# A NEW FUNCTIONAL FORM FOR ESTIMATING
## LORENZ CURVES

BY P. ORTEGA, G. MARTÍN, A. FERNÁNDEZ, M. LADOUX, AND A. GARCÍA

*University of Málaga, Spain*

There are several functional forms for estimating Lorenz curves from grouped data. Based on studies of the Spanish distribution of income, we propose a new functional form which provides very good fits. Our specification contains the Pareto Lorenz curve as a particular case, and allows one to compute easily, with the provided formulae, the Gini, Kakwani, and Chakravarty Inequality Indexes.

In past years, many papers have been written about the estimation of Lorenz curves from grouped data. In them, their authors propose functional forms for the Lorenz curve. From our point of view, the more relevant are:

$$(1) \qquad f(x) = x^{\alpha} e^{\beta(x-1)}, \qquad \alpha > 0, \beta > 0,$$

$$(2) \qquad f(x) = [1 - (1-x)^{\alpha}]^{1/\beta}, \qquad \alpha > 0, 0 < \beta \le 1$$

which are due to Kakwani and Podder (Kakwani and Podder, 1973), and Rasche *et al.* (Rasche, Gaffney, Koo, and Obst, 1980), respectively.

Other well-known proposed functional forms for the Lorenz curve are:

$$(3) \qquad f(x) = xA^{x-1}, \qquad A > 0,$$

$$(4) \qquad f(x) = x - Ax^{\alpha}(1-x)^{\beta}$$

due to Gupta (Gupta, 1984) and Kakwani (Kakwani, 1980) respectively. However, function (3) is only a particular case of (1), and (4) can not be considered a Lorenz curve because it is not positive in many cases.

Functional forms (1) and (2) have been successfully used in (Fernández Morales *et al.*, 1989; García Lizana *et al.*, 1989) to estimate the Lorenz curves of the income distributions of Spanish provinces and then to compute the Gini's Inequality Index and several poverty indexes.

The purpose of this note is to give a new functional form that provides good fits to the Lorenz curve of many income distributions, as we shall see below, and allows one to compute easily Gini's Index and other inequality measures.

A function $f$ is a Lorenz curve if it verifies:
  (i) $f(x) \ge 0$ for every $x \in [0, 1]$
  (ii) $f(0) = 0$ and $f(1) = 1$,
  (iii) $f$ is derivable in $(0, 1)$ and $f'(x) \ge 0$ for every $x \in (0, 1)$,
  (iv) $f$ admits second derivative in $(0, 1)$ and $f'(x) \ge 0$ for every $x \in (0, 1)$.

The functional form that we propose is the following:

$$(5) \qquad f(x) = x^{\alpha}[1 - (1-x)^{\beta}], \qquad \alpha \ge 0, 0 < \beta \le 1.$$

PROPOSITION. *The functional form defined by* (5) *is a Lorenz curve, and its Gini Index is*

(6)
$$G = \frac{\alpha-1}{\alpha+1} + 2B(\alpha+1, \beta+1),$$

*where B represents the beta function.*

PROOF. Conditions i and ii are trivially verified by $f$. The first and second derivatives of $f$ are respectively:

$$f'(x) = \alpha x^{\alpha-1}(1-(1-x)^{\beta}) + \beta x^{\alpha}(1-x)^{\beta-1},$$

and

$$f''(x) = \alpha(\alpha-1)x^{\alpha-2}[1-(1-x)^{\beta}] + 2\alpha\beta x^{\alpha-1}(1-x)^{\beta-1} + \beta(1-\beta)x^{\alpha}(1-x)^{\beta-2}.$$

