

# Gastos com saúde: uma análise por domicílios para a cidade de São Paulo\*

DENISARD ALVES\*\*

*Este artigo procura responder à seguinte questão: o que ocorre com os gastos com saúde quando a população de um país em desenvolvimento envelhece? Uma implicação dessa análise é que quanto mais idosa a população maior a demanda pelo sistema público de saúde. A análise baseia-se em uma pesquisa domiciliar sobre consumo para a cidade de São Paulo, a maior do Brasil. Os procedimentos de estimação consistem em métodos paramétricos, como Ordinary Least Squared, Tobit Censored Regression, e Least Absolute Deviation, um método semiparamétrico. Ao contrário do Ordinary Least Squared e do Tobit Censored Regression, o Least Absolute Deviation não requer a hipótese de normalidade da distribuição dos erros para que o estimador seja consistente. A hipótese de um impacto positivo da idade sobre os gastos com saúde é sustentada pelos dados com um estimador que apresenta propriedades de grandes amostras.*

## 1 - Introdução

No Brasil existe um sistema de saúde dual. Pessoas com recursos financeiros, ou cujo empregador proporciona assistência médica, têm acesso ao sistema de saúde privado que atende, com qualidade, à demanda. O resto do país, ao contrário, conta com um sistema público de clínicas e hospitais. E, assim, como na maioria dos casos de sistemas públicos de saúde no mundo, o sistema brasileiro é caracterizado por uma longa espera e uma qualidade questionável, com a implicação prática de que aqueles que dependem desse sistema nunca conseguem resolver razoavelmente seus problemas de saúde e, conseqüentemente, passam mais tempo doentes.

Esse sistema de saúde dual é particularmente relevante, no que diz respeito ao Brasil, à luz das recentes mudanças na estrutura sociodemográfica do país. Em 1990, apenas 6,7% da população brasileira tinham idade acima de 60 anos. Contudo, para 2010, a expectativa é de que essa parcela seja de 9,7% e, para 2030, de 16,9% [World Bank (1994, 1998 e 2000)]. Durante os últimos 20 anos, o tamanho das famílias dentro dos segmentos mais pobres da sociedade brasileira

---

\* O autor agradece o apoio financeiro do Nemesys, um projeto CNPq-Pronