

O desempenho de duas funções de Lorenz com dados de renda do Brasil

José W. Rossi*

Neste artigo comparamos o desempenho de duas funções de Lorenz, propostas recentemente na literatura, quando aplicadas aos dados dos censos de 1970 e 1980, para a distribuição de renda no Brasil. A primeira, sugerida por Kakwani e Podder (1976), é definida em termos de um sistema de novas coordenadas para a curva de Lorenz, os parâmetros da qual poderão ser estimados pelo método dos mínimos quadrados ordinários, já que a função é log-linear. A segunda função, proposta por Rasche et alii (1980), é intrinsecamente não-linear, apresentando, portanto, maior complexidade quanto à estimação dos parâmetros. Em ambos os casos o índice de Gini poderá ser prontamente obtido a partir dos parâmetros da função.

As duas funções apresentaram resultados muito bons com os dados dos censos, não permitindo sequer estabelecer, do ponto de vista estatístico pelo menos, qualquer superioridade de uma sobre a outra.

1. Introdução; 2. Sobre as funções; 3. Aplicação.

1. Introdução

A preocupação com a mensuração na desigualdade da distribuição das rendas pessoais vem de longa data. Vilfredo Pareto com o seu *Cours d'économie politique*, publicado em 1896, por exemplo, representa um esforço pioneiro para medir a desigualdade por meio do ajustamento de uma função matemática a dados empíricos de renda. Ainda outros métodos têm sido desenvolvidos para avaliar o grau de concentração da renda, sem a preocupação com qualquer função matemática específica. Na verdade, esses dois procedimentos têm florescido lado a lado.

Nestas notas, entretanto, nos ocuparemos tão-somente com o desempenho de duas funções de Lorenz sugeridas recentemente por Kakwani e Podder (1976) e

* Professor de economia e econometria na Coppe da UFRJ.

Rasche et alii (1980), quando aplicadas aos dados dos censos de 1970 e 1980 para a distribuição de renda no Brasil. No item a seguir, discutiremos os detalhes dessas duas funções, para no item 3 apresentarmos uma comparação sobre os seus respectivos desempenhos com os dados de distribuição de renda dos dois censos.

2. Sobre as funções

2.1 *Considerações gerais*

É geralmente reconhecido na literatura – *e.g.*, Cramer (1971) – que a tradicional função de Pareto só produz resultados razoáveis quando ajustada à parte superior da distribuição das rendas pessoais que, freqüentemente, representa menos da metade dos recipientes da renda. Nestas circunstâncias, não é de surpreender que, em se tratando de ajustamento para toda a distribuição da renda, a distribuição log-normal fornece, em geral, melhores resultados que a função de Pareto.¹

Várias outras funções matemáticas têm sido sugeridas recentemente na literatura para descrever a distribuição das rendas pessoais. Por exemplo, Salem & Mount (1974), após ajustarem uma função gama aos dados das rendas familiares nos EUA, no período 1967-70, concluem que esta fornece melhores resultados que a distribuição log-normal.² Por outro lado, Singh & Maddala (1976), tendo ajustado a esses dados uma função de Burr, representando uma generalização da função de Pareto e da distribuição de Weibull, concluem que os seus resultados são ainda melhores que aqueles da distribuição gama. Aliás, a distribuição ajustada produziu índices de Gini³ que se situaram em todos os anos dentro do intervalo estabelecido por Gastwirth (1972) e que, por sugestão de Gastwirth e Smith (1972), tem sido geralmente utilizado como teste para a precisão do ajustamento de uma dada função a dados de distribuição de renda.

Não é nosso propósito, entretanto, rever aqui as várias funções que têm sido propostas na literatura para descrever a distribuição das rendas pessoais. O objetivo destas notas é, conforme já dito, tão-somente avaliar o desempenho das duas funções discutidas a seguir, quando aplicadas aos dados dos censos de 1970 e 1980.

2.2 *A função de Kakwani & Podder (1976)*

A função de Kakwani & Podder (1976) é dada em termos de variáveis que são definidas no contexto de um sistema de novas coordenadas para a curva de Lorenz. Vejamos os pontos principais do seu desenvolvimento.

¹ Sobre este ponto, veja Hart (1975), por exemplo.

² Os autores mostram ainda que o índice de Gini pode ser facilmente calculado a partir dos parâmetros da função.