It is clear that $f'(x) \geq 0$ for every $x \in (0,1)$ and that $f''(x) \geq 0$ if $\alpha \geq 1$. To see the non-negativity of $f''$ for $0 \leq \alpha < 1$, let

$$u(x) = (1-\alpha)\alpha[1-(1-x)^{\beta}],$$
$$v(x) = 2\alpha\beta x(1-x)^{\beta-1} + \beta(1-\beta)x^{2}(1-x)^{\beta-2}$$

where $u$ is the first term in the expression of $f''$ divided by $(-x^{\alpha-2})$, and $v$ is the sum of the last two terms also divided by $x^{\alpha-2}$. Thus, to prove $f'' \geq 0$ in $(0,1)$ is equivalent to prove $u \leq v$ in $(0,1)$. To do this, let's compute $u'$ and $v'$:

$$u'(x) = (1-\alpha)\alpha\beta(1-x)^{\beta-1}$$
$$v'(x) = 2\alpha\beta(1-x)^{\beta-1} + 2\alpha\beta(1-\beta)x(1-x)^{\beta-2} + 2\beta(1-\beta)x(1-x)^{\beta-2}$$
$$+ \beta(1-\beta)(2-\beta)x^{2}(1-x)^{\beta-3}.$$

Since $u'$ is clearly smaller than the first term of $v'$, and the other terms of $v'$ are non-negative in $(0,1)$ we obtain $u'(x) \leq v'(x)$ for every $x \in (0,1)$. This relation together with $u(0) = v(0) = 0$ assures that $u(x) \leq v(x)$ for every $x \in (0,1)$.

Finally, the Gini's Index is:

$$G = 1 - 2\int_{0}^{1} x^{\alpha}[1-(1-x)^{\beta}]\, dx$$

$$= 1 - 2\int_{0}^{1} x^{\alpha}\, dx + 2\int_{0}^{1} x^{\alpha}(1-x)^{\beta}\, dx$$

$$= 1 - \frac{2}{\alpha+1} + 2B(\alpha+1, \beta+1)$$

$$= \frac{\alpha-1}{\alpha+1} + 2B(\alpha+1, \beta+1).$$

Function (5) includes as a particular case ($\alpha = 0$) the Lorenz curve of the Pareto's income distribution. In addition, for $\beta = 1$ and $\alpha = 0$ we obtain the equaliatarian line.

The functional form (5) has demonstrated to give good fits for the Lorenz curves of a wide range of income distributions. We have done estimations corresponding to 21 countries [from data in (Shorrocks, 1983)] and to the 50 Spanish provinces (from official data in [I.N.E., 1981]). Table 1 contains the sum of squared residuals (SSR) of the fitted Lorenz curve for the Spanish provinces with five different functional forms. The proposed functional form (5) gives in all provinces smaller SSR than Pareto, Gupta (3) and Kakwani and Podder (1) ones, and in many cases than the Rasche *et al.* (2) functional form.

Another advantage of (5) is the possibility of generating easy formulations for inequality measures associated with the Lorenz curve. Here, we give formulae for two of these inequality indexes.

Kakwani (1980) introduced the inequality measure $K_r$ defined by

$$(7) \qquad K_r = 1 - r(r+1) \int_0^1 L(x)(1-x)^{r-1}\, dx$$

where $L(x)$ is the Lorenz curve.

On the other hand, Chakravarty (1988) introduced the generalized inequality indices $I_r$ defined by

$$(8) \qquad I_r = 2 \left( \int_0^1 (x - L(x))^r\, dx \right)^{1/r}.$$

Both indexes are parametric generalizations of the Gini's Index. It is clear that $K_1 = I_1 = G$.

In the next proposition we give formulae for $K_r$ and $I_r$ corresponding to the Lorenz curve defined in (5).

PROPOSITION. *If* $L(x) = x^\alpha[1 - (1-x)^\beta]$, $\alpha \geq 0$, $0 < \beta \leq 1$, *then*

$$(9) \qquad K_r = 1 - r(r+1)[B(\alpha+1, r) - B(\alpha+1, \beta+r)]$$

*and, if r is natural,*

$$(10) \qquad I_r = 2 \left( \sum_{i=0}^{r} \sum_{k=0}^{i} (-1)^{i+k} \binom{r}{i}\binom{i}{k} B[r+1+i(\alpha-1), \beta k+1] \right)^{1/r}.$$

PROOF. By definition (7),

$$K_r = 1 - r(r+1) \int_0^1 x^\alpha[1 - (1-x)^\beta](1-x)^{r-1}\, dx$$

$$= 1 - r(r+1) \left[ \int_0^1 x^\alpha(1-x)^{r-1}\, dx - \int_0^1 x^\alpha(1-x)^{\beta+r-1}\, dx \right]$$

$$= 1 - r(r+1)[B(\alpha+1, r) - B(\alpha+1, \beta+r)].$$

To prove (10), we will develop $(x - x^\alpha(1-(1-x)^\beta))^r$ by Newton's formula.

