

Modelling Lorenz curve

Johan Fellman^{1,2}

Abstract

There has been a large number of studies in which the scientists have built models for income distributions. As an alternative, some have built models for Lorenz curves. The step from Lorenz curve to distribution function is more difficult than the step from distribution function to Lorenz curve. There is a difference between advanced and simple Lorenz models. Advanced models with several parameters yield a better fit to data, but are difficult to connect to exact income distributions. Simple one-parameter models can more easily be associated with the corresponding income distribution, but when statistical analyses are performed the goodness of fit is often poor. In this study, simple models are considered and the results compared with results obtained by numerical methods without any assumptions concerning the Lorenz model.

Keywords: Gini coefficients, income distributions, Golden's method, Lagrange polynomials, Simpson's rule, trapezium rule

1 Introduction

There has been a large number of studies in which the scientists have built models for income distributions. As an alternative, some have built models for the Lorenz curve. Among these we may list the following studies: Kakwani and Podder ([1], [2]), Kakwani [3], Rasche et al. [4], Gupta [5], Rao and Tam [6],

¹ Hanken School of Economics, Helsinki, Finland, e-mail: fellman@hanken.fi

² Folkhälsan Institute of Genetics, Helsinki, Finland

Chotikabanich [7], Ogowang and Rao [8], Cheong [9] and Rohde [10]. The step from Lorenz curve to distribution function is more difficult than that from distribution function to Lorenz curve. There is a difference between advanced and simple Lorenz models. Advanced models yield a better fit to data, but are difficult to connect to exact income distributions. Simple one-parameter models can more easily be associated with the corresponding income distribution, but when statistical analyses are performed the goodness of fit is often poor.

2 Methods

In this study, we use the standard notation $L(p)$ for the Lorenz curve. Rao and Tam [6] compared five different models. The first was the Kakwani and Podder (KP) model [1],

$$\eta = a\pi^b (\sqrt{2} - \pi)^c \quad \begin{array}{l} a > 0 \\ 0 \leq b \leq 1, \\ 0 \leq c \leq 1 \end{array} \quad (1)$$

where $\pi = \frac{L+p}{\sqrt{2}}$ and $\eta = \frac{L-p}{\sqrt{2}}$.

The second is the generalised Pareto model (GP) analysed by Rasche et al. [4]

$$L_{GP} = \left(1 - (1-p)^a\right)^{1/b} \quad \begin{array}{l} 0 \leq a \leq 1 \\ 0 \leq b \leq 1 \end{array} \quad (2)$$

and the third is the Gupta (G) model [5]

$$L_G = p\beta^{p-1}. \quad (3)$$

In addition, Rao and Tam constructed a generalized Gupta model (RT)

$$L_{RT} = p^a \beta^{p-1} \quad a > 1. \quad (4)$$

Finally, Rao and Tam introduced a simplified version (S) of the RT model ($\beta = 1$)

$$L_S = p^a \quad a > 1. \quad (5)$$

Chotikabanich [7] defined an alternative Lorenz curve (C):

$$L_C(p) = \frac{e^{kp} - 1}{e^k - 1}, \quad k > 0. \quad (6)$$

The Gupta model and models S and C contain only one parameter. They are so simple that it is impossible to distinguish between the length of the range for the income distribution function and the Gini coefficient. Since with only one parameter to estimate, these distribution properties cannot be independently

estimated, we pay these models special attention and analyse them in more detail.

In general, the step from the Lorenz curve to the income distribution starts from the formula

$$L'(p) = \frac{x_p}{\mu}, \quad (7)$$

where x_p is the p -percentile and μ is the mean of the corresponding distribution $F(x)$. We define $M(\cdot)$ as the inverse function of $L'(\cdot)$. From (7) we obtain

$$p = M\left(\frac{x_p}{\mu}\right). \quad (8)$$

Equation (8) indicates that $M(\cdot)$ is the income distribution function corresponding to the given Lorenz curve, that is, $F(x) = M\left(\frac{x}{\mu}\right)$. This connection

between the Lorenz curve and the distribution function is easily defined, but for most of the exact Lorenz curves it is difficult or even impossible to obtain the income distribution mathematically.

The Gupta model. Examples of Lorenz curves for the Gupta model (3) are given in Figure 1.

