Journal of Econometrics 40 (1989) 327–338. North-Holland

ELLIPTICAL LORENZ CURVES

José A. VILLASEÑOR

Colegio de Postgraduados, Chapingo, Mexico

Barry C. ARNOLD

University of California, Riverside, CA 92521, USA

Received May 1986, final version received July 1988

A family of elliptic Lorenz curves is proposed for fitting grouped income data. The associated distribution and density functions are displayed together with the Gini indices. Estimation procedures are discussed. Comparisons are made with alternative models using Australian 1967–68 income data.

1. Introduction

The Lorenz curve (LC) provides a convenient summarization of data in the study of income distributions. It graphically displays the relation between the cumulative proportion of income units and the cumulative proportion of income received when the units are arranged in the order of increasing income. Several models have been proposed for income size distributions. See Arnold (1983) and Macdonald (1984) for surveys. An alternative approach, proposed by Kakwani and Podder (1973), is to consider parametric families of LC's which will fit sample LC's arising from actual data reasonably well. A variety of parametric families of LC's have been subsequently considered. The list includes classical Pareto LC's, the aforementioned proposal of Kakwani and Podder (1973, 1976) and models due to Rasche et al. (1980) and Gupta (1984). Pakes (1981) discusses and compares many of these models and, in addition, proposes an incomplete beta LC.

Very few of the models allow for explicit expression of both the LC and the density of the corresponding size distribution. In addition, even when the sample LC's and the fitted LC's are in close agreement, serious shortcomings are observable in the fit of the corresponding densities.

The present research began with the observation that the Lorenz curve for Mexican income in 1977 was remarkably similar to a quarter segment of a circle of radius 1. Preliminary investigation using several well-known data sets indicated that segments of ellipses frequently fit data surprisingly well.

No economic motivation for the model was or is available but the fact remains: segments of ellipses provide a flexible family of Lorenz curves whose corresponding unimodal densities are readily described and which perform remarkably well in fitting data.

General properties of LC's are reviewed in section 2. The class of general quadratic Lorenz curves is introduced in section 3 and the elliptic subclass is identified. Estimation procedures are discussed in section 4. The last section treats an example first discussed by Kakwani and Podder (1973) and then by Pakes (1981) using data from the Australian Survey of Consumer Expenditure and Finances (ASCEF) 1967–68.

2. Notation and previous results

Here we consider the Lorenz curve as defined by Gastwirth (1971). To any distribution function $F(y)$ with positive finite expectation μ and whose support is a subset of the non-negative real numbers R^+, we associate a Lorenz curve $L(x)$ defined by

$$L(x) = \int_0^x F^{-1}(y)\,\mathrm{d}y/\mu, \qquad x \in [0,1], \tag{1}$$

where

$$F^{-1}(y) = \sup\{z\colon F(z) \le y\}.$$

A characterization of a Lorenz curve attributed to Gaffney and Anstis by Pakes (1981) is the following:

Theorem 1. Suppose $L(x)$ is defined and continuous on $[0,1]$ *with second derivative $L''(x)$. The function $L(x)$ is a Lorenz curve if and only if*

$$L(0)=0, \quad L(1)=1, \quad L'(0^+) \ge 0, \quad L''(x) \ge 0 \quad \text{in} \quad (0,1). \tag{2}$$

Concerning the density $f(x)$ associated with a Lorenz curve $L(x)$ we have the following:

Theorem 2. If $L''(x)$ exists and is positive everywhere in an interval (x_1, x_2), then F has a finite positive density in the interval $(\mu L'(x_1^+), \mu L'(x_2^-))$ which is given by

$$f(x) = 1/\mu L''(F(x)). \tag{3}$$

3. Quadratic Lorenz curves

The general quadratic form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{4}$$

includes many curves passing through (0,0) and (1,1) which satisfy condition (2) and hence may be considered to be Lorenz curves. The subfamily of curves (4) satisfying (2) may be called the class of general quadratic Lorenz curves. The curve is a segment of an ellipse, a parabola or a hyperbola according as $b^2 - 4ac$ is < 0, $= 0$ or > 0. If (4) is to pass through (0,0) and (1,1) we must have

