R. de Econometria	Rio de Janeiro	v. VI nº 2	p. 83 - 98	nov 1986

UTILIZAÇÃO DA CURVA DE PARETO TRUNCADA PARA ESTIMAR A RENDA TOTAL DE ESTRATOS*

Rodolfo Hoffmann**

RESUMO

Freqüentemente é necessário estimar a renda total de cada estrato quando se dispõe de dados sobre o número de pessoas distribuídas em estratos de renda. A obtenção dessa estimativa é particularmente difícil para o estrato de rendas mais altas, sem limite superior finito. O uso da curva de Pareto com dois parâmetros muitas vezes leva a estimativas exageradamente altas da renda média nesse estrato. Este trabalho mostra como a curva de Pareto truncada pode ser utilizada para estimar a renda total dos estratos de rendas relativamente elevadas. Em seguida o método proposto é utilizado para analisar a distribuição da renda entre as pessoas economicamente ativas na agropecuária, no Brasil, em 1980, considerando a classificação dessas pessoas como empregados, empregadores ou autônomos.

---

* Trabalho apresentado no V Encontro de Econometria da Região Sudeste, em Piracicaba, SP, nos dias 22 e 23 de maio de 1986.

** Professor da ESALQ-USP, 13400 - Piracicaba, SP.

ABSTRACT

Frequently we need to estimate total income for each income class from data reporting the number of persons in each class. This estimation is particularly difficult for the highest income class, without a finite upper limit. In many cases the use of the two-parameter Pareto curve will overstimate the mean income of this class. This paper shows how the truncated Pareto curve can be used to more correctly estimate the total income of the higher income classes. The method is applied in an analysis of the income distribution among economically active persons in brazilian agriculture in 1980, classified as employees, employers or autonomous (self-employed).

1. INTRODUÇÃO

Em trabalho recente, em co-autoria com Angela Kageyama, analisamos a distribuição da renda entre pessoas economicamente ativas (PEA) na agricultura brasileira, considerando o tempo de trabalho semanal dessas pessoas e sua "posição na ocupação", isto é, sua classificação como empregados, empregadores ou autônomos (ver Hoffmann e Kageyama, 1985). Dispúnhamos de dados, fornecidos pelo IBGE e originários do recenseamento de 1980, em que as pessoas ativas são distribuídas em 10 estratos de rendimento mensal. Como não dispúnhamos do rendimento total ou médio por estrato, esses valores médios foram fixados, tendo por base outros conjuntos de dados, em que a renda média por estrato era conhecida. Houve, evidentemente, um certo grau de arbitrariedade ao fixar aqueles valores. O problema é mais grave quando se trata de fixar o rendimento médio das pessoas ativas que ganham mais de 20 salários mínimos (S.M.). Naquele trabalho esse rendimento médio foi fixado em 40 S.M.

Ao fixar a média do último estrato em 40 S.M., certamente estamos subestimando a desigualdade entre as "posições na ocupação" pois os empregadores que ganham mais de 20 S.M. certamente têm um rendimento médio maior do que o dos empregados no mesmo estrato.

Para superar esse problema, é necessário um método que permita estimar a renda média desse estrato a partir dos próprios dados. O método "clássico" consiste em utilizar a função de Pareto com dois parâmetros:

$$U = \frac{\theta}{x^{\alpha}}, \tag{1}$$

onde U é o número de pessoas com renda maior do que x. Os parâmetros α e Θ podem ser estimados a partir de informações sobre o número de pessoas por estrato de renda. A renda média das pessoas com renda maior do que x_o é dada por

$$\mu_o = \frac{1}{U_o} \int_{x_o}^{\infty} x \left(- \frac{dU}{dx} \right) dx$$

com $U_o = \dfrac{\Theta}{x_o^{\alpha}}$

Segue-se que

$$\mu_o = \frac{x_o^{\alpha}}{\Theta} \int_{x_o}^{\infty} x \cdot \frac{\alpha\Theta}{x^{\alpha+1}} dx$$

ou

$$\mu_o = \frac{\alpha}{\alpha-1} x_o \qquad (2)$$

Esse resultado só vale, evidentemente, para $\alpha > 1$. Quando α é apenas ligeiramente maior do que um, o valor da média μ_o pode se tornar exageradamente grande.