Following Gupta, we observe that

$$L'_G(p) = p\beta^{p-1} \log \beta + \beta^{p-1} = \frac{x_p}{\mu}. \quad (9)$$

Consequently,

$$\lim_{p \rightarrow 0^+} x_p = \mu \lim_{p \rightarrow 0^+} L'_G = \frac{\mu}{\beta} \quad (10)$$

and

$$\lim_{p \rightarrow 1^-} x_p = \mu \lim_{p \rightarrow 1^-} L'_G = \mu(1 + \log \beta) \quad (11)$$

From this it follows that Gupta's model corresponds to distributions defined on a finite interval $(\mu\beta^{-1}, \mu(1 + \log \beta))$. In spite of the fact that the Gupta model is relatively simple, the corresponding income distribution is not attainable. The equation (9) cannot be solved exactly with respect to variable p . In the formula, variable p can be found both as a factor and in the exponent.

For the Gupta model, the Gini coefficient is

$$G_G = 1 - 2 \int_0^1 p\beta^{p-1} dp = 1 - \frac{2}{\log \beta} \left(1 - \frac{\beta-1}{\beta \log \beta}\right). \quad (12)$$

Figure 1 shows that the Gini coefficient tends towards 1, when $\beta \rightarrow \infty$.

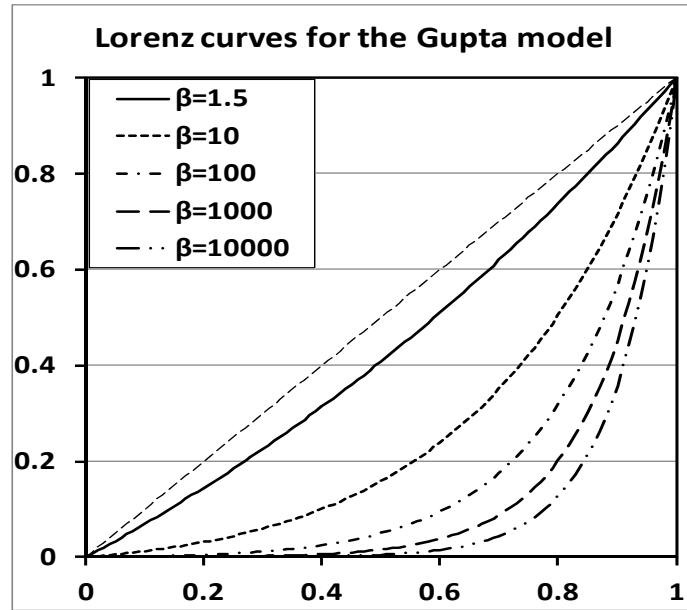


Figure 1: The Lorenz curves for the Gupta model for various β values

Figure 1 shows that the Gini coefficient tends towards 1, when $\beta \rightarrow \infty$.

Remark: In Rao and Tam [6], the formula for the Gini coefficient based on the Gupta model contains a misprint (c.f. our equation (12)), but our numerical check indicates that the authors have used the correct formula.

Following Gupta, the variable $\log \beta$ can be estimated by using the logarithm of the model in (3), that is, from the equation

$$\log\left(\frac{L}{p}\right) = (p-1)\log(\beta).$$

The generalized Gupta model (RT). For the generalized Gupta model, we obtain

$$\frac{x_p}{\mu} = L'_G(p) = p^{\alpha-1} \left(p\beta^{p-1} \log \beta + \alpha\beta^{p-1} \right). \quad (13)$$

The income distribution is defined on the interval $(0, \mu(\alpha + \log \beta))$. It can be observed that if $\beta \rightarrow \infty$ the range of the income distribution then tends

towards $(0, \infty)$ for both the Gupta and the generalized Gupta.

Following Gradsheteyn and Ryshnik [11], Rao and Tam gives the Gini coefficient

$$G_{RT} = 1 - 2e^{-\frac{\log \beta}{1+\alpha}} {}_1F_1(1+\alpha; 2+\alpha; \log \beta)$$

for the generalised Gupta model where ${}_1F_1$ denotes the confluent hyper-geometric function with the parameters indicated in the parentheses.

The simplified RT model (S). The simplified RT model is obtained for $\beta = 1$ and is given in (5), being $L_S(p) = p^\alpha$, $\alpha > 1$. The Lorenz curves for various α values are given in Figure 2.