$$f = 0, \tag{5a}$$

and

$$e = -(a + b + c + d). \tag{5b}$$

If $c = 0$, the resulting LC is necessarily a hyperbola. This class of LC's has been investigated by Aggarwal and Singh (1984) [see also Arnold (1986)]. If $c \neq 0$, we may standardize and assume $c = 1$. The quadratic Lorenz curves, with $c = 1$, will satisfy (4) and have $f = 0$ and $e = -(a + b + d + 1)$. Two classes of curves can be identified by solving for y in (4), namely,

$$y = L_1(x) = \left\{-(bx + e) + \left(\alpha x^2 + \beta x + e^2\right)^{1/2}\right\}/2, \tag{6a}$$

and

$$y = L_2(x) = \left\{-(bx + e) - \left(\alpha x^2 + \beta x + e^2\right)^{1/2}\right\}/2, \tag{6b}$$

where $e = -(a + b + d + 1)$, $\alpha = b^2 - 4a$ and $\beta = 2be - 4d$. The density functions corresponding to the first class turn out to be either monotone on their support or bathtub shaped with local maxima at both end points. Their appropriateness for fitting income distributions is thus questionable and it is reasonable to focus attention only on the second class. The class (6b) contains hyperbolic LC's (when $\alpha > 0$), elliptical LC's (when $\alpha < 0$) and parabolic LC's (when $\alpha = 0$).

It is possible though tedious to enumerate completely the conditions that a, b and d must satisfy to ensure that the Lorenz curve (6b) is a true Lorenz curve, i.e., satisfies (2). The reader is referred to Villaseñor and Arnold (1984a) for details. In the present work we will focus on elliptical LC's. For them the necessary and sufficient conditions are relatively simply described.

Lemma 1. In order that (6b) should represent an elliptical Lorenz curve it is necessary and sufficient that $\alpha = b^2 - 4a < 0$, $a+b+d+1>0$, $d \geq 0$ *and* $a+d-1 \leq 0$.

Lorenz curves belonging to the family (6b) are twice differentiable. If we let F_2 denote the distribution function associated with L_2 then, using (1), $L_2'(x) = F_2^{-1}(x)/\mu$ (where μ is the mean of F_2). From this F_2 may be determined. It may be differentiated to yield the density f_2 whose support will be the interval $(\mu L'(0+), \mu L'(1-))$ using Theorem 2. In this manner we may verify the following:

Theorem 3. If L_2 *is an elliptical Lorenz curve, then its associated distribution function is absolutely continuous with density*

$$f_2(y) = (2\alpha K/\mu)\left\{[b+2(y/\mu)]^2 - \alpha\right\}^{3/2}, \tag{7}$$

on the interval

$$(\mu d/(a+b+d+1),\ \mu(2a+b+d)/(a+d-1)). \tag{8}$$

where

$$\mu = \int_0^\infty y f_2(y)\,\mathrm{d}y \quad (\textit{the mean of } f_2),$$

$$K = (\beta^2 - 4\alpha e^2)^{1/2}/2\alpha,$$

$$\alpha = b^2 - 4a,$$

$$\beta = 2be - 4d,$$

$$e = -(a+b+d+1),$$

in which the parameters a, b *and* d *satisfy the conditions of Lemma 1.*

Note that in (7), the lower end point of the support can be 0 (when $d=0$) and the upper end point can be ∞ (when $a+d=1$). The support will be $(0,\infty)$ when $a=1$ and $d=0$. The parameterization using a, b and d subject to constraints of Lemma 1 is natural if we think of describing the LC's but is eminently inconvenient for describing the densities in Theorem 3. An alternative parameterization is available. Observe that rescaling and translation will transform the density (7) into one of the form

$$f_0(z) \propto (1+z^2)^{-3/2}, \qquad \eta_1 < z < \eta_2. \tag{9}$$

It is not difficult to verify that densities of the form (9) where $0 \leq \eta_1 < \eta_2 < \infty$ do indeed have elliptical LC's. If Z has density (9), then $X = \tau Z + \nu$ will also have an elliptical Lorenz curve for any $\tau > 0$ and any μ such that $\tau\eta_1 + \nu > 0$ (to ensure that X is non-negative and has a well defined LC). We are thus led to the following alternative specification of the class of densities with elliptical LC's.