³ Conforme demonstrado por Cronin (1979), o índice de Gini poderá ser também calculado facilmente a partir dos parâmetros da função de Singh & Maddala (1976).

Seja x a variável renda e $f(x)$ a sua função de densidade. Se denominarmos F e F_1 , respectivamente, as percentagens acumuladas das pessoas e percentagens acumuladas das rendas recebidas por tais pessoas, no diagrama da curva de Lorenz, então temos:

$$F = \int_0^x f(x) dx, \quad (1)$$

com

$$\frac{dF}{dx} = f(x), \quad (2)$$

e

$$F_1 = \frac{\int_0^x x f(x) dx}{\int_0^\infty x f(x) dx} = \frac{1}{\mu} \int_0^x x f(x) dx, \quad (3)$$

com

$$\frac{dF_1}{dx} = \frac{x f(x)}{\mu} \quad (4)$$

onde μ representa o valor médio da variável renda, x .⁴

Considere agora que um sistema de novas coordenadas seja definido como⁵

$$n = \frac{1}{\sqrt{2}} (F - F_1) \quad \text{e} \quad \pi = \frac{1}{\sqrt{2}} (F + F_1) \quad (5)$$

cuja representação gráfica é mostrada na figura 1.

Note que π varia entre zero e $\sqrt{2}$, e $n \leq \pi$.

Destes resultados, verifica-se facilmente que

$$\frac{dn}{dx} = \frac{1}{\sqrt{2}} \left[f(x) - \frac{x f(x)}{\mu} \right] \quad \text{e} \quad \frac{d\pi}{dx} = \frac{1}{\sqrt{2}} \left[f(x) + \frac{x f(x)}{\mu} \right] \quad (6)$$

permitindo obter-se

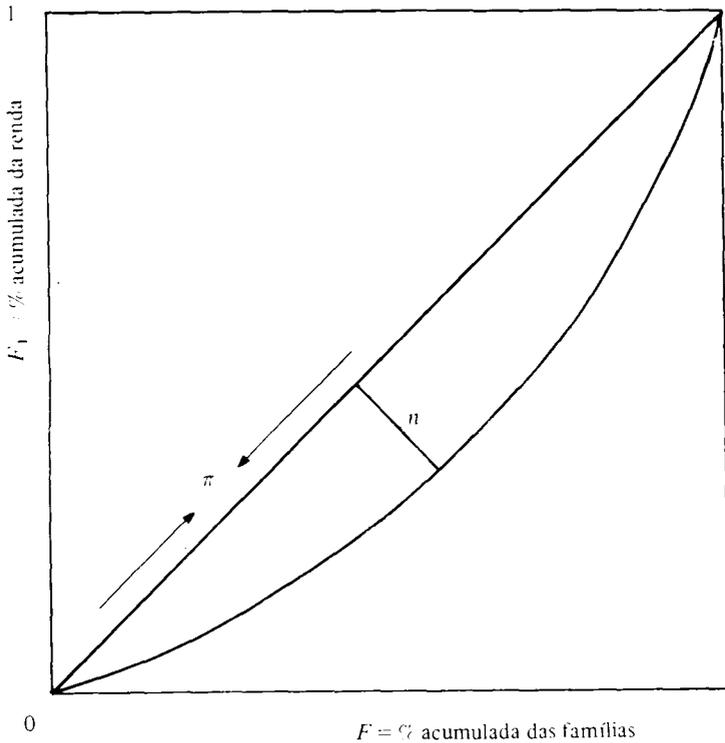
$$\frac{dn}{d\pi} = g'(\pi) = \frac{\mu - x}{\mu + x} \quad (7)$$

⁴ Observe que $dF_1/dF = x/\mu > 0$ e $d^2F_1/dF^2 = 1/\mu f(x) > 0$, indicando pois ser a curva de Lorenz convexa na direção do eixo de F (i.e., $F_1 < F$).

⁵ Veja a prova deste resultado no anexo.

Figura 1

Curva de Lorenz para a Função $n = a\pi^\alpha (\sqrt{2} - \pi)^\beta$



onde $g'(\pi)$ é obviamente a derivada da função de Lorenz (definida no contexto das novas coordenadas) com relação à variável π . Não é difícil mostrar ainda ser a derivada segunda dada por:

$$g''(\pi) = -2\mu^2 \sqrt{2} / f(x)(\mu + x)^3 < 0 \quad (8)$$

isto é, a curva de Lorenz é **convexa**.