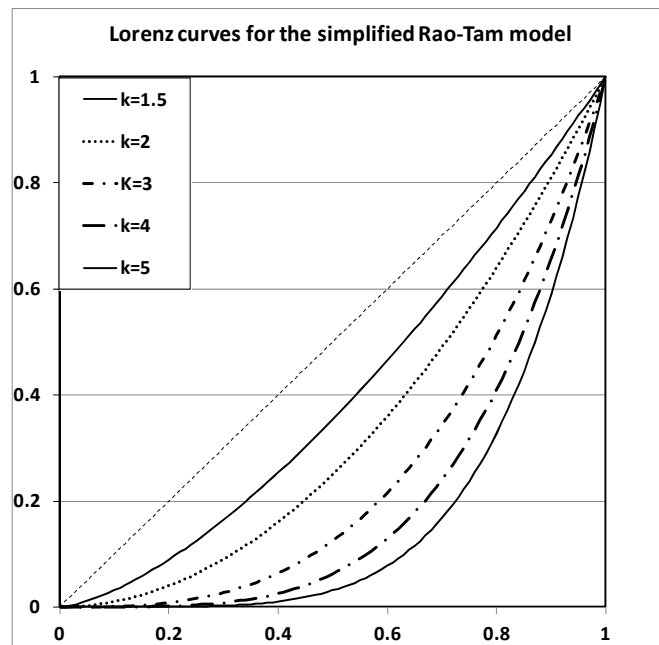


Figure 2: Rao-Tam simplified Lorenz curves

The Gini coefficient is $G_S = \frac{\alpha - 1}{\alpha + 1}$. The income distribution corresponding to the RT model can be found. The derivative of $L_S(p) = p^\alpha$ is $L'_S(p) = \alpha p^{\alpha-1}$. We obtain $\frac{x_p}{\mu} = L'_S(p) = \alpha p^{\alpha-1}$, $p^{\alpha-1} = \frac{x_p}{\alpha \mu}$ and $p = \left(\frac{x_p}{\alpha \mu}\right)^{\frac{1}{\alpha-1}}$. Hence, the income

distribution is $F(x) = \left(\frac{x}{\alpha\mu}\right)^{\frac{1}{\alpha-1}}$ defined on the interval $(0, \alpha\mu)$. Income distributions are given in Figure 3.

The Chotikabanich model. Chotikabanich (1993) introduced the alternative one-parameter Lorenz curve

$$L_C(p) = \frac{e^{kp} - 1}{e^k - 1}, \text{ where } k > 0.$$

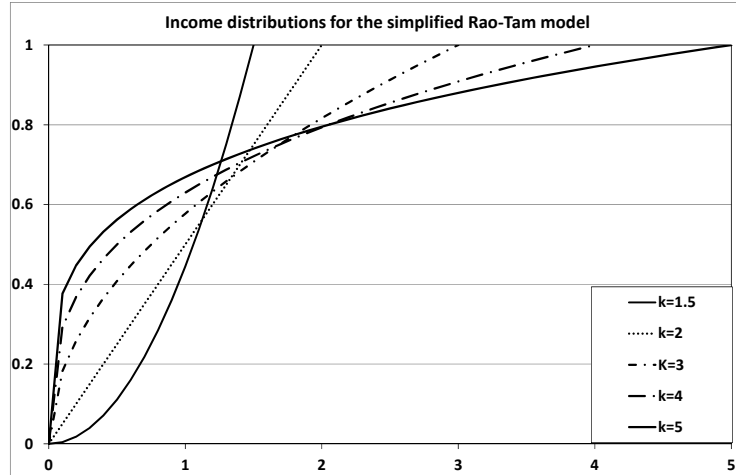


Figure 3: Income distributions corresponding to the Rao-Tam simplified Lorenz curve

It is easily found that

$$L_C(0) = 0,$$

$$L_C(1) = 1,$$

$$\frac{dL_C(p)}{dp} = \frac{ke^{kp}}{e^k - 1} > 0,$$

$$\frac{d^2L_C(p)}{dp^2} = \frac{k^2e^{kp}}{e^k - 1} > 0.$$

The second derivative is positive and hence the Lorenz curve is convex. Consequently, the first derivative is increasing from the minimum

$$\frac{dL_C(0)}{dp} = \frac{k}{e^k - 1} > 0 \text{ to } \frac{dL_C(1)}{dp} = \frac{ke^k}{e^k - 1}.$$

Lorenz curves as functions of parameter k are given in Figure 4.

