Estimating Lorenz Curves Using a Dirichlet Distribution

Duangkamon Chotikapanich

Department of Economics, Curtin University of Technology, Perth, WA 6845 (*chotikapanichd@cbs.curtin.edu.au*)

William E. Griffiths

Department of Economics, University of Melbourne, Vic 3010, Australia (*b.griffiths@unimelb.edu.au*)

Abstract

The Lorenz curve relates the cumulative proportion of income to the cumulative proportion of population. When a particular functional form of the Lorenz curve is specified it is typically estimated by linear or nonlinear least squares, estimation techniques that have good properties when the error terms are independently and normally distributed. Observations on cumulative proportions are clearly neither independent nor normally distributed. This paper proposes and applies a new methodology that recognizes the cumulative proportional nature of the Lorenz curve data by assuming that the income proportions are distributed as a Dirichlet distribution. Five Lorenz-curve specifications are used to demonstrate the technique. Maximum likelihood estimates under the Dirichlet distribution assumption provide better-fitting Lorenz curves than nonlinear least squares and another estimation technique that has appeared in the literature.

Keywords: Gini coefficient; maximum likelihood estimation.

1. INTRODUCTION

The Lorenz curve is one of the most important tools upon which the measurement of income inequality is based. For a given economy or region, it relates the cumulative proportion of income to the cumulative proportion of population, after ordering the population according to increasing level of income. A number of approaches to Lorenz curve estimation have been adopted. In one approach, a particular assumption about the statistical distribution of income is made, the parameters of this income distribution are estimated, and a Lorenz curve consistent with the distributional assumption, and consistent with the parameter estimates for that distribution, is obtained. See, for example, McDonald (1984) and McDonald and Xu (1995). Ryu and Slottje (1996) suggest another approach. They approximate the Lorenz curve from any income distribution by expanding the inverse distribution function in terms of (a) an exponential polynomial series and (b) a sequence of Bernstein polynomial functions. When micro-data are available, nonparameteric estimation of the Lorenz curve and related inequality measures is possible. See, for example, Beach and Davidson (1983), Gastwirth and Gail (1985), and Bishop et al (1989). An alternative approach, more suited to grouped data, is to specify a particular functional form for the Lorenz curve and estimate it directly. It is this approach that is the focus of this paper.

Early breakthroughs on Lorenz curve estimation were those of Gastwirth (1972) and Kakwani and Podder (1973, 1976). Kakwani and Podder recognized the multinomial nature of grouped data and used a Lorenz curve specification that, after transformation, could be placed in an approximate linear model framework.

Other specifications have typically been estimated by linear or nonlinear least squares (Kakwani 1980, Basmann et al 1990, Chotikapanich 1993). Such exercises are useful for fitting Lorenz curves, but, because the covariance matrix estimates they provide are only relevant for independent normally distributed errors, they do not provide a basis for inference about Lorenz curve parameters or any inequality measures derived from them. Clearly, observations on cumulative proportions, or even their logarithms if such a transformation is convenient, will be neither independent nor normally distributed. Sarabia et al (1999) overcome this problem by suggesting a distribution-free method of estimation. Suppose that a Lorenz curve has n unknown parameters, and that M observations on the cumulative proportions are available. They find a set of parameter estimates for each of the $K = \binom{M}{n}$ subsets of n observations. Since each of the subsets yields n equations in n unknown parameters, a set of parameter estimates is obtained by solving these equations. The medians of the sets of parameter estimates are recommended as the final set of estimates. No distribution theory is available for this procedure, but the authors do provide some bootstrap standard errors.

An alternative way to proceed, and the approach adopted in this paper, is to choose a distributional assumption that is consistent with the proportional nature of the data and to pursue maximum likelihood estimation. A suitable distribution is the Dirichlet distribution. It is a multivariate distribution for a vector of random variables that are shares that sum to unity. By relating the parameters of the Dirichlet distribution to Lorenz curve differences, we can accommodate the cumulative proportional nature of the Lorenz curve data, and set up a likelihood function dependent on the unknown parameters of the Lorenz curve. A similar approach was adopted by Woodland (1979) for estimation of share equations that arise in demand and production theory. To further motivate the choice of a Dirichlet distribution, note that, with random sampling, the number of households in each of a number of income classes can be viewed as an observation from the multinomial distribution (Aigner and Goldberger 1970, Kakwani and Podder 1973). Furthermore, by using a transformation from cell numbers to cell proportions, the multinomial distribution can be approximated by a Dirichlet distribution (Johnson 1960, Johnson and Kotz 1969, p.285). Thus, the Dirichlet distribution is a reasonable choice for share data, irrespective of the original income distribution from which the observations were drawn. The choice of a Dirichlet distribution for income shares is much less arbitrary than choosing a specific income distribution. In addition, the number of recognized multivariate distributions that are directly applicable to share data is very limited. Apart from the Dirichlet distribution, only two other possibly-relevant generalized beta distributions are described in Johnson and Kotz (1972). These facts and the general lack of recognition of the share nature of the data in much of the literature on Lorenz curve estimation, make the Dirichlet distribution a useful alternative to pursue.