Infelizmente, ao estimar o valor de α, não é muito raro encontrar valores de α apenas ligeiramente maiores do que um ou mesmo menores do que um.

Optamos, então, pelo uso da curva de Pareto truncada, descrita a seguir.

2. A CURVA DE PARETO TRUNCADA

Na curva de Pareto truncada o número de pessoas com renda maior do que x é dado por

$$U = \frac{\Theta}{x^{\alpha}} - \frac{\Theta}{M^{\alpha}} = \Theta\left(\frac{1}{x^{\alpha}} - \frac{1}{M^{\alpha}}\right), \tag{3}$$

onde M representa, teoricamente, a renda máxima.

Fazendo

$$\frac{\Theta}{M^{\alpha}} = N,$$

obtemos

$$U = \frac{\Theta}{x^{\alpha}} - N \tag{4}$$

A renda média das pessoas com renda maior ou igual a x_o é dada por

$$\mu_o = \frac{1}{U_o}\int_{x_o}^{M} x\left(-\frac{dU}{dx}\right) dx,$$

com $U_o = \Theta\left(\frac{1}{x_o^{\alpha}} - \frac{1}{M^{\alpha}}\right)$

Segue-se que

$$\mu_o = \frac{1}{U_o}\int_{x_o}^{M} x\,\frac{\alpha\Theta}{x^{\alpha+1}}\, dx$$

Integrando e simplificando obtém-se, para $0 < \alpha \neq 1$.

$$\mu_o = \frac{\alpha}{\alpha-1} \cdot \frac{\dfrac{1}{x_o^{\alpha-1}} - \dfrac{1}{M^{\alpha-1}}}{\dfrac{1}{x_o^{\alpha}} - \dfrac{1}{M^{\alpha}}} \tag{5}$$

e, para $\alpha = 1$,

$$\mu_o = \frac{\ln \dfrac{M}{x_o}}{\dfrac{1}{x_o} - \dfrac{1}{M}} \tag{6}$$

É fácil ver que com $M \to \infty$ a expressão (5) se transforma em (2).

O uso das expressão (5) ou (6) para determinar a renda média exige a determinação prévia do valor de α.

3. DETERMINAÇÃO DO VALOR DO PARÂMETRO α DA CURVA DE PARETO TRUNCADA COM BASE EM DOIS PONTOS

Vamos considerar apenas os dois estratos de rendas mais altas. Trata-se, no caso, dos estratos cujos limites inferiores são $x_j = 10$ S.M. e $x_k = 20$ S.M.. Seja U_j o número de pessoas nos dois estratos, isto é, o número de pessoas com renda maior do que $x_j = 10$ S.M., e seja U_k o número de pessoas no último estrato. De acordo com (3) temos

$$U_j = \Theta\left(\frac{1}{x_j^{\alpha}} - \frac{1}{M^{\alpha}}\right)$$

$$U_k = \Theta\left(\frac{1}{x_k^{\alpha}} - \frac{1}{M^{\alpha}}\right)$$

Dividindo membro-a-membro, obtemos

$$\frac{U_j}{U_k} = \frac{\dfrac{1}{x_j^\alpha} - \dfrac{1}{M^\alpha}}{\dfrac{1}{x_k^\alpha} - \dfrac{1}{M^\alpha}} \qquad (7)$$

Uma vez que os valores de x_j, x_k, U_j e U_k são conhecidos, essa expressão permite determinar o valor de α, tendo-se previamente fixado o valor de M. A escolha do valor de M será discutida adiante.

