k!]$. However, the series converges so fast that only a few terms of the series expansion are sufficient for accurate estimates of the Gini index to be obtained.^c Based on the model $y = [(\exp(kp) - 1)/(\exp(k) - 1)]$. The corresponding expression for the Gini index is $[(k-2)\exp(k) + (k+2)]/[k(\exp(k) - 1)]$.^d Based on the model $y = \delta[p^a b^{p-1}] + (1-\delta)[(\exp(kp) - 1)/(\exp(k) - 1)]$.^e Based on the model $y = \{p^a b^{p-1}\}^\gamma \{(\exp(kp) - 1)/(\exp(k) - 1)\}^\lambda$.^f SSE denotes the sum of squares of model estimation errors.^g MAVE denotes the maximum absolute value of model estimation errors.^h G denotes the Gini index.

case, the Gini indices are very comparable to those obtained by Basmann et al. (1990), Ryu and Slottje (1996), and Ogwang and Rao (1996).

4. Conclusions

In this paper we have proposed hybrid LCs as a way of circumventing an important drawback of traditional models of the LC, namely lack of satisfactory fit over the entire range of a given income distribution. Two categories of hybrid models are identified, namely the additive models and the multiplicative models. Whereas the additive models are obtained by taking convex combinations of the traditional models, the multiplicative models are obtained by taking their weighted products. A comparison of the performances of the hybrid LCs with those of the constituent LCs shows that both the additive and multiplicative models perform generally better than the constituent LCs.

References

- Arnold, B.C., 1986. A class of hyperbolic Lorenz curves. *Sankhya* 48 Series B, 427–436.
- Basmann, R.L., Hayes, K.J., Slottje, D., Johnson, J.D., 1990. A general functional form for approximating the Lorenz curve. *Journal of Econometrics* 43, 77–90.

- Chotikapanich, D., 1993. A comparison of alternative functional forms for the Lorenz curve. *Economics Letters* 3, 187–192.
- Gupta, M.R., 1984. Functional forms for fitting 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.
- Kakwani, N.C., Podder, N., 1976. Efficient estimation of the Lorenz curve and the associated inequality measures from grouped data. *Econometrica* 44, 137–148.
- Ogowang, T., Rao, U.L.G., 1996. A new functional form for approximating the Lorenz curve. *Economics Letters* 52, 21–29.
- Ortega, P., Martin, G., Fernandez, A., Ladoux, M., Garcia, A., 1991. A new functional form for estimating Lorenz curves. *Review of Income and Wealth* 37, 447–452.
- Rao, U.L.G., Tam, A.Y., 1987. An empirical study of selection and estimation of alternative models of the Lorenz curve. *Journal of Applied Statistics* 14, 275–280.
- Rasche, R.H., Gaffney, J., Koo, A.Y.C., Obst, N., 1980. Functional forms for estimating the Lorenz curve. *Econometrica* 48, 1061–1062.
- Rossi, J.W., 1985. Notes on a new functional form for the Lorenz curve. *Economics Letters* 17, 193–197.
- Ryu, H.K., Slottje, D.J., 1996. Two flexible functional form approaches for approximating the Lorenz curve. *Journal of Econometrics* 72, 251–274.
- Sarabia, J., 1997. A hierarchy of Lorenz curves based on the generalized Tukey's lambda distribution. *Econometric Reviews* 16, 305–320.
- Schader, M., Schmid, F., 1994. Fitting parametric Lorenz curves to grouped income distributions—a critical note. *Empirical Economics* 19, 361–370.
- Villaseñor, J.A., Arnold, B.C., 1989. Elliptical Lorenz curves. *Journal of Econometrics* 40, 327–338.