Considerando as novas coordenadas definidas, Kakwani e Podder (1976) sugerem que a seguinte forma funcional seja utilizada para a curva de Lorenz⁶

$$n = a\pi^\alpha (\sqrt{2} - \pi)^\beta, \quad \text{para } \hat{a} \geq 0, \alpha > 0 \text{ e } \beta > 0 \quad (9)$$

⁶ Além da aplicação para problemas de distribuição de renda, esta função tem outras importantes aplicações em análise econômica, uma destas sendo a estimativa das elasticidade-renda nos dispêndios familiares. Sobre isto veja Kakwani (1978), e, para uma aplicação a dados brasileiros, veja Rossi (1982).

e que tem a propriedade de ser: a) simétrica, se $\alpha = \beta$; b) enviesada na direção (1,1), se $\beta > \alpha$; c) enviesada na direção (0,0), se $\beta < \alpha$. Adicionalmente, a aplicação das condições em (7) e (8) permitem mostrar que se $0 < \alpha \leq 1$ e $0 < \beta \leq 1$, então fica excluída qualquer possibilidade de a função apresentar pontos de inflexão.⁷

O índice de Gini da função em (9) poderá ser derivado de maneira simples a partir dos parâmetros α e β . Mais precisamente, escreva-se o índice como

$$G = \int_0^{\sqrt{2}} g(\pi) d\pi. \quad (10)$$

Então a substituição da função em (9) nesta relação, e considerando ainda a definição da função beta como

$$B(1 + \alpha; 1 + \beta) = \int_0^1 y^\alpha (1 - y)^\beta dy, \quad (11)$$

tem-se que:⁸

$$G = 2a(\sqrt{2})^{1+\alpha+\beta} \cdot B(1 + \alpha; 1 + \beta) \quad (12)$$

2.3 A função sugerida por Rasche, Gaffney, Koo & Obst (1980)

Como ponto de partida, esses autores observam que as restrições $a \geq 0$, $0 \leq \alpha \leq 1$ e $0 \leq \beta \leq 1$, apresentadas para a função em (9), não satisfazem todas as condições seguintes, necessárias para uma curva de Lorenz apropriada: a) se $F = 0$, então $F_1 = 0$; b) se $F = 1$, então $F_1 = 1$; c) $F_1 < F$; d) a inclinação da curva deve ser não-negativa e monotonicamente crescente. De acordo com os autores, aquelas restrições foram presumivelmente estabelecidas para que a função tivesse um máximo em termos de π e n e não pudesse ter pontos de inflexão no intervalo de zero a $\sqrt{2}$, para π . Ocorre, como alegam, que a condição necessária para tal é que o limite da inclinação da função iguale-se a 1 e -1 quando π se aproxima das extremidades zero e $\sqrt{2}$. Porém, este não é o caso da função em (9), já que a sua inclinação não tem limites naqueles extremos (mais especificamente, $g'(\pi) \rightarrow \infty$ quando $\pi \rightarrow 0$ e $g'(\pi) \rightarrow -\infty$ quando $\pi \rightarrow \sqrt{2}$).

Essas críticas são, entretanto, rebatidas por Kakwani (1980) com o argumento de que "... computamos os valores de $g'(\pi)$ para um grande número de conjuntos alternativos para os valores de α e β no intervalo $0 < \alpha < 1$ e $0 < \beta < 1$.

⁷ Este ponto poderá ser visto claramente após escrevermos a derivada segunda da função como (Kakwani & Podder, 1976):

$$g''(\pi) = -n \left[\frac{\alpha(1-\alpha)}{\pi^2} + \frac{\beta(1-\beta)}{(\sqrt{2}-\pi)^2} + \frac{2\alpha\beta}{\pi(\sqrt{2}-\pi)} \right]$$

⁸ Faça $y = \pi/\sqrt{2}$, para facilitar a demonstração.