In Section 2, we outline the distributional assumptions and how they relate to Lorenz curve estimation. The likelihood function for a set of unknown Lorenz curve parameters is derived. To illustrate our suggested techniques we use data on Sweden and Brazil considered earlier by Shorrocks (1983) and revisited by Sarabia et al (1999). These data are described in Section 3; five different Lorenz functions that we use in the empirical work are presented. The results are given and discussed in Section 4. Several questions are investigated. To examine whether the results are sensitive to the chosen estimation technique we compare our estimates and their standard errors to those obtained by Sarabia et al (1999), and those obtained using nonlinear least squares. Since Lorenz-curve estimation is usually a first step towards estimating inequality, maximum likelihood (ML) and nonlinear least squares estimates for the Gini coefficient are obtained for each Lorenz-curve specification. Finally, we examine which estimation technique leads to the best fitting Lorenz curve.

2. MODELS, ASSUMPTIONS AND ESTIMATION

Suppose we have available observations on cumulative proportions of population ($\pi_1, \pi_2, \ldots, \pi_M$ with $\pi_M = 1$) and corresponding cumulative proportions of income ($\eta_1, \eta_2, \ldots, \eta_M$ with $\eta_M = 1$) obtained after ordering population units according to increasing income. We wish to use these observations to estimate a parametric version of a Lorenz curve that we write as $\eta = L(\pi;\beta)$ where β is an $(n \times 1)$ vector of unknown parameters. Clearly, one would not expect all data points to lie exactly on the curve $\eta_i = L(\pi_i;\beta)$. It seems reasonable to assume, however, that conditional on the population proportions π_i, the income shares $q_i = \eta_i - \eta_{i-1}$ are random variables with means

$$E(q_i) \; = \; E(\eta_i) - E(\eta_{i-1}) \; = \; L(\pi_i;\beta) - L(\pi_{i-1};\beta) \qquad (1)$$

Our proposal is to also assume $q = (q_1, q_2, \ldots, q_M)'$ follows a Dirichlet distribution which is a distribution consistent with the share nature of the random vector q. The probability density function (pdf) for the Dirichlet distribution is given by

$$f(q \mid \alpha) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_M)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\ldots\Gamma(\alpha_M)} q_1^{\alpha_1 - 1} q_2^{\alpha_2 - 1} \ldots q_M^{\alpha_M - 1} \quad (2)$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_M)'$ are the parameters of the pdf and $\Gamma(.)$ is the gamma function. By relating the α_i to the Lorenz function, we can find a pdf for q which has the mean given in equation (1) and which is a function of the Lorenz curve parameters. Working in this direction, we set

$$\alpha_i = \lambda[L(\pi_i;\beta) - L(\pi_{i-1};\beta)] \quad (3)$$

where λ is an additional unknown parameter. This definition for α_i gives the desired result because the mean of the Dirichlet distribution is given by

$$\begin{aligned} E(q_i) &= \frac{\alpha_i}{\alpha_1 + \alpha_2 + \cdots + \alpha_M} \\ &= \frac{\lambda[L(\pi_i;\beta) - L(\pi_{i-1};\beta)]}{\lambda \sum_{i=1}^{M} [L(\pi_i;\beta) - L(\pi_{i-1};\beta)]} \\ &= L(\pi_i;\beta) - L(\pi_{i-1};\beta) \quad (4) \end{aligned}$$

since $L(\pi_M;\beta) = 1$ and $L(\pi_0;\beta) = 0$. We can now write the pdf for q as

$$f(q \mid \theta) = \Gamma(\lambda) \prod_{i=1}^{M} \frac{q_i^{\lambda[L(\pi_i;\beta) - L(\pi_{i-1};\beta)] - 1}}{\Gamma(\lambda[L(\pi_i;\beta) - L(\pi_{i-1};\beta)])} \quad (5)$$

where $\theta = (\beta', \lambda)'$.