Theorem 4. The class of distributions with elliptical Lorenz curves consists of all absolutely continuous distributions whose densities are of the form

$$f(x) \propto \left[1 + [(x-\nu)/\tau]^2\right]^{-3/2}, \qquad \tau\eta_1 + \nu < x < \tau\eta_2 + \nu, \tag{10}$$

where $0 \leq \eta_1 < \eta_2 < \infty$, $\tau > 0$ *and* ν *is such that* $\tau\eta_1 + \nu \geq 0$.

The parameters $(\tau, \nu, \eta_1, \eta_2)$ of (10) are related to the parameters a, b, d of the Lorenz curve (6b) by

$$\begin{aligned}
\nu &= -b\mu/2,\\
\tau &= \mu(4a-b^2)^{1/2}/2,\\
\eta_1 &= 2[d/(a+b+d+1) + b/2](4a-b^2)^{-1/2},\\
\eta_2 &= 2[(2a+b+d)/(a+d-1) + b/a](4a-b^2)^{-1/2}.
\end{aligned} \tag{11}$$

From the representation (10) it is evident that the density is either non-decreasing or, if it has a mode, its mode is located at $\nu(= -b\mu/2)$.

One measure of inequality which is widely used among economists is the Gini index. It is defined by

$$G_L = 2\int_0^1 \{x - L(x)\}\, \mathrm{d}x, \tag{12}$$

for a given LC $L(x)$. For the Lorenz curves $L_2(x)$ whose densities are given by (7) [or (10)] we have

$$\begin{aligned}
G_2 &= 2\int_0^1 \{x - L_2(x)\}\, \mathrm{d}x\\
&= 1 + (b+2e)/2 + \{(2\alpha+\beta)(a+d-1) + \beta e\}/4\alpha\\
&\quad + \{(\beta^2 - 4\alpha e^2)/8\alpha\sqrt{-\alpha}\}\Big\{\sin^{-1}\Big([2\alpha+\beta]/[\beta^2-4\alpha e^2]^{1/2}\Big)\\
&\quad - \sin^{-1}\Big(\beta/[\beta^2-4\alpha e^2]^{1/2}\Big)\Big\}.
\end{aligned} \tag{13}$$

4. Parameter estimation for quadratic Lorenz curves

From eqs. (4) and (5) with $c = 1$ we have

$$ax^2 + bxy + y^2 + dx + ey = 0. \tag{14}$$

Eq. (14) is linear in the parameters a, b, d, since $e = -(a+b+d+1)$. In fact, (14) is equivalent to

$$y(1-y) = a(x^2 - y) + by(x-1) + d(x-y). \tag{15}$$

If we let $t = y(1-y)$, $u = x^2 - y$, $v = y(x-1)$ and $w = x - y$, then t is a linear function through the origin of u, v and w. Hence given pairs of observations $(x_i, y_i) \in (0,1) \times (0,1)$ $(i = 1, 2, \ldots, m)$ of cumulative proportions of income units and cumulative proportion of income received, a linear regression through the origin of the t_i's onto the u_i's, v_i's and w_i's produces least-squares-like estimates for a, b and d.

If we impose the restrictions $d = 0$ and/or $a + d = 1$, eq. (14) remains linear in the remaining parameters. Hence an equation analogous to eq. (15) can be obtained and then the same estimation procedure outlined above can be used.

In order to estimate μ, we might use the sample mean income as a natural estimate if it is available. An alternative estimate of μ would be the value of μ which minimizes the sum of squared differences $\hat{F}(u_i) - x_i$, where the u_i's are the upper limits of income classes, the x_i's are the cumulative proportion of income units and $\hat{F}(\cdot)$ is $F_2(x)$ corresponding to the density (7) in which parameters a, b, d are replaced by their estimates. This minimization problem does not seem to have a simple solution. A related and more tractable criterion is to minimize the sum of squared differences $\sum_{i=1}^{m-1}(u_i - \hat{F}^{-1}(x_i))^2$. This has a simple solution, since $\hat{F}^{-1}(x_i) = \mu\hat{L}'(x_i)$ where $\hat{L}'(\cdot)$ is the derivative of $L_2(x)$ of eq. (6b) in which the parameters a, b, d are replaced by their estimates. A suitable estimate of μ is thus obtained by least squares. That is

$$\hat{\mu} = \sum_{i=1}^{m-1} u_i \hat{L}'(x_i) \Big/ \sum_{i=1}^{m-1} \{\hat{L}'(x_i)\}^2,$$

where m is the number of income classes.