Infelizmente (7) é uma equação em α sem solução explícita. Podemos escrever

$$F(\alpha) = U_j\left(\frac{1}{x_k^\alpha} - \frac{1}{M^\alpha}\right) - U_k\left(\frac{1}{x_j^\alpha} - \frac{1}{M^\alpha}\right) = 0$$

Considerando apenas os dois primeiros termos do desenvolvimento de $F(\alpha)$ de acordo com a série de Taylor, temos

$$F(\alpha_o) + \Delta\alpha\, F'(\alpha_o) = 0 \qquad (8)$$

onde α_o é uma estimativa preliminar de α e $\Delta\alpha = \alpha - \alpha_o$.

De (8) segue-se que

$$\Delta\alpha = -\frac{F(\alpha_o)}{F'(\alpha_o)} \qquad (9)$$

A estimativa preliminar de α pode ser obtida facilmente fazendo M = 0. Então

$$\alpha_o = \frac{\log U_j - \log U_k}{\log x_k - \log x_j} \tag{10}$$

Em seguida calcula-se o valor de $F(\alpha_o)$, de $F'(\alpha_o)$ e de $\Delta\alpha$. O novo valor de α será $\alpha_1 = \alpha_o + \Delta\alpha$. Utilizando esse valor corrigido calcula-se $F(\alpha_1)$, $F'(\alpha_1)$ e uma nova correção $\Delta\alpha$. O processo é repetido até que a correção possa ser considerada desprezível. É importante assinalar que o processo iterativo foi convergente em todos os casos empíricos analisados. No programa para computador que foi elaborado, o processo iterativo só era interrompido quando o valor absoluto da correção se tornava igual ou menor do que 10^{-5}.

Uma vez obtido o valor de α para os dois últimos estratos, calcula-se o valor de $N = \theta/M^{\alpha}$. Sejam x_i e x_{i+1} os limites inferior e superior de um estrato anterior aos dois últimos. Seja U_i o número de pessoas com renda maior do que x_i e seja U_{i+1} o número de pessoas com renda maior do que x_{i+1}. Então, de acordo com (4) temos que

$$U_i + N = \frac{\theta}{x_i^{\alpha}} \quad \text{e} \quad U_{i+1} + N = \frac{\theta}{x_{i+1}^{\alpha}}$$

Dividindo membro-a-membro e aplicando logaritmos, obtemos

$$\alpha = \frac{\log (U_i + N) - \log (U_{i+1} + N)}{\log x_{k+1} - \log x_i} \tag{11}$$

É claro que esse valor de α não será, necessariamente, igual ao valor obtido para os dois últimos estratos.

4. ESCOLHA DO VALOR DE <u>M</u>

Quando o valor de α é elevado, a renda média do último estrato não é muito afetada pelo valor de M. Com $x_o = 20$ e $\alpha = 2$, de (5) obtemos $\mu_o = 36{,}4$ para $M = 200$ e $\mu_o = 38{,}1$ para $M = 400$.

Entretanto, quando o valor de α é pequeno (ao redor de 1), a renda média do último estrato se mostra muito sensível ao valor de M. Com $x_o = 20$ e $\alpha = 1$, de (6) obtemos $\mu_o = 51,2$ para M = 200 e $\mu_o = 63,1$ para M = 400.

Inversamente, quanto mais alto for fixado o valor de M, maior será a variação do valor da renda média do último estrato em função de α. Se M = 200 o valor da renda média do último estrato para $\alpha = 1$ ($\mu_o = 51,2$) é 41% maior do que o valor para $\alpha = 2$ ($\mu_o = 36,4$). No entanto, se M = 400 o valor da renda média do último estrato para $\alpha = 1$ ($\mu_o = 63,1$) é 66% maior do que o valor para $\alpha = 2$ ($\mu_o = 38,1$). Assim, ao fixar o valor de M estaremos limitando a intensidade da variação do valor da renda média do último estrato em função do parâmetro α.