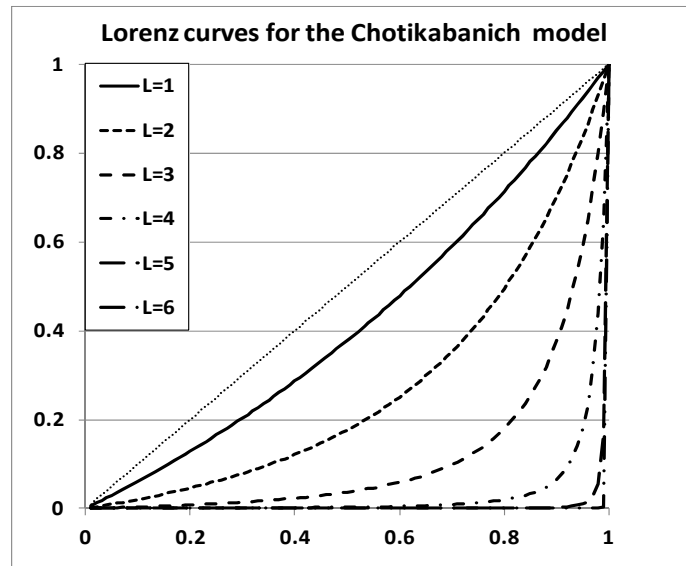


Figure 4: Lorenz curves for the Chotikabanich models

The Gini coefficient is

$$\begin{aligned}
 G_C &= 1 - 2 \int_0^1 L_C(p) dp = 1 - 2 \int_0^1 \frac{e^{kp} - 1}{e^k - 1} dp = 1 - 2 \left(\left(\frac{\frac{1}{k} e^{kp} - 1}{e^k - 1} \right)_{p=1} - \left(\frac{\frac{1}{k} e^{kp} - 1}{e^k - 1} \right)_{p=0} \right) \\
 &= 1 - 2 \left(\left(\frac{\frac{1}{k} e^k - 1}{e^k - 1} \right) - \left(\frac{\frac{1}{k} - 1}{e^k - 1} \right) \right) = 1 - 2 \left(\frac{e^k - 1}{k(e^k - 1)} \right) \\
 &= \frac{k(e^k - 1) - 2e^k + 2}{k(e^k - 1)} = \frac{(k - 2)e^k + k + 2}{k(e^k - 1)}
 \end{aligned}$$

The Gini coefficient increases toward 1 when $k \rightarrow \infty$.

If we consider an income distribution with the mean μ , then income is distributed over the interval $(\frac{\mu k}{e^k - 1}, \frac{\mu k e^k}{e^k - 1})$. When $k \rightarrow \infty$, this interval converges towards $(0, \infty)$.

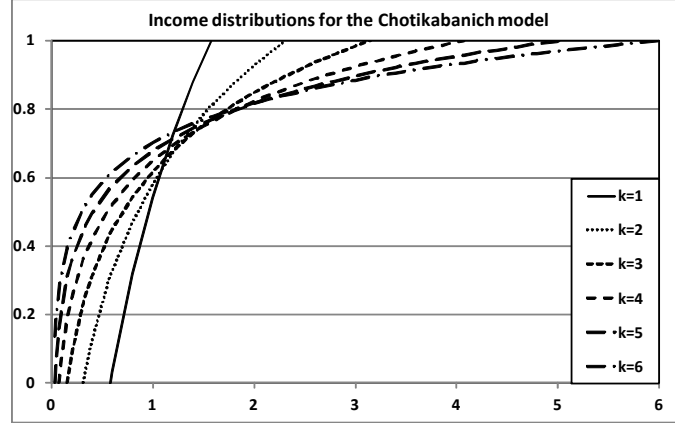


Figure 5: Income distributions for the Chotikabanich models

If we assume an arbitrary μ , then $x_p = \frac{\mu dL_c(p)}{dp} = \frac{\mu k e^{kp}}{e^k - 1}$ and we get

$$\frac{\mu k e^{kp}}{e^k - 1} = x_p.$$

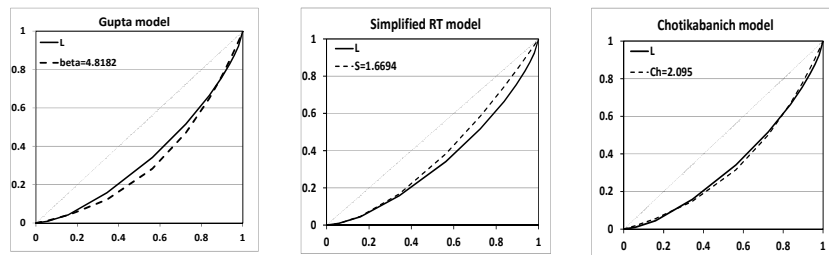
Hence, $F(x) = \frac{1}{k} \log\left(\frac{x(e^k - 1)}{\mu k}\right)$ and the theoretical income distribution is obtained. Figure 5 presents income distributions for various k values.