The above techniques would be resorted to if the available data consisted (as it often does) of just a few points on the sample Lorenz curve and perhaps some convenient estimate of μ. If instead the data may be considered to represent a sample of size n from the distribution F_2, consistent estimates of the parameters are available. It is convenient to use the parameterization provided in Theorem 4 [eq. (10)]. Denote the ith-order statistic of the sample by $X_{i:n}$. If the X_i's have common density (10), then

$$X_{1:n} \xrightarrow{\text{a.s.}} \tau\eta_1 + \nu,$$

$$X_{n/3:n} \xrightarrow{\text{a.s.}} \tau\xi_{1/3} + \nu,$$

$$X_{2n/3:n} \xrightarrow{\text{a.s.}} \tau\xi_{2/3} + \nu,$$

$$X_{n:n} \xrightarrow{\text{a.s.}} \tau\eta_2 + \nu,$$

where

$$\xi_{1/3} = (2\lambda_1 + \lambda_2)\Big/\sqrt{3 - (2\lambda_1 + \lambda_2)^2},$$

$$\xi_{2/3} = (\lambda_1 + 2\lambda_2)\Big/\sqrt{3 - (\lambda_1 + 2\lambda_2)^2},$$

in which

$$\lambda_1 = \eta_1\Big/\sqrt{1 + \eta_1^2} \quad \text{and} \quad \lambda_2 = \eta_2\Big/\sqrt{1 + \eta_2^2}.$$

If we equate the observed order statistics in (18) to their asymptotic limits and solve, we obtain consistent estimates of the parameters τ, ν, η_1, η_2. Note that, using this approach, the fitted density will have as its support the range of the observed sample.

Maximum likelihood estimation could also be used in such a setting. Use of $X_{1:n}$ and $X_{n:n}$ to estimate the upper and lower limits of the support is defensible and the problem reduces to a two-parameter maximization problem. The resulting likelihood equations may be solved numerically. Data in such detailed form is often not available and we focus on techniques, like the least squares approach outlined above, which are applicable to data sets which consist of just a few points on the sample Lorenz curve.

5. An application and comparisons

We will illustrate the proposed fitting technique using data from the ASCEF first analysed by Kakwani and Podder (1973, 1976) and then by Pakes (1981). These data appear in table 1 together with the class proportions and cumulative proportions. The corresponding sample LC appears in table 2 together with a variety of fitted Lorenz curves, described below.

The model proposed by Kakwani and Podder (1973) produces a LC of the type

$$L_{\text{KP}}(x) = x^{\delta} \exp\{-\eta(1 - x)\}, \qquad 1 < \delta < 2, \quad \eta > 0.$$

Table 1

The ASCEF (1967–68) data.

Income range	Number of income units	Class proportions	Cumulative proportions
< 1000	310	0.05698	0.05698
1000 – 1999	552	0.10147	0.15847
2000 – 2999	1007	0.18511	0.34358
3000 – 3999	1193	0.21930	0.56288
4000 – 4999	884	0.16250	0.72538
5000 – 5999	608	0.11176	0.83714
6000 – 6999	314	0.05772	0.89486
7000 – 7999	222	0.04081	0.93567
8000 – 8999	128	0.02353	0.95920
9000 –10,999	112	0.02059	0.97979
> 11,000	110	0.02022	—

Table 2

Fitted Lorenz curves.

x	Sample Lorenz curve	Estimated Lorenz curves			
		$\hat{L}_{\mathrm{KP}}(x)$	$\hat{L}_{\mathrm{P}}(x)$	$\hat{L}_{\mathrm{C}}(x)$	$\hat{L}_2(x)$
0.1	0.0213	0.02198	0.0331	0.05961	0.021634
0.2	0.0657	0.06368	0.0843	0.12206	0.066451
0.3	0.1273	0.12112	0.1470	0.18785	0.126676
0.4	0.2001	0.19393	0.2200	0.25769	0.199337
0.5	0.2833	0.28255	0.3025	0.33259	0.283347
0.6	0.3781	0.38781	0.3958	0.41405	0.378803
0.7	0.4867	0.51080	0.5013	0.50458	0.487053
0.8	0.6119	0.65283	0.6225	0.60893	0.611588
0.9	0.7624	0.81535	0.7683	0.73900	0.761931
SSE		0.00523	0.00231	0.01622	0.0000025
SAE		0.14349	0.138	0.3423	0.00435