Após algumas experiências , fixamos o valor de M em 200 salários mínimos. Cabe ressaltar que isso não significa que acreditemos que esse seja o limite superior para as rendas individuais em 1980. Trata-se apenas de limitar, de uma maneira que nos pareceu razoável, a possibilidade de variação do valor da renda média do último estrato.

Todos os cálculos foram refeitos com M = 400. As rendas médias são um pouco maiores, crescem os índices de desigualdade, especialmente a desigualdade entre as três posições na ocupação, mas os resultados gerais da análise não são alterados.

Ao escolher o valor de M levou-se em consideração a coerência na agregação dos resultados, ou seja, verificou-se que com M = 200 a renda total estimada para o Brasil como um todo, era praticamente igual à renda total obtida somando os resultados obtidos aplicando o método aos dados de cada região do país ou somando os resultados obtidos aplicando o método a cada categoria de "posição na ocupação".

Note-se que a opção pelo valor M = 200 foi feita com base na análise de dados sobre a distribuição de renda entre a PEA na agricultura. É possível que um valor maior deva ser adotado ao se aplicar o método para a PEA dos setores urbanos (secundário e terciário).

5. OUTRAS CARACTERÍSTICAS DO MÉTODO

Sabe-se que a curva de Pareto só se ajusta bem para rendas relativamente elevadas. Assim, a curva de Pareto truncada só foi utilizada para os estratos acima de 3 S.M. na distribuição das pessoas economicamente ativas de acordo com seu rendimento mensal.

Para os seis primeiros estratos, cujos limites inferiores são 0, 0,25, 0,5, 1, 1,5 e 2 S.M., os rendimentos médios foram fixados em 0,16, 0,41, 0,80, 1,25, 1,75 e 2,5 S.M. Note-se que nos três primeiros estratos a média é fixada acima do ponto central e nos três estratos seguintes a média coincide com o ponto central. Nesses seis primeiros estratos a estimativa da desigualdade dentro dos estratos e a interpolação de percentis foi feita considerando, em cada estrato, uma distribuição com função de densidade linear (ver Hoffmann, 1979 e Hoffmann, 1984).

Nos estratos onde a renda média foi estimada por meio da distribuição de Pareto truncada, essa mesma distribuição é utilizada para estimar a desigualdade dentro do estrato e para interpolar percentis. A seguir, apresentamos as fórmulas necessárias.

A função de densidade de uma distribuição de Pareto truncada é

$$f(x) = \frac{\alpha\theta}{x^{\alpha+1}} \text{ , com } a \leq x \leq b \tag{12}$$

A correspondente função de distribuição é

$$p = F(x) = \theta\left(\frac{1}{a^{\alpha}} - \frac{1}{x^{\alpha}}\right) \tag{13}$$

O valor de θ deve ser tal que

$$\theta \left(\frac{1}{a^{\alpha}} - \frac{1}{b^{\alpha}} \right) = 1$$

Segue-se que

$$\theta = \frac{b^{\alpha}}{\rho^{\alpha}-1}, \tag{14}$$

onde $\rho = \dfrac{b}{a}$

A média da distribuição é

$$\mu = \frac{\alpha\, a}{\alpha-1} \cdot \frac{\rho^{\alpha}-\rho}{\rho^{\alpha}-1} \text{ para } 0 < \alpha \neq 1 \tag{15}$$

e

$$\mu = \frac{b}{\rho-1} \ln \rho \quad \text{para } \alpha = 1 \tag{16}$$

É claro que essas expressões são equivalentes a (5) e (6), bastando fazer $x_o = a$ e $M = b$.