Os cálculos mostraram que em quase todos os casos a função $g'(\pi)$ variara de 1 a -1 para π variando de 1.2×10^{-7} (aproximadamente zero) a $\sqrt{2} - 10^{-4}$ (aproximadamente $\sqrt{2}$), mas tendo assíntotas em $\pi = 0$ e $\pi = \sqrt{2}$. Isto sugere que as singularidades nos pontos extremos não deverão afetar os resultados empíricos” (Kakwani, 1980, p. 1.063). O autor prossegue no seu argumento, recorrendo à evidência empírica fornecida por Jain (1975): “Ela [Jain] ajustou a função de Kakwani & Podder a cerca de 500 distribuições observadas da renda para 70 países. O estudo mostrou claramente que a função de densidade subjacente a esta equação da curva de Lorenz fornece boa precisão no ajustamento para um largo número das distribuições observadas de renda nos diversos países” (Kakwani, 1980, p. 1.063).

De qualquer maneira, Rasche et alii (1980) propõem uma forma funcional alternativa que satisfaz as condições de a a d mencionadas, e cuja equação é dada por:

$$F_1 = [1 - (1 - F)^\alpha]^{1/\beta} \quad (13)$$

onde $\alpha > 0$ e $\beta \leq 1$.

É mostrado pelos autores que, se $\alpha < 1$: a) a primeira derivada da função é nula no ponto zero e, quando avaliada no ponto 1, é infinita; b) a derivada segunda será sempre positiva no campo de variação $0 < F < 1$. Isto é, a função possui a convexidade apropriada e as restrições da inclinação garantem que a curva sempre se situará dentro do triângulo abaixo da linha da perfeita igualdade do diagrama de Lorenz.

Finalmente, se definirmos o índice de Gini como

$$G = 1 - 2 \int_0^1 [1 - (1 - F)^\alpha]^{1/\beta} dF \quad (14)$$

e se fizermos, seguindo Rasche et alii (1980), $\mu = 1 - (1 - F)^\alpha$ então com procedimento semelhante àquele em relação ao resultado em (12), tem-se que:

$$G = 1 - \frac{2}{\alpha} B\left(\frac{1}{\beta} + 1; \frac{1}{\alpha}\right), \quad (15)$$

onde B é a função beta, definida em (11).

3. Aplicação

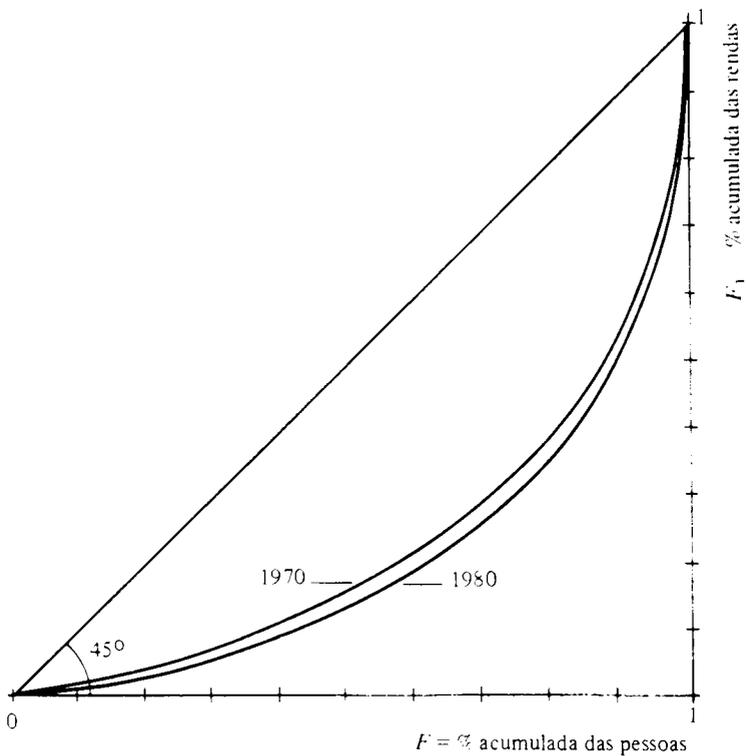
Com o objetivo de avaliar o desempenho das duas funções definidas em (9) e (13), utilizaremos aqui os dados sobre as distribuições de renda no Brasil, como fornecidas pelos censos de 1970 e 1980. Tais distribuições são dadas no quadro 1.

Quadro 1
Distribuição de renda da população economicamente ativa – Brasil

Classes	População (%)		Rendimentos (%)	
	1970	1980	1970	1980
Até 1/2 salário mínimo (SM)	28,4	12,5	5,7	1,4
> 1/2 até 1 SM	32,2	20,8	15,9	5,7
> 1 até 2 SM	21,7	31,1	19,6	15,2
> 2 até 5 SM	12,7	23,6	24,7	25,7
> 5 até 10 SM	3,3	7,2	14,6	17,7
> 10 até 20 SM	1,3	3,2	11,8	15,2
Maior que 20 SM	0,4	1,6	7,7	19,1

Fonte: *Jornal do Brasil*, 23 set. 1981, p. 19.