If we let $(x_i, y_i)_{i=1}^m$ denote the m available points on the sample Lorenz curve, then Kakwani and Podder's least squares estimates are obtained by minimizing

$$\sum_{i=1}^{m} [\log y_i - \log L_{\mathrm{KP}}(x_i)]^2 = \sum_{i=1}^{m} [\log y_i - \delta \log x_i + \eta(1-x_i)]^2.$$

The resulting estimates are $\hat{\delta} = 1.462$ and $\hat{\eta} = 0.501$. The corresponding estimated Lorenz curve is denoted by $\hat{L}_{\mathrm{KP}}(x)$ in table 2.

The family proposed by Pakes (1981) has a LC of the form

$$L_{\mathrm{P}}(x)=\int_0^x z^{\delta-1}(1-z)^{\eta-1}\,\mathrm{d}z\Big/B(\delta,\eta),$$

where $\delta \geq 1$, $0<\eta\leq 1$ and B is the beta function. A hybrid estimation scheme was used by Pakes in this situation. First he observed that the corresponding distribution function F has a Paretian upper tail, i.e.,

$$1-F(x)\sim (x/x_0)^{-1/(\eta-1)} \quad [x\to\infty],$$

where $x_0=\mu/B(\delta,\eta)$. He then plots $\log x$ vs. $\log(1-F(x))$ and focusses on the approximately linear upper tail of the graph. Fitting a line to this by least squares gives estimates of $1/(\eta-1)$ and x_0. With these values of η and x_0, the estimated value of δ is obtained by matching the sample and population means. The resulting estimates are $\hat{\delta}=1.33$ and $\hat{\eta}=0.727$. The corresponding estimated Lorenz curve is denoted by $\hat{L}_{\mathrm{P}}(x)$ in table 2.

The well-known classical Pareto LC is

$$L_{\mathrm{C}}(x)=1-(1-x)^{(\gamma-1)/\gamma}, \qquad \gamma>1.$$

A least squares estimate of γ based on sample Lorenz curve points $(x_i, y_i)_{i=1}^m$ is obtained by minimizing

$$\sum_{i=1}^m \left[\log(1-y_i)-\log(1-F_{\mathrm{C}}(x_i))\right]^2$$

$$=\sum_{i=1}^m \left(\log(1-y_i)-[(\gamma-1)/\gamma]\log(1-x_i)\right)^2.$$

The resulting estimate is $\hat{\gamma}=2.4002$ and the corresponding estimated Lorenz curve is denoted by $\hat{L}_{\mathrm{C}}(x)$ in table 2.

The final fitted Lorenz curve in table 2 is an elliptical Lorenz curve $\hat{L}_2(x)$ fitted using the estimation procedure described in section 4. The corresponding parameter estimates are

$$\hat{a}=1.08742, \quad \hat{b}=-1.6601, \quad \hat{d}=0.01906.$$

We observe in table 2 that on the basis of the sum of squared errors (SSE) and the sum of absolute errors (SAE), $\hat{L}_{\mathrm{C}}$ is improved upon by $\hat{L}_{\mathrm{KP}}$ which is improved upon by $\hat{L}_{\mathrm{P}}$ which is improved upon by $\hat{L}_2$. It is also interesting to notice that the estimated values of $\hat{L}_2(x)$ differ from the observed values by less than 0.0008 for every x.

Table 3

Means and supports of fitted densities.