The variances and covariances between the shares are given by (Johnson and Kotz, 1972, p.231-234)

$$\text{var}(q_i) = \frac{E(q_i)[1 - E(q_i)]}{\lambda + 1} \quad (6)$$

$$\text{cov}(q_i, q_j) = -\frac{E(q_i)E(q_j)}{\lambda + 1} \tag{7}$$

Thus, the income shares are correlated, with correlations given by

$$r_{ij} = -\left(\frac{E(q_i)E(q_j)}{[1 - E(q_i)][1 - E(q_j)]}\right)^{1/2} \tag{8}$$

Since the variances depend on $E(q_i)$, the shares are also heteroskedastic. The parameter λ acts as an inverse variance parameter. The larger the value of λ, the better the fit of the Lorenz curve to the data.

The maximum likelihood estimate for θ can be found by maximizing the log-likelihood function

$$\begin{aligned} \log[f(q \mid \theta)] = \log\Gamma(\lambda) &+ \sum_{i=1}^{M} (\lambda[L(\pi_i;\beta) - L(\pi_{i-1};\beta)] - 1)\log q_i \\ &- \sum_{i=1}^{M} \log\Gamma(\lambda[L(\pi_i;\beta) - L(\pi_{i-1};\beta)]) \end{aligned} \tag{9}$$

3. DATA AND LORENZ CURVES

To illustrate our suggested techniques we use income distribution data on national samples of income recipients for a year close to 1970, for two countries: Sweden and Brazil. These data were used by Sarabia et al (1999). They were derived from Jain (1975) and first published in Shorrocks (1983). The data are in the form of decile cumulative income shares. Shorrocks used the data on these two countries as part of a group of twenty countries to examine the ranking of income distributions given different social states. Sarabia et al (1999) used the data to illustrate their proposed method for the estimation of Lorenz curves. The data on these two countries were chosen because of their differences in the degree of inequality in income distributions.

A large number of functional forms have been suggested in the literature for modelling the Lorenz curve. For details of the various alternatives, see Sarabia et al (1999), and references therein. To keep our study manageable, we chose only 5, ranging from one simple function with only one unknown parameter, to two three-parameter functions which are more flexible, but also harder to estimate precisely. The 5 different Lorenz functions to which we applied the two data sets are:

$$L_1(\pi;k) = \frac{e^{k\pi}-1}{e^k-1} \qquad k>0 \qquad (10)$$

$$L_2(\pi;\alpha,\delta) = \pi^{\alpha}[1-(1-\pi)^{\delta}] \qquad \alpha \geq 0,\ 0<\delta \leq 1 \qquad (11)$$

$$L_3(\pi;\delta,\gamma) = [1-(1-\pi)^{\delta}]^{\gamma} \qquad \gamma \geq 1,\ 0<\delta \leq 1 \qquad (12)$$

$$L_4(\pi;\alpha,\delta,\gamma) = \pi^{\alpha}[1-(1-\pi)^{\delta}]^{\gamma} \qquad \alpha \geq 0, \gamma \geq 1,\ 0<\delta \leq 1 \qquad (13)$$

$$L_5(\pi;a,b,d) = \pi - a\pi^{d}(1-\pi)^{b} \qquad a>0,\ 0<d \leq 1,\ 0<b \leq 1 \qquad (14)$$

The function L_1 is the relatively simple one-parameter function suggested by Chotikapanich (1993); L_2 coincides with the proposal of Ortega et al (1991). L_3 is a well-known form of Lorenz curve suggested by Rasche et al (1980) and L_4 is an extension of L_3 and L_2 introduced by Sarabia et al (1999). Note that L_4 nests both L_2 and L_3, with L_2 being L_4 with $\gamma = 1$ and L_3 being L_4 with $\alpha = 0$. Setting both $\gamma = 1$ and $\alpha = 0$ yields the Lorenz curve $L = 1-(1-\pi)^{\delta}$ which originates from the classical Pareto distribution. The function L_5 is the "beta function" proposed by Kakwani (1980). It is considered one of the best performers among a number of different functional forms for Lorenz curves. See, for example, Datt (1998). Note that, when $a = 1$ and $d = 1$, L_5 is the same as L_2 with $\alpha = 1$.