Figura 2
Curvas de Lorenz para as distribuições de renda no Brasil em 1970 e 1980



As curvas de Lorenz correspondentes a essas duas distribuições estão na figura 2. Observe que a curva para 1980 está claramente abaixo daquela para 1970, indicando portanto não haver qualquer ambigüidade, de acordo com Atkinson (1970), sobre qual distribuição é a mais concentrada. Não é nosso propósito, entretanto, investigar esta questão aqui.⁹ Na verdade, o gráfico das curvas de Lorenz é aqui apresentado mais como um elemento de verificação para avaliar se as funções discutidas produzem pelo menos índices de Gini com valores na direção correta, nos dois anos considerados.

O ajustamento de mínimos quadrados da função em (9) aos dados do quadro 1, forneceu os resultados mostrados no quadro 2.

Quadro 2
Parâmetros estimados para $n = a\pi^\alpha (\sqrt{2} - \pi)^\beta$

Ano	a	α	β	R^2	G
1970	0,6072	1,0552	0,9679	0,9985	0,5678
1980	0,5978	0,9941	0,9312	0,9974	0,5848

Observe que é muito bom o ajustamento como dado pelo coeficiente de determinação (R^2). E isto, apesar de termos uma ligeira violação de uma das restrições para os parâmetros, já que em 1970 temos α maior que 1. Foram calculados ainda os índices de Gini, conforme a expressão em (12), cujos valores encontram-se também no quadro 2, e aos quais retornaremos mais adiante.

A função em (13), por outro lado, que exige estimação por métodos não lineares, quando aplicada aos mesmos dados dos censos de 1970 e 1980, forneceu os resultados do quadro 3.¹⁰

Quadro 3
Parâmetros estimados para $F_1 = [1 - (1 - F)^\alpha]^\frac{1}{\beta}$

Ano	α	β	R^2	G	G_I	G_S
1970	0,5055	0,6219	0,9987	0,5705	0,545	0,5727
1980	0,4936	0,6089	0,9992	0,5900	0,565	0,5927

Note, primeiramente, que são bastante bons os resultados como dados pelo coeficiente de determinação. Ressalte-se aqui que tais coeficientes não devem ser comparados com aqueles apresentados no quadro 2, para a função em (9).

⁹ Para uma análise destas distribuições, veja Rossi (1981).

¹⁰ O método de estimação, aqui utilizado, foi aquele sugerido por Maiquardt (1963).

Isto porque aqueles referem-se ao ajustamento de mínimos quadrados da forma logarítmica das variáveis π e n . Naturalmente, para tornar a comparação válida, teríamos que recalculer o coeficiente de determinação em termos das variáveis originais (i.e., F e F_1). A adoção de tal procedimento permitiu obter coeficientes em torno do valor 0,9995 para ambos os anos. Conclui-se portanto que, em termos do coeficiente de determinação, pelo menos, parece difícil dizer-se qual função apresenta o melhor ajustamento.

No quadro 3, fornecemos ainda os índices de Gini, calculados pela expressão em (15), bem como o intervalo de Gastwirth (1972)¹¹ para o índice, dado pelos valores G_I e G_S . Note que, também com relação a isto, ambas as funções apresentam bons ajustamentos, pois os seus índices se situam dentro do intervalo (G_I , G_S),¹² conforme recomendado pelo teste de Gastwirth & Smith (1972).

Quadro 4
Distribuição de renda da população economicamente ativa — Brasil

Frequências relativas das rendas						
Classes	Observadas		Previstas pela função			
			Kakwani & Podder		Rasche et alii	
	1970	1980	1970	1980	1970	1980
1	0,057	0,014	0,061	0,016	0,050	0,011
2	0,159	0,057	0,140	0,045	0,157	0,050
3	0,196	0,152	0,212	0,164	0,213	0,161
4	0,247	0,257	0,254	0,266	0,251	0,270
5	0,146	0,177	0,141	0,170	0,132	0,169
6	0,118	0,152	0,114	0,144	0,100	0,136
7	0,077	0,191	0,078	0,195	0,097	0,204