3 Applications

Kakwani and Podder [2] applied their Lorenz model to Australian data, comparing four alternatives, of which all resulted in accurate estimates. The estimates varied between 0.3195 and 0.3208 when the actual value was 0.3196. Rao and Tam [6] applied the Kakwani-Podder, the generalised Pareto, the RT, the Gupta and the simplified RT models to the same data. Their comparison of the models indicates that the Kakwani-Podder, the generalised Pareto and the RT model yielded the best estimates. The Gupta and the simplified RT models resulted in estimates with the largest errors. For the Gupta model, the estimate was too high (0.3691) and for the simplified RT model it was too low (0.2508). The magnitude of these errors was comparable. These findings support the criticism of the use of simple one-parameter models.

We applied the Chotikabanich model and obtained the following results. We considered $\min_k \sum (f_{obs} - f(k))^2$ and estimated the parameter k . We performed the minimization by using $f = L$ and $f = \log(L)$. We fitted the model to the

Kakwani and Podder data obtained, $k = 0.2095$ and $G = 0.3262$ and $k = 0.2097$ and $G = 0.3263$, respectively. We observe that the one-parameter Chotikabanich model yields slightly better but still less exact results. As a comparison, we present Lorenz models fitted to the Australian data graphically in Figure 6, observing that the Chotikabanich model is closest to the empirical curve. The simplified RT and the Gupta models show larger but comparable discrepancies. These findings support the results obtained by Rao and Tam. In Figure 6, we also observe that Gupta model yields too high an estimate and the simplified model too low an estimate.



Note that the Chotikabanich gives the best fit.

Figure 6: Graphical presentation of the goodness of fit obtained by the Gupta, RT and Chotikabanich models

Fellman [12] studied the numerical estimation of the Gini coefficient based on Lorenz curves. The alternative methods were the trapezium rule, Simpson's rule, a modified version of Golden's method [13] and the Lagrange method. In this study, the Simpson rule cannot be performed because it demands equidistant points. In general, the trapezium rule yields Gini coefficients which are too low. For the Australian data, the result is 0.3134, which is slightly below the correct value. Since the Lagrange method demands an even number of sub-intervals, we had to modify the method slightly. We applied the Lagrange method for the ten last sub-intervals and added a small correction from the first sub-interval. The estimate obtained is 0.3199, a result which is closest to the correct value. Fellman [12] presented a modified version of Golden's method. When we apply this method to the Australian data, we obtain the estimate of 0.3075. This is too low, but still greater than the extremely low value obtained by the simplified RT model. Summing up, our opinion is that one has to choose the Lorenz model with due consideration. This is especially important if the selection should be performed among simple one-parameter models.

ACKNOWLEDGEMENTS. We are grateful to two anonymous referees for comments and suggestions on a previous version of the manuscript. This study was in part supported by a grant from the “Magnus Ehrnrooths Stiftelse” Foundation.

References

- [1] N.C. Kakwani and N. Podder, On the Estimation of Lorenz Curves from Grouped Observations, *International Economic Review*, **14**, (1973), 278-292.
- [2] N.C. Kakwani and N. Podder, Efficient estimation of the Lorenz curve and the associated inequality measures from grouped observations, *Econometrica*, **44**, (1976), 137-148.
- [3] N. Kakwani, On a Class of Poverty Measures, *Econometrica*, **4**, (1980), 437-446.
- [4] R.H. Rasche, J. Gaffney, A.Y.C. Koo and N. Obst, Functional Forms for Estimating the Lorenz Curve, *Econometrica*, **48**, (1980), 1061-1062.
- [5] M.R. Gupta, Functional form for estimating the Lorenz curve, *Econometrica*, **52**, (1984), 1313-1314.
- [6] U.L.G. Rao and A.Y.-P. Tam, An empirical study of selection and estimation of alternative models of the Lorenz curve, *J. of Applied Statistics*, **14**, (1987), 275-280.
- [7] D. Chotikabanich, A comparison of alternative functional forms for the Lorenz curve, *Economics Letters*, **41**, (1993), 129-138.
- [8] T. Ogwang and U.L.G. Rao, Hybrid models of the Lorenz curve, *Economics Letters*, **69**, (2000), 39-44.
- [9] K.S. Cheong, An empirical comparison of alternative functional forms for the Lorenz curve, *Applied Economics Letters*, **9**, (2002), 171-176.
- [10] N. Rohde, An alternative functional form for estimating the Lorenz curve, *Economics Letters*, **105**, (2009), 61-63.
- [11] I.S. Gradsheteyn, and I.M. Ryshnik, *Tables of Integral Series and Products*, New York, Academic Press, 1965.
- [12] J. Fellman, Estimation of Gini coefficients using Lorenz curves. *Journal of Statistical and Econometric Methods*, **1**(2), (2012), 31-38.
- [13] J. Golden, A simple geometric approach to approximating the Gini coefficient, *J. Economic Education*, **39**(1), (2008), 68-77.