Once a Lorenz curve has been estimated, one is usually interested in various inequality measures that are related to it. As an example, we compute maximum likelihood estimates for the Gini coefficients that can be derived from each of the Lorenz functions. In each case the Gini coefficient is defined as

$$G = 1 - 2\int_0^1 L(\pi;\beta)\, d\pi \tag{15}$$

Alternative expressions for G can be found for some of the Lorenz curves. However, with the exception of L_1, they still generally involve a numerical integral. We obtain ML estimates by numerically evaluating (15) in each case with β replaced by the ML estimate $\hat{\beta}$.

4. RESULTS

In addition to ML estimation using the assumption of a Dirichlet distribution, we also estimated each function using nonlinear least squares. Because nonlinear least squares has been popular in the literature, it is useful to compare its estimates and standard errors to those from ML estimation. However, conventional nonlinear least squares (NL) standard errors are computed assuming independent identically distributed error terms, an assumption that is unrealistic with share data. Thus, for NL standard errors we report those suggested by Newey and West (1987). The estimates and standard errors obtained by Sarabia et al (1999), for L_2, L_3 and L_4 are also reported; they provide further evidence on the sensitivity of estimates to choice of estimation technique. However, 'Sarabia estimates' for L_1 and L_5 are not available; nor are the standard errors for the 'Sarabia-based' Gini coefficient estimates for all functions.

Point estimates and standard errors of the Lorenz curve parameters and the corresponding Gini coefficients for Sweden are presented in Table 1. With the exception of the function L_4, the estimates of the Lorenz parameters and the Gini coefficient are not sensitive to the estimation technique. Nonlinear least squares, ML and 'Sarabia' lead to almost identical estimates. For L_4 there is considerable variation in the Lorenz parameter estimates, and the Sarabia-estimated Gini coefficient is noticeably different from the others. A somewhat remarkable outcome is that, with the exception of the Sarabia et al estimate from L_4, the point estimates of the Gini coefficient are relatively insensitive to estimation technique *and* functional form specification.

[Table 1 near here]

Although point estimation is robust with respect to choice of estimation technique (and functional form), assessment of the reliability of the estimates, via their standard errors, is heavily dependent on estimation technique. Choosing a maximum likelihood technique that is consistent with the share nature of the data can have a big impact on the perceived precision of the estimates. In Table 1 the standard errors for ML are generally higher than those for nonlinear least squares; those reported by Sarabia et al are higher for some coefficients and lower for others. The standard errors of the Gini coefficient were calculated using the asymptotic approximation

$$\mathrm{var}(\hat{G}) = \frac{\partial G}{\partial \beta'} V_\beta \frac{\partial G}{\partial \beta} \tag{16}$$

where V_β is the asymptotic covariance matrix for the ML or NL estimator for β. Expressions derived using (16) for each of the Lorenz curves are given in the Appendix.

The remarks made about Sweden also hold for the estimates for Brazil given in Table 2, with some minor exceptions. Once again, there are vastly different estimates for L_4, confirming considerable instability in the estimation of this function. In contrast to Sweden, estimates of the L_1 parameter and corresponding Gini coefficient are also sensitive to choice of estimation technique. The other functions remain insensitive to choice of estimation technique. Except for L_1 the Gini coefficient estimates are insensitive with respect to both estimation technique and choice of functional form. Despite yielding similar point estimates, the three estimation techniques yield very different standard errors.

[Table 2 near here]

We turn now to questions of goodness of fit, and choice between alternative Lorenz functions. For a straight goodness-of-fit comparison, we compare values of information inaccuracy (Theil 1967, 1975). For testing nested functional forms we use likelihood ratio tests and the ML estimates.

Let $\hat{q}_i$ denote the predicted income shares obtained from an estimated model. Theil's (1967) measure of information inaccuracy is defined as

$$I = \sum_{i=1}^{M} q_i \log\left(\frac{q_i}{\hat{q}_i}\right) \tag{17}$$

Estimated functions with smaller values of I are better fits than those with larger values. If the q_i are similar to the $\hat{q}_i$, then knowing their values provides little information relative to knowledge of the predictions. The function is a good fit.