¹¹ Deve ser ressaltado, aqui, que este intervalo fora calculado de maneira aproximada, já que: a) em vez das rendas médias das classes, usamos (por desconhecê-las) os seus pontos médios; b) na última classe, que é aberta, consideramos o valor de 40 salários mínimos como limite superior da renda; e finalmente c) para a renda média total, consideramos o mesmo valor de 2,73 salários mínimos, prevalescente em 1980, como válido em 1970, por desconhecermos o valor deste último. Note que todas essas informações são necessárias, além dos próprios limites de renda das várias classes, para o cálculo do intervalo de Gastwirth, conforme pode-se constatar da nota seguinte.

¹² Estes índices são dados por:

$$G_S = G_I + D,$$

onde $D = \sum \lambda_i^2 (\mu_i - a_{i-1}) (a_i - \mu_i) \div \mu (a_i - a_{i-1})$, sendo λ_i , a proporção de pessoas na classe i ; a_i e a_{i-1} , os limites superior e inferior da renda na classe i , e μ_i e μ são as rendas médias das classe i e de toda a distribuição da renda, respectivamente. Quanto ao G_I , é calculado pelo método tradicional da soma das áreas dos trapézios sob a curva de Lorenz (isto é, ignorando as desigualdades das rendas dentro das classes). Para maiores detalhes, veja Gastwirth (1972).

Finalmente, como forma alternativa para comparar o desempenho das duas funções, apresentamos no quadro 4 as frequências relativas das rendas, tanto observadas quanto estimadas pelas duas funções, nas várias classes da renda, para os mesmos dados dos censos: observe que, uma vez mais, parece difícil poder-se dizer qual função produz melhor resultado.¹³ Deve ser reconhecido aqui, entretanto, que a função em (13) apresenta uma desvantagem em relação àquela em (9). É que, sendo intrinsecamente não-linear (isto é, não há transformação das variáveis que permita colocar a função em forma linear nos parâmetros), apresenta maior complexidade quanto aos aspectos da estimação, pois a solução é obtida de maneira iterativa, podendo apresentar dificuldade no que diz respeito à convergência do estimador para a solução ótima.

Anexo

Prova de que:

$$a) \pi = \frac{1}{\sqrt{2}} (F + F_1)$$

Da figura 3, temos que:

$$\pi^2 = OL^2 + LQ^2 \quad \text{ou} \quad \pi^2 = 2LQ^2, \quad \text{já que} \quad OL = LQ.$$

$$\text{Então} \quad \frac{1}{\sqrt{2}} \pi = LQ = OL.$$

Note que: 1) $OL = F - LM$; 2) $LQ = F_1 + NQ$. Então, 1) + 2) fornece $2LQ = F + F_1 + NQ - LM$.

$$\text{Como} \quad NQ = LM, \quad \text{temos que} \quad 2LQ = F + F_1 \Rightarrow \pi = \frac{1}{\sqrt{2}} (F + F_1).$$

$$b) n = \frac{1}{\sqrt{2}} (F - F_1)$$

Ainda da figura, temos:

$$n^2 = NQ^2 + NP^2 \quad \text{ou} \quad n^2 = 2NP^2. \quad \text{já que} \quad NQ = NP.$$

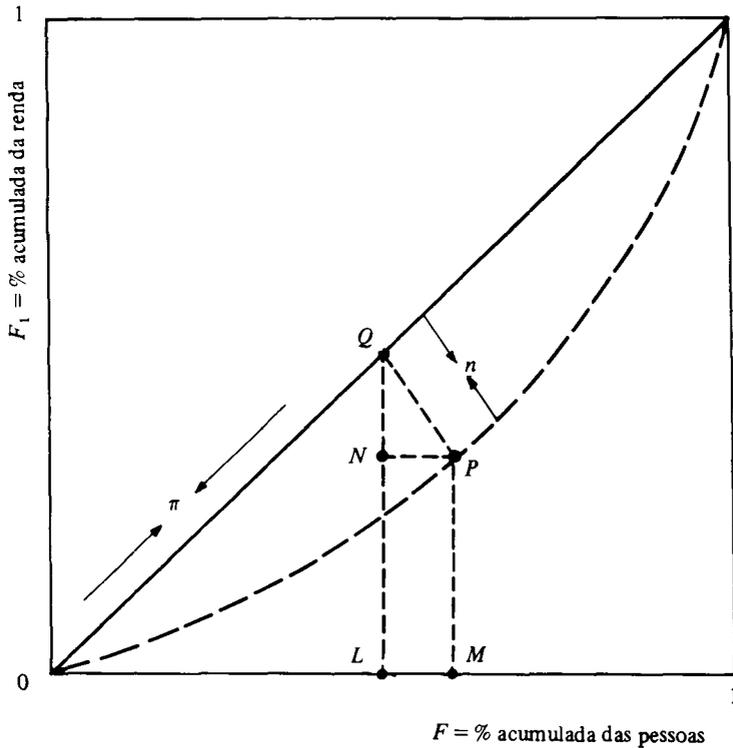
$$\text{Então} \quad \frac{1}{\sqrt{2}} n = NP = NQ.$$