	$\hat{F}'_{KP}$	$\hat{F}'_{P}$	$\hat{F}'_{C}$	$\hat{F}'_{2}$
$\hat{\mu}$	4144	4144	4144	3971
Support	(0, 8135)	$(0, +\infty)$	$(0, +\infty)$	(169, 19967)

Addressing attention to the problem of density fitting, we analyzed the density functions corresponding to the fitted LC's of table 2 which we denote by $\hat{F}'_{KP}$, $\hat{F}'_{P}$, $\hat{F}'_{C}$ and $\hat{F}'_{2}$, respectively. From the original data reported in Kakwani and Podder (1973) we have that the mean observed income is 4144. This value was used as an estimate of μ in $\hat{F}'_{KP}$, $\hat{F}'_{P}$ and $\hat{F}'_{C}$, while the least squares estimate in section 4 for μ was used in $\hat{F}'_{2}$. Table 3 presents the corresponding supports together with their estimates of μ.

The observed mean income per class of the data that appeared in Kakwani and Podder (1976) were used to evaluate the fitted densities and compare the values with the observed frequencies. We also calculated the *SSE* and the *SAE* for the whole range as well as for the right-hand tail from the mode of the observed frequencies (*SSE** and *SAE*). The comparison restricted to the right of the mode was included because this is the region for which Pareto models seem to perform well. The information is displayed in table 4.

The estimates of $\hat{F}'_{KP}(y)$ and $\hat{F}'_{P}(y)$ in table 4 were obtained numerically since $F'_{KP}(y)$ does not admit an explicit form and $F'_{P}(y)$ is obtainable explicitly only if $\delta + \eta = 2$. The classical Pareto density is

$$\hat{F}'_{C}(y) = \hat{\lambda}(\hat{\lambda} y/\hat{\gamma})^{-\hat{\gamma}-1} \quad \text{with} \quad \hat{\lambda} = \hat{\gamma}^2/\hat{\mu}(\hat{\gamma}-1).$$

The general quadratic density $\hat{F}'_{2}(y)$ is given by (7).

From table 4 we notice that the general quadratic density provides a better fit than the density corresponding to the beta LC based upon an *SSE* or *SAE* criterion over the whole range. In the same sense the densities associated with $\hat{L}_{KP}$ and $\hat{L}_{C}$ provide very poor fits. On the basis of *SSE**, $\hat{F}'_{C}$ still provides the best fit (slightly better than $\hat{F}'_{2}$), while on the basis of *SAE**, $\hat{F}'_{P}$ gives the best fit (slightly better than $\hat{F}'_{2}$).

We remark that the poor fit of $\hat{F}'_{KP}$ proposed by Kakwani and Podder (1973) is due in part to its support whose right end point is far below the observed mean income of the upper class (cf. tables 3 and 4). In 1976 the same authors proposed a family of curves for fitting the sample LC under discussion; strictly by numerical means they obtain a curve $\hat{h}_{KP}$ fitted to the observed frequencies. Under the *SSE* criterion our $\hat{F}'_{2}$ is still better than their $\hat{h}_{KP}$ (see table 4); although under the *SAE* criterion $\hat{h}_{KP}$ is better than $\hat{F}'_{2}$. The

Table 4

Comparison of fitted densities.

Mean income y	Relative frequency					
	Observed	$\hat{F}'_{KP}(y)$	$\hat{F}'_{P}(y)$	$\hat{F}'_{C}(y)$	$\hat{F}'_{2}(y)$	$\hat{h}_{KP}$
674.39	0.05698	0.0759	0.0126	76.1940	0.0744	0.0403
1,426.10	0.10147	0.1354	0.0754	5.9718	0.1162	0.1188
2,545.79	0.18511	0.1579	0.2127	0.8325	0.1983	0.2089
3,469.35	0.21930	0.1515	0.2338	0.2906	0.2239	0.2029
4,470.33	0.16250	0.1394	0.1687	0.1227	0.1675	0.1527
5,446.60	0.11176	0.1273	0.1025	0.0627	0.0986	0.1033
6,460.93	0.05772	0.1161	0.0585	0.0351	0.0540	0.0661
7,459.14	0.04081	0.1066	0.0342	0.0215	0.0309	0.0406
8,456.66	0.02353	0.1009[a]	0.0207	0.0140	0.0181	0.0245
9,788.38	0.02059	0.0000	0.0113	0.0085	0.0105	0.0236
15,617.69	0.02022	0.0000	0.0015	0.0017	0.0018	0.0183
SSE		0.02219	0.00423		0.00149	0.00166
SAE		0.42885	0.16623		0.11567	0.10695
*SSE**		0.01992	0.00082	0.01054	0.00080	0.00052
*SAE**		0.34879	0.06819	0.24223	0.07033	0.04915

[a]Evaluated at 8135, the right end point of the support.

curve they fitted to the sample LC is not an LC [cf. Rasche et al. (1980) and Pakes (1981)]. Due to this fact this model was not included in tables 1–3. We have been informed that the Dagum (1980) distribution provides a slightly better fit to the present data set ($SSE = 0.00119$) than does $\hat{F}_2$.