On the other hand, q_i quite different from the $\hat{q}_i$ convey considerable information, leading to a large value of I and a poor fit. The information inaccuracy measure was computed using predictions from the nonlinear and ML estimates, and for the Sarabia et al estimates for functions L_2, L_3 and L_4. The outcomes are presented in Table 3.

[Table 3 near here]

For the Swedish data, ML estimation provides a better fit than nonlinear least squares for all functional forms. It also provides better fits than those from the technique suggested by Sarabia et al for the functions they considered. The differences are not great for L_1, L_2 and L_3; they are most noticeable for L_4 and L_5. The large improvement of ML over nonlinear least squares in the case of L_5 is perhaps surprising, given the apparent similarity of the two sets of Lorenz curve estimates. A closer examination of the two sets of predictions for this case revealed that they were not as close as one might suspect by comparing parameter estimates. Also, nonlinear least squares led to some relatively large over predictions that were penalised heavily by the information criterion. Finally, it is interesting that a ranking of the relative magnitudes of the ML standard errors for the Gini coefficient corresponds exactly to a goodness-of-fit ranking of the ML-estimated Lorenz functions.

The information inaccuracies for the Brazilian data lead to the same conclusions with two small modifications. Nonlinear least squares and ML estimation of L_5 had the same fit. Nonlinear least squares provided a better fit than ML for L_1.

To provide information about choice of functional form we examined whether likelihood ratio tests suggested nested versions of L_4 and L_5 would be adequate. The availability of these tests is one of the advantages of the maximum likelihood methodology that we have proposed. Table 4 contains χ^2 values for likelihood ratio tests for various hypotheses. These results suggest that L_3 is an acceptable restricted version of L_4 for both Sweden and Brazil. Also, L_2 is an acceptable restricted version of L_4 for Sweden, but not for Brazil. Finally, a restricted version of L_2, obtained by setting $\alpha = 1$, is clearly rejected relative to the best-fitting L_5.

[Table 4 near here.]

5. CONCLUSIONS AND SUMMARY

One way of estimating a Lorenz curve is to assume a particular distribution for income, estimate the parameters of that distribution, and derive the corresponding Lorenz curve. Another way is to assume a particular Lorenz curve, and estimate its parameters. For this second approach we have suggested a distributional assumption and a corresponding estimation technique which is consistent with the proportional nature of Lorenz-curve data, can be used to approximate share data from any income distribution, and can be employed with any Lorenz-curve specification.

Our model and estimation technique was applied to two data sets that have been the subject of past analyses, one for Sweden, a country with relatively low inequality, and one for Brazil, a country with relatively high inequality. Results were obtained for 5 different Lorenz-curve specifications. Our findings do not necessarily carry over to other data sets and other functions. With this fact kept in mind, we reached the following conclusions. Point estimation of the Gini coefficient was generally insensitive to choice of distributional assumption, estimation technique and Lorenz-curve specification. There were two exceptions to this conclusion. One was for the function L_1 applied to the Brazilian data, using the Dirichlet distribution. The second exception was the estimate from L_4 with the Swedish data and the estimation technique of Sarabia et al. The discrepancy obtained in this case appears to be a consequence of estimation instability associated with this function.

Although point estimation of the Gini coefficient was robust, assessment of the precision of estimation was not. It depended heavily on choice of functional form and choice of estimation technique. With respect to estimation technique, we found that ML estimation, under our proposal to use the Dirichlet distribution, provided the best fit. Useful future work would be a Monte Carlo study to assess whether the standard errors produced by each estimation technique are an accurate reflection of finite-sample variability of the estimates.

APPENDIX: EXPRESSIONS FOR VARIANCES OF THE GINI COEFFICIENT

For L_1:

$$\operatorname{var}(\hat{G}) = \left(\frac{2(e^{\hat{k}}(e^2 - \hat{k}^2 - 2) + 1)}{(\hat{k}(e^{\hat{k}} - 1))^2} \right)^2 \operatorname{var}(\hat{k})$$