De (13) e (14) obtemos

$$x = a \left[1 - p \left(1 - \frac{1}{\rho^{\alpha}} \right) \right]^{-\frac{1}{\alpha}} \tag{17}$$

Essa expressão permite determinar qualquer percentil da distribuição. A correspondente proporção acumulada da renda (a ordenada da curva de Lorenz) é dada por

$$\Phi(x) = \frac{b^{\alpha-1}}{\rho^{\alpha-1}-1}\left(\frac{1}{a^{\alpha-1}} - \frac{1}{x^{\alpha-1}}\right) \text{ para } 0 < \alpha \neq 1 \tag{18}$$

e

$$\Phi(x) = \frac{\ln(x/a)}{\ln\rho} \text{ para } \alpha = 1$$

O índice de Gini da distribuição de Pareto truncada, com α positivo e diferente de 1 e 1/2, é

$$G = \frac{2}{\rho^{\alpha}-1}\left(\rho^{\alpha} - \frac{\alpha-1}{2\alpha-1} \cdot \frac{\rho^{2\alpha-1}-1}{\rho^{\alpha-1}-1}\right) - 1 \tag{19}$$

Finalmente, a redundância da distribuição, com $0 < \alpha \neq 1$, é

$$R = \frac{1}{\alpha-1} - \ln\left(\frac{\alpha}{\alpha-1} \cdot \frac{\rho^{\alpha}-\rho}{\rho^{\alpha}-1}\right) - \frac{\ln\rho}{\rho^{\alpha-1}-1} \tag{20}$$

6. ANÁLISE DA DISTRIBUIÇÃO DA RENDA ENTRE PESSOAS ECONOMICAMENTE ATIVAS NA AGROPECUÁRIA, NO BRASIL, EM 1980

Para ilustrar a aplicação do método proposto, apresentamos, nesta seção, resultados obtidos com dados do Censo Demográfico de 1980, em que as pessoas ativas na agropecuária, extração vegetal e pesca são classificadas em 10 estratos de acordo com o rendimento de todas as ocupações. Foram excluídas as pessoas ativas sem rendimento.

A Tabela 1 mostra os valores obtidos para algumas medidas de tendência central e de desigualdade no Brasil e em 6 regiões. Trata-se de 10.397 mil pessoas ativas na agricultura, recebendo algum rendimento em suas ocupações. Aplicando o método para os dados do Brasil como um todo, a renda total estimada foi 16.040 milhares de S.M., ao passo que a soma das rendas totais estimadas para cada região foi 16.048 milhares de S.M. O erro de agregação é de apenas 0,05%.

Tabela 1

RENDA MÉDIA (μ) E RENDA MEDIANA (D), EM SALÁRIOS MÍNIMOS, ÍNDICE DE GINI (G), ÍNDICE DE THEIL (T) E PORCENTAGENS DA RENDA TOTAL CORRESPONDENTES AOS 50% MAIS POBRES (50^-), AOS 10% MAIS RICOS (10^+) E AOS 5% MAIS RICOS (5^+) PARA A DISTRIBUIÇÃO DA RENDA ENTRE A PEA AGROPECUÁRIA, EXCLUSIVE OS SEM RENDIMENTO, CONSIDERANDO OS RENDIMENTOS DE TODAS AS OCUPAÇÕES, NO BRASIL E SUAS REGIÕES, EM 1980.

Região	μ	D	G	T	50^-	10^+	5^+
Norte	1,66	1,20	0,410	0,325	24,1	33,8	24,0
Nordeste	0,92	0,72	0,441	0,404	22,1	36,4	27,4
Sudeste excl. SP	1,71	0,93	0,544	0,559	18,4	49,1	39,1
São Paulo	2,24	1,17	0,560	0,589	17,7	51,3	41,7
Sul	2,24	1,13	0,574	0,554	15,6	49,1	38,0
C. Oeste	2,06	1,08	0,557	0,570	17,8	50,1	40,1
Brasil	1,54	0,90	0,542	0,547	17,9	47,4	37,3

Verifica-se, na Tabela 1, que as regiões com maior renda média e mediana são o Sul e São Paulo. A região mais pobre é o Nordeste, com 42,4% das pessoas e apenas 25,4% da renda total.

Note-se que há uma correlação positiva entre os valores das medidas de tendência central e o grau de desigualdade.