449

## TABLE 1

SUM OF SQUARED RESIDUALS OF DIFFERENT ESTIMATED FUNCTIONAL FORMS FOR SPANISH PROVINCES LORENZ CURVES

| Province | Pareto | Gupta | Kakwani | Rasche | Ortega |
|---|---|---|---|---|---|
| Alava | 0.0105 | 0.0049 | 0.0050 | 0.000109 | 0.000093 |
| Albacete | 0.0169 | 0.0031 | 0.0018 | 0.000081 | 0.000109 |
| Alicante | 0.0116 | 0.0031 | 0.0064 | 0.000009 | 0.000009 |
| Almeria | 0.0126 | 0.0100 | 0.0097 | 0.000009 | 0.000003 |
| Asturias | 0.0189 | 0.0032 | 0.0017 | 0.000014 | 0.000013 |
| Avila | 0.0128 | 0.0066 | 0.0068 | 0.000099 | 0.000086 |
| Badajoz | 0.0132 | 0.0032 | 0.0036 | 0.000288 | 0.000365 |
| Baleares | 0.0146 | 0.0031 | 0.0033 | 0.000039 | 0.000064 |
| Barcelona | 0.0114 | 0.0036 | 0.0045 | 0.000025 | 0.000012 |
| Burgos | 0.0203 | 0.0054 | 0.0022 | 0.000018 | 0.000019 |
| Caceres | 0.0099 | 0.0033 | 0.0073 | 0.000003 | 0.000007 |
| Cadiz | 0.0190 | 0.0034 | 0.0019 | 0.000020 | 0.000012 |
| Cantabria | 0.0104 | 0.0032 | 0.0056 | 0.000023 | 0.000008 |
| Castellon | 0.0155 | 0.0036 | 0.0026 | 0.000072 | 0.000098 |
| Ciudad Real | 0.0196 | 0.0096 | 0.0046 | 0.000028 | 0.000014 |
| Cordoba | 0.0105 | 0.0032 | 0.0057 | 0.000065 | 0.000100 |
| Coruña | 0.0140 | 0.0031 | 0.0037 | 0.000004 | 0.000013 |
| Cuenca | 0.0085 | 0.0074 | 0.0122 | 0.000026 | 0.000024 |
| Gerona | 0.0138 | 0.0031 | 0.0041 | 0.000061 | 0.000039 |
| Granada | 0.0168 | 0.0067 | 0.0048 | 0.000005 | 0.000019 |
| Guadalajara | 0.0186 | 0.0031 | 0.0012 | 0.000071 | 0.000085 |
| Guipuzcoa | 0.0106 | 0.0070 | 0.0041 | 0.000026 | 0.000028 |
| Huelva | 0.0181 | 0.0065 | 0.0032 | 0.000046 | 0.000065 |
| Huesca | 0.0078 | 0.0059 | 0.0126 | 0.000078 | 0.000039 |
| Jaen | 0.0179 | 0.0045 | 0.0030 | 0.000071 | 0.000108 |
| Leon | 0.0159 | 0.0045 | 0.0039 | 0.000020 | 0.000026 |
| Lerida | 0.0159 | 0.0041 | 0.0020 | 0.000017 | 0.000023 |
| Lugo | 0.0140 | 0.0032 | 0.0044 | 0.000048 | 0.000079 |
| Madrid | 0.0132 | 0.0036 | 0.0058 | 0.000012 | 0.000025 |
| Malaga | 0.0154 | 0.0034 | 0.0032 | 0.000054 | 0.000045 |
| Murcia | 0.0144 | 0.0031 | 0.0039 | 0.000003 | 0.000010 |
| Navarra | 0.0114 | 0.0034 | 0.0070 | 0.000084 | 0.000058 |
| Orense | 0.0157 | 0.0038 | 0.0040 | 0.000022 | 0.000041 |
| Palencia | 0.0219 | 0.0067 | 0.0023 | 0.000014 | 0.000008 |
| Palmas Las | 0.0144 | 0.0035 | 0.0042 | 0.000009 | 0.000013 |
| Pontevedra | 0.0183 | 0.0032 | 0.0019 | 0.000009 | 0.000010 |
| Rioja, La | 0.0166 | 0.0071 | 0.0007 | 0.000030 | 0.000040 |
| Salamanca | 0.0145 | 0.0051 | 0.0052 | 0.000007 | 0.000012 |
| Santa Cruz | 0.0134 | 0.0035 | 0.0048 | 0.000010 | 0.000004 |
| Segovia | 0.0088 | 0.0031 | 0.0070 | 0.000030 | 0.000061 |
| Sevilla | 0.0136 | 0.0031 | 0.0039 | 0.000010 | 0.000018 |
| Soria | 0.0165 | 0.0056 | 0.0040 | 0.000046 | 0.000075 |
| Tarragona | 0.0185 | 0.0033 | 0.0025 | 0.000032 | 0.000046 |
| Teruel | 0.0141 | 0.0034 | 0.0023 | 0.000109 | 0.000141 |
| Toledo | 0.0122 | 0.0036 | 0.0034 | 0.000066 | 0.000098 |
| Valencia | 0.0107 | 0.0032 | 0.0065 | 0.000009 | 0.000001 |
| Valladolid | 0.0134 | 0.0037 | 0.0035 | 0.000081 | 0.000061 |
| Vizcaya | 0.0120 | 0.0035 | 0.0052 | 0.000052 | 0.000058 |
| Zamora | 0.0097 | 0.0034 | 0.0071 | 0.000034 | 0.000038 |
| Zaragoza | 0.0193 | 0.0058 | 0.0033 | 0.000022 | 0.000013 |
| Ceuta Y M. | 0.0186 | 0.0070 | 0.0040 | 0.000022 | 0.000047 |
| España | 0.0158 | 0.0046 | 0.0041 | 0.000003 | 0.000004 |