For L_2:

$$G = 1 - 2\int_0^1 \pi^{\alpha}[1 - (1-\pi)^{\delta}]\, d\pi$$

$$\operatorname{var}(\hat{G}) = \begin{bmatrix} \frac{\partial G}{\partial \alpha} & \frac{\partial G}{\partial \delta} \end{bmatrix} \begin{bmatrix} \operatorname{var}(\hat{\alpha}) & \operatorname{cov}(\hat{\alpha}, \hat{\delta}) \\ \operatorname{cov}(\hat{\alpha}, \hat{\delta}) & \operatorname{var}(\hat{\delta}) \end{bmatrix} \begin{bmatrix} \frac{\partial G}{\partial \alpha} \\ \frac{\partial G}{\partial \delta} \end{bmatrix}$$

where $\frac{\partial G}{\partial \alpha} = -2\int_0^1 \pi^{\alpha} \log(\pi)[1 - (1-\pi)^{\delta}]\, d\pi$

and $\frac{\partial G}{\partial \delta} = 2\int_0^1 \pi^{\alpha} (1-\pi)^{\delta} \log(1-\pi)\, d\pi$

For L_3:

$$G = 1 - 2\int_0^1 [1 - (1-\pi)^{\delta}]^{\gamma}\, d\pi$$

$$\operatorname{var}(\hat{G}) = \begin{bmatrix} \frac{\partial G}{\partial \delta} & \frac{\partial G}{\partial \gamma} \end{bmatrix} \begin{bmatrix} \operatorname{var}(\hat{\delta}) & \operatorname{cov}(\hat{\delta}, \hat{\gamma}) \\ \operatorname{cov}(\hat{\delta}, \hat{\gamma}) & \operatorname{var}(\hat{\gamma}) \end{bmatrix} \begin{bmatrix} \frac{\partial G}{\partial \delta} \\ \frac{\partial G}{\partial \gamma} \end{bmatrix}$$

where $\frac{\partial G}{\partial \delta} = 2\int_0^1 \gamma[1 - (1-\pi)^{\delta}]^{\gamma-1} (1-\pi)^{\delta} \log(1-\pi)\, d\pi$

and $\frac{\partial G}{\partial \gamma} = -2\int_0^1 [1 - (1-\pi)^{\delta}]^{\gamma} \log[1 - (1-\pi)^{\delta}]\, d\pi$

For L_4 : $$G = 1 - 2\int_0^1 \pi^{\alpha}[1-(1-\pi)^{\delta}]^{\gamma}\, d\pi$$

$$\operatorname{var}(\hat{G}) = \begin{bmatrix} \frac{\partial G}{\partial \alpha} & \frac{\partial G}{\partial \delta} & \frac{\partial G}{\partial \gamma} \end{bmatrix} \begin{bmatrix} \operatorname{var}(\hat{\alpha}) & \operatorname{cov}(\hat{\alpha},\hat{\delta}) & \operatorname{cov}(\hat{\alpha},\hat{\gamma}) \\ \operatorname{cov}(\hat{\alpha},\hat{\delta}) & \operatorname{var}(\hat{\delta}) & \operatorname{cov}(\hat{\delta},\hat{\gamma}) \\ \operatorname{cov}(\hat{\alpha},\hat{\gamma}) & \operatorname{cov}(\hat{\delta},\hat{\gamma}) & \operatorname{var}(\hat{\gamma}) \end{bmatrix} \begin{bmatrix} \frac{\partial G}{\partial \alpha} \\ \frac{\partial G}{\partial \delta} \\ \frac{\partial G}{\partial \gamma} \end{bmatrix}$$

where $$\frac{\partial G}{\partial \alpha} = -2\int_0^1 \pi^{\alpha} \log(\pi)[1-(1-\pi)^{\delta}]^{\gamma}\, d\pi$$

$$\frac{\partial G}{\partial \gamma} = -2\int_0^1 \pi^{\alpha} [1-(1-\pi)^{\delta}]^{\gamma} \log[1-(1-\pi)^{\delta}]\, d\pi$$

$$\frac{\partial G}{\partial \delta} = 2\int_0^1 \pi^{\alpha}\gamma[1-(1-\pi)^{\delta}]^{\gamma-1} (1-\pi)^{\delta} \log(1-\pi)]\, d\pi$$