Verifica-se que da redundância total da desigualdade no país apenas 8,7% correspondem à redundância entre as 6 regiões.

A Tabela 2 mostra os resultados obtidos para as pessoas classificadas em três posições na ocupação principal: empregado, empregador e autônomo. Tratam-se de 10.388 mil pessoas, das quais 46,1% são empregados, 3,1% são empregadores e 50,8% são autônomos. O total de pessoas é menor do que o considerado na Tabela 1 porque foram excluídas as pessoas não classificadas em uma das três categorias.

Tabela 2

RENDA MÉDIA (μ) E RENDA MEDIANA (D), EM SALÁRIOS MÍNIMOS, ÍNDICE DE GINI (G), ÍNDICE DE THEIL (T) E PORCENTAGENS DA RENDA TOTAL CORRESPONDENTES AOS 50% MAIS POBRES (50^-), AOS 10% MAIS RICOS (10^+) E AOS 5% MAIS RICOS (5^+) PARA A DISTRIBUIÇÃO DA RENDA ENTRE A PEA AGROPECUÁRIA, EXCLUSIVE OS SEM RENDIMENTO, CONSIDERANDO OS RENDIMENTOS DE TODAS AS OCUPAÇÕES, NO BRASIL, EM 1980, CONFORME A POSIÇÃO NA OCUPAÇÃO PRINCIPAL

Posição na ocupação	μ	D	G	T	50^-	10^+	5^+
Empregado	0,99	0,86	0,349	0,233	27,4	28,2	19,2
Empregador	9,48	3,62	0,663	0,597	9,5	55,0	40,5
Autônomo	1,58	0,93	0,534	0,486	17,3	44,7	33,0
Total	1,54	0,90	0,542	0,547	17,9	47,4	37,3

Aplicando o método para o número total de pessoas nas três categorias, a renda total estimada foi 15.967 milhares de S.M., ao passo que a soma das rendas totais estimadas para cada categoria foi 16.053 milhares de S.M. O erro de agregação é de 0,5%.

Como seria de se esperar, as rendas médias e medianas são mais altas para a categoria dos empregadores e mais baixas para a categoria dos empregados.

A categoria dos empregadores certamente inclui, além dos empresários agrícolas propriamente ditos, pequenos agricultores que apesar de contratarem assalariados em caráter eventual ou complementar, declararam ser essa sua posição na ocupação principal. A presença de empregadores relativamente pobres ao lado de grandes empresários capitalistas na mesma categoria reflete-se no maior grau de desigualdade da distribuição de renda (G = 0,663), especialmente em comparação com os empregados, que mostram maior homogeneidade de nível de rendimentos (G = 0,349).

Apesar da forte desigualdade dentro da categoria dos empregadores, a decomposição da redundância mostra que a desigualdade entre categorias explica 26,8% da redundância total. Em outro trabalho mostramos que esse componente é mais importante nas regiões de agricultura mais desenvolvida (ver Hoffmann e Kageyama, 1986). Para a PEA agrícola de São Paulo a desigualdade entre as três categorias corresponde a 40,7% da redundância total.

BIBLIOGRAFIA CITADA

HOFFMANN, R. Estimação da Desigualdade dentro dos Estratos no Cálculo do Índice de Gini e da Redundância. *Pesquisa e Planejamento Econômico* 9(3): 719-738, 1979.

HOFFMANN, R. Estimation of Inequality and Concentration Measures From Grouped Observations. *Revista de Econometria* 4(1):5-21, 1984.

HOFFMANN, R. & KAGEYAMA, A.A. Posição na Ocupação, Tempo de Trabalho e Renda na Agricultura Brasileira em 1980. Trabalho apresentado ao VII *Encontro Brasileiro de Econometria*, em Vitória, ES, dez. de 1985 (pp.213-231 dos anais), 1985 e 1986. Uma nova versão do trabalho será publicada na *Revista de Economia Política*, 1986.