¹³ Kakwani (1980) chegou a conclusões semelhantes com dados relativos à Austrália, pois os resultados por ele obtidos não permitiram determinar qual fora o melhor ajustamento entre essas duas funções.

Note que: 3) $NP = F - OL$; 4) $NQ = OL - F_1$. Então, 3) + 4) fornece $2NQ = F - F_1 \Rightarrow n = \frac{1}{\sqrt{2}} (F - F_1)$.

Figura 3

Curva de Lorenz para a Função $n = a\pi^\alpha (\sqrt{2} - \pi)^\beta$



Abstract

In this paper we compare the performance of two Lorenz functions, suggested recently in the literature, when applied to income data for Brazil. The first of these functions, proposed by Kakwani & Podder (1976), is defined in terms of a new coordinate system for the Lorenz curve, the parameters of which can be estimated by Ordinary Least Square, as the function is linear in the logs. The second function, suggested by Rasche et alii (1980), is intrinsically non-linear, thus presenting more complexity on its estimation aspect. In both cases the Gini index can be readily calculated from the parameters of the function.

The two functions produced very good results with the Brazilian data. In fact, we were even unable to establish, on statistical grounds at least, any superiority of one over the other.

Referências bibliográficas

- Atkinson, A. B. "On the measurement of inequality. *Journal of Economic Theory*, 2 (3): 244-63, Sept. 1970.
- Cramer, J. S. *Empirical econometrics*. North-Holland, 1971.
- Cronin, D. C. A function for size distribution of incomes: a further comment. *Econometrica*, 47 (3): 773-4, May 1979.
- Gastwirth, J. L. The estimation of the Lorenz curve and Gini index. *The Review of Economics and Statistics* (3): 306-16, Aug. 1972.
- _____. & Smith, J. T. A new Goodness-of-Fit Test. The American Statistical Association, Proceedings of the Business and Economic Statistics Section, 1972. p. 320-2.
- Hart, P. E. Moment distribution in economics: an exposition. *Journal of the Royal Statistical Society*, Series A, 138, part 3, p. 423-34, 1975.
- Jain, S. *Size distribution of income: a compilation of data*. Washington, DC, USA. A World Bank Publication, 1975.
- Kakwani, N. C. A new method of estimating engel elasticities. *Journal of Econometrics*, 8: 103-10, 1978.
- _____. Functional forms for estimating the Lorenz curve: a reply. *Econometrica*, 48 (4): 1.063-4, May 1980.
- _____. & Podder, N. Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. *Econometrica*, 44, (1): 137-48, Jan. 1976.
- Marquardt, D. W. An algorithm for least square estimation of non-linear parameters. *Journal of the Society of Industrial and Applied Mathematics*, 2: 431-41, 1963.
- Rasche, R. H.; Gaffney, J., Koo, A. Y. & Obst, N. Functional forms for estimating the Lorenz curve. *Econometrica*, 48 (4): 1.061-2, May 1980.
- Rossi, J. W. O agravamento da concentração de renda. *Jornal do Brasil*, Caderno Especial, 22 nov. 1981. p. 3.
- _____. *Índices de desigualdade de renda e medidas de concentração industrial: aplicações a casos brasileiros*. Rio de Janeiro, Zahar, 1982.
- Salem, B. Z. & Mount, T. D. A convenient descriptive model of income distribution: the Gamma Density. *Econometrica*, 42 (6): 1.115-27, nov. 1974.
- Singh, S. K. & Maddala, G. S. A function for size distribution of incomes. *Econometrica*, 44 (5): 963-70, Sept. 1976.