Source: Data from I.N.E., 1980–81.

450

$$\int_0^1 (x - x^\alpha[1-(1-x)^\beta])^r \, dx = \int_0^1 \left( \sum_{i=0}^r (-1)^i \binom{r}{i} x^{r-i+i\alpha}[1-(1-x)^\beta]^i \right) dx.$$

If we use again Newton's formula to develop the expression $[1-(1-x)^\beta]^i$, we obtain:

$$\int_0^1 (x - x^\alpha[1-(1-x)^\beta])^r \, dx = \int_0^1 \sum_{i=0}^r \sum_{k=0}^i (-1)^{i+k} \binom{r}{i}\binom{i}{k} x^{r-i+i\alpha}(1-x)^{\beta k} \, dx$$

$$= \sum_{i=0}^r \sum_{k=0}^i (-1)^{i+k} \binom{r}{i}\binom{i}{k} B[r+1+i(\alpha-1), \beta k+1].$$

We have estimated the inequality indexes mentioned above for 21 countries and different values of $r$, using the estimated coefficients of (5) and formulae (6), (7) and (8). The results appear in Table 2. For all the indexes, Japan results to be the country with less inequality while Brasil is the one with higher values.

TABLE 2

ESTIMATED INEQUALITY INDEXES FOR 21 COUNTRIES

|  | G | $K_{0.5}$ | $K_{1.5}$ | $K_2$ | $K_3$ | $I_2$ | $I_3$ |
|---|---|---|---|---|---|---|---|
| Australia | 0.3180 | 0.2182 | 0.3891 | 0.4404 | 0.5116 | 0.3412 | 0.3561 |
| Brasil | 0.6370 | 0.5155 | 0.7019 | 0.7420 | 0.7907 | 0.6972 | 0.7351 |
| Columbia | 0.5572 | 0.4254 | 0.6335 | 0.6829 | 0.7446 | 0.6042 | 0.6338 |
| Denmark | 0.3673 | 0.2495 | 0.4500 | 0.5092 | 0.5900 | 0.3955 | 0.4134 |
| Finland | 0.4712 | 0.3303 | 0.5625 | 0.6249 | 0.7057 | 0.5089 | 0.5328 |
| India | 0.4563 | 0.3517 | 0.5167 | 0.5554 | 0.6039 | 0.4956 | 0.5204 |
| Indonesia | 0.4436 | 0.3544 | 0.4908 | 0.5187 | 0.5506 | 0.4881 | 0.5132 |
| Japan | 0.3116 | 0.2182 | 0.3768 | 0.4232 | 0.4870 | 0.3341 | 0.3486 |
| Kenya | 0.6233 | 0.5139 | 0.6797 | 0.7140 | 0.7550 | 0.6851 | 0.7242 |
| Malaysia | 0.5116 | 0.3780 | 0.5927 | 0.6465 | 0.7150 | 0.5528 | 0.5789 |
| Netherl. | 0.4479 | 0.3201 | 0.5300 | 0.5861 | 0.6596 | 0.4825 | 0.5045 |
| N. Zealand | 0.3693 | 0.2514 | 0.4516 | 0.5106 | 0.5908 | 0.3974 | 0.4155 |
| Norway | 0.3594 | 0.2434 | 0.4412 | 0.5001 | 0.5807 | 0.3868 | 0.4044 |