For L_5 : $$G = 1 - 2\int_0^1 [\pi - a\pi^{d}(1-\pi)^{b}]\, d\pi$$

$$\operatorname{var}(\hat{G}) = \begin{bmatrix} \frac{\partial G}{\partial a} & \frac{\partial G}{\partial d} & \frac{\partial G}{\partial b} \end{bmatrix} \begin{bmatrix} \operatorname{var}(\hat{a}) & \operatorname{cov}(\hat{a},\hat{d}) & \operatorname{cov}(\hat{a},\hat{b}) \\ \operatorname{cov}(\hat{a},\hat{d}) & \operatorname{var}(\hat{d}) & \operatorname{cov}(\hat{d},\hat{b}) \\ \operatorname{cov}(\hat{a},\hat{b}) & \operatorname{cov}(\hat{d},\hat{b}) & \operatorname{var}(\hat{b}) \end{bmatrix} \begin{bmatrix} \frac{\partial G}{\partial a} \\ \frac{\partial G}{\partial d} \\ \frac{\partial G}{\partial b} \end{bmatrix}$$

where $$\frac{\partial G}{\partial a} = 2\int_0^1 \pi^{d}(1-\pi)^{b}\, d\pi$$

$$\frac{\partial G}{\partial d} = 2\int_0^1 a\pi^{d}(1-\pi)^{b} \log(\pi)\, d\pi$$

$$\frac{\partial G}{\partial b} = 2\int_0^1 a\pi^{d}(1-\pi)^{b} \log(1-\pi)\, d\pi$$

REFERENCES

Aigner, D.J. and Goldberger, A.S. (1970), "Estimation of Pareto's Law from Grouped Observations", *Journal of American Statistical Association*, 65, 712-723.

Basmann, R.L., Hayes, K.J., Slottje, D.J. and Johnson J.D. (1990), "A General Functional Form for Approximating the Lorenz Curve", *Journal of Econometrics*, 43, 77-90.

Beach, C.M. and Davidson, R. (1983), "Distribution-Free Statistical Inference with Lorenz Curves and Income Shares", *Review of Economic Studies*, 50, 723-735.

Bishop, J.A., Chakraborti, S. and Thistle, P.D. (1989), "Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves", *Review of Economics and Statistics*, 71, 725-727.

Chotikapanich, D. (1993), "A Comparison of Alternative Functional Forms for the Lorenz Curve", *Economics Letters*, 41, 129-138.

Datt, G. (1998), "Computational Tools for Poverty Measurement and Analysis", *FCND Discussion Paper No. 50*, International Food Policy Research Institute, World Bank.

Gastwirth, J.L. (1972), "The Estimation of the Lorenz Curve and Gini Index", *Review of Economics and Statistics*, 54, 306-316.

Gastwirth, J.L. and Gail, M.H. (1985), "Simple Asymptotically Distribution-Free Methods Comparing Lorenz Curves and Gini Indices Obtained from Complete Data", in R. L. Basmann and G.F. Rhodes, Jr., editors, *Advances in Econometrics*, 4, JAI Press, Greenwich, CT.

Jain, S. (1975), *Size Distribution of Income*, World Bank, Washington.

Johnson, N.L. (1960), "An Approximation to the Multinomial Distribution; Some Properties and Applications", *Biometrika*, 47, 93-102.

Johnson, N.L. and Kotz, S. (1969), *Discrete Distributions*, John Wiley & Sons, New York.

Johnson, N.L. and Kotz, S. (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley & Sons, New York.

Kakwani, N.C. (1980), "On a Class of Poverty Measures", *Econometrica*, 48, 437-446.

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

Kakwani, N.C. and Podder, N. (1976), "Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations", *Econometrica*, 44, 137-148.

McDonald, J. B. (1984), "Some Generalized Functions for the Size Distribution of Income", *Econometrica*, 52, 647-663.

McDonald J.B. and Xu, Y.J. (1995), "A Generalization of the Beta Distribution with Applications", *Journal of Econometrics*, 66, 133-152.

Newey, W. and West, K. (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.

Ortega, P., Fernandez, M.A., Lodoux, M. and Garcia, A. (1991), "A New Funational Form for Estimating Lorenz Curves", *Review of Income and Wealth*, 37, 447-452.

Rasche, R.H., Gaffney, J., Koo, A. and Obst, N. (1980), "Functional Forms for Estimating the Lorenz Curve", *Econometrica*, 48, 1061-1062.

Ryu, H.K. and Slottje, D.J. (1996), "Two Flexible Functional Form Approaches for Approximating the Lorenz Curve, *Journal of Econometrics*, 72, 251-274.

Sarabia, J-M, Castillo, E. and Slottje, D.J. (1999), "An Ordered Family of Lorenz Curves", *Journal of Econometrics*, 91, 43-60.

Shorrocks, A.F. (1983), "Ranking Income Distributions", *Economica*, 50, 3-17.