The model proposed by Rasche et al. (1980) for a LC given by

$$L(x) = \{1-(1-x)^{\delta}\}^{\eta}, \qquad 0<\delta<1<\eta,$$

was not considered either in tables 1–3 because, to our knowledge, there is not a simple way to estimate the parameters [cf. Gupta (1984)]; besides, its density function does not admit an explicit form.

It should be pointed out that $\hat{\nu} = -\hat{b}\hat{\mu}/2 = 3{,}296.12$ lies within the support of $\hat{F}_2'$, hence $\hat{F}_2'$ is a unimodal density (cf. table 4). Also, using eq. (13), the Gini index of the fitted LC $\hat{L}_2$ is given by $\hat{G}_2 = 0.3183$ which is very close to the actual value 0.3196 reported in Kakwani and Podder (1976). Besides $\hat{G}_2$ lies within the lower and upper limits, 0.3134 and 0.3223 respectively, of the goodness-of-fit test criterion proposed by Gastwirth and Smith (1972). These limits were computed by Kakwani and Podder (1976).

6. Further comments

The observed frequencies in the example discussed above seem to be reasonably well fitted overall by a density corresponding to an elliptical LC, however, at a first glance we might think that the upper tail is not thick enough. In order to find out how flexible this family of densities is, a study of different sets of data was prepared providing graphical comparisons of observed and fitted LC's and observed and fitted densities in each case. The results are collected in a separate technical report [Villaseñor and Arnold (1984b)]. From this study we observe the following: (a) In most unimodal frequency histograms, the density corresponding to an elliptical LC produced quite satisfactory fits even along the upper tail. (b) The greater the number of classes the better the fit. When the number of classes was small the shape of the fitted density was still in accordance with the histogram. (c) The cases with bimodal histograms proved to be difficult to fit with our unimodal family of densities, even though the corresponding LC fit may still be satisfactory. In one case the estimated value of α was positive indicating a better fit with a hyperbolic LC. Elliptical LC's seem to be very common however.

References

Aggarwal, V. and R. Singh, 1984, On optimum stratification with proportional allocation for a class of Pareto distributions, Communications in Statistics A 13, 3107–3116.

Arnold, B.C., 1983, Pareto distributions (International Cooperative Publishing House, Fairland, MD).

Arnold, B.C., 1986, A class of hyperbolic Lorenz curves, Sankhya 48 B, 427–436.

Dagum, C., 1980, The generation and distribution of income, the Lorenz curve and the Gini ratio, Economie Appliquée 33, 327–367.

Gastwirth, J.L., 1971, A general definition of the Lorenz curve, Econometrica 39, 1037–1039.

Gastwirth, J.L. and J.T. Smith, 1972, A new goodness-of-fit test, Proceedings of the American Statistical Association, 320–322.

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

Kakwani, N.C. and N. Podder, 1973, On estimation of Lorenz curves from grouped observations, International Economic Review 14, 137–148.

Kakwani, N.C. and N. Podder, 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.

Pakes, A.G., 1981, On income distributions and their Lorenz curves, Technical report (Department of Mathematics, University of Western Australia, Nedlands, W.A.).

Rasche, R.H., J. Gaffney, A.Y.C. Koo and N. Obst, 1980, Functional forms for estimating the Lorenz curve, Econometrica 48, 1061–1062.

Villaseñor, J.A. and B.C. Arnold, 1984a, The general quadratic Lorenz curve, Technical report (Colegio de Postgraduados, Chapingo, Mexico).

Villaseñor, J.A. and B.C. Arnold, 1984b, Some examples of fitted general quadratic Lorenz curves, Technical report no. 130 (Statistics Department, University of California, Riverside, CA).