| Panama | 0.4466 | 0.3199 | 0.5278 | 0.5832 | 0.6558 | 0.4810 | 0.5029 |
| Spain | 0.3436 | 0.2382 | 0.4167 | 0.4688 | 0.5401 | 0.3688 | 0.3850 |
| Sri Lanka | 0.4106 | 0.2878 | 0.4924 | 0.5493 | 0.6252 | 0.4419 | 0.4618 |
| Sweden | 0.3873 | 0.2679 | 0.4687 | 0.5262 | 0.6037 | 0.4166 | 0.4353 |
| Tanzania | 0.5389 | 0.4291 | 0.5987 | 0.6359 | 0.6815 | 0.5885 | 0.6198 |
| Tunisia | 0.5094 | 0.3678 | 0.5974 | 0.6561 | 0.7310 | 0.5505 | 0.5765 |
| Uruguay | 0.5011 | 0.3515 | 0.5966 | 0.6610 | 0.7430 | 0.5424 | 0.5687 |
| U. Kingdom | 0.3618 | 0.2571 | 0.4319 | 0.4807 | 0.5465 | 0.3886 | 0.4057 |

*Sources*: Data from Shorrocks, A., 1983, and *I.N.E.*, 1980-81.

Kakwani's and Chakravarty Indexes are highly correlated with Gini Index, but this is even more evident for the latter, with correlations above 0.99. In addition, Kakwani's Indexes show decreasing dispersion with $r$, in contrast with Chakravarty's Indexes, which show increasing dispersion with $r$.

## REFERENCES

Chakravarty, S. R., Extended Gini Indices of Inequality, *International Economic Review*, 29, 147–156, 1988.

Fernández Morales, A., García Lizana, A., and Martín Reyes, G., Técnicas para la medición de la pobreza y su distribución territorial en España, *Comunicaciones presentadas a la XIV Reunión de Estudios Regionales*, A.E.C.R., Málaga, 1988.

García Lizana, A., Ladoux Aguilar, M., Martín Reyes, G., Fernández Morales, A., and Cabello Gonzalez, J. M., El contexto socioeconómico de la pobreza en España, *Cáritas*, 288, 15–26, 1989.

Gupta, M. R., Functional Form for Estimating the Lorenz Curve, *Econometrica*, 52, 1313–1314, 1984.

Kakwani, N., On a Class of Poverty Measures, *Econometrica*, 48, 437–446, 1980.

Kakwani, N., and Podder, N., On the Estimation of Lorenz Curves from Grouped Observations, *International Economic Review*, 14-2, 278-291, 1973.

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

Shorrocks, A. F., Ranking Income Distributions, *Economica*, 50, 3–17, 1983.