Theil, H. (1967), *Economics and Information Theory*, Amsterdam, North Holland.

Theil, H. (1975), *Theory and Measurement of Consumer Demand*, Amsterdam, North Holland.

Woodland, A.D. (1979), "Stochastic Specification and the Estimation of Share Equations", *Journal of Econometrics*, 10, 361-383.

Table 1
Estimates and Standard Errors for Lorenz Parameters and Gini Coefficients
Sweden

		α	δ	γ	Gini
L_2					
	NL	0.5954 (0.0136)	0.6352 (0.0052)		0.3880 (0.0013)
	ML	0.6068 (0.0206)	0.6412 (0.0085)		0.3872 (0.0041)
	Sarabia	0.5960 (0.0018)	0.6400 (0.0303)		0.3850
L_3					
	NL		0.7269 (0.0032)	1.5602 (0.0076)	0.3871 (0.0007)
	ML		0.7335 (0.0072)	1.5767 (0.0176)	0.3877 (0.0036)
	Sarabia		0.7300 (0.0263)	1.5620 (0.0022)	0.3860
L_4					
	NL	-0.7552 (0.5638)	0.7931 (0.0366)	2.2893 (0.5458)	0.3864 (0.00004)
	ML	0.0048 (0.6612)	0.7330 (0.0756)	1.5721 (0.6369)	0.3876 (0.0036)
	Sarabia	0.0769 (0.0003)	0.6490 (0.0977)	1.1740 (0.0002)	0.3210

L_1		k	Gini
	NL	2.5029 (0.0826)	0.3792 (0.0292)
	ML	2.5313 (0.1831)	0.3828 (0.0228)

L_5		a	d	b	Gini
	NL	0.7664 (0.0148)	0.9397 (0.0138)	0.5929 (0.0108)	0.3876 (0.0010)
	ML	0.7492 (0.0143)	0.9199 (0.0093)	0.5862 (0.0109)	0.3870 (0.0031)

Table 2
Estimates and Standard Errors for Lorenz Parameters and Gini Coefficients
Brazil

		α	δ	γ	Gini
L_2					
	NL	0.5727	0.2876		0.6361
		(0.0223)	(0.0019)		(0.0012)
	ML	0.5270	0.2857		0.6326
		(0.0383)	(0.0053)		(0.0052)
	Sarabia	0.4900	0.2780		0.6350
		(0.0038)	(0.0662)		
L_3					
	NL		0.3782	1.4357	0.6328
			(0.0038)	(0.0127)	(0.0010)
	ML		0.3721	1.4160	0.6325
			(0.0068)	(0.0225)	(0.0040)
	Sarabia		0.3640	1.3960	0.6340
			(0.0713)	(0.0004)	
L_4					
	NL	0.2169	0.3467	1.2674	0.6339
		(0.1950)	(0.0289)	(0.1473)	(0.0013)
	ML	0.0262	0.3683	1.3950	0.6325
		(0.2148)	(0.0318)	(0.1734)	(0.0039)
	Sarabia	0.0770	0.6170	1.1740	0.6440
		(0.0001)	(0.1041)	(0.0091)	

L_1		*k*	Gini
	NL	5.3685	0.6368
		(0.6726)	(0.1647)
	ML	3.8438	0.5234
		(0.8237)	(0.0747)

L_5		*a*	*d*	*b*	Gini
	NL	0.9151	1.0001	0.2698	0.6349
		(0.0030)	(0.0024)	(0.0016)	(0.0003)
	ML	0.9131	0.9990	0.2685	0.6349
		(0.0044)	(0.0024)	(0.0021)	(0.0013)

Table 3
Information Inaccuracy Measure

	Sweden			Brazil		
	ML	NL	Sarabia	ML	NL	Sarabia
L_1	0.00888	0.00892		0.10851	0.08791	
L_2	0.00029	0.00031	0.00030	0.00056	0.00067	0.00070
L_3	0.00025	0.00027	0.00026	0.00031	0.00034	0.00035
L_4	0.00025	0.00029	0.01259	0.00031	0.00038	0.09710
L_5	0.00017	0.00032		0.00003	0.00003	

Table 4
The Likelihood Ratio Test

	Sweden	Brazil	Critical Value
L_4 VS L_2	1.351	5.333	3.841
L_4 VS L_3	0.000	0.015	3.841
L_5 VS L_2 (with $\alpha = 1$)	36.907	31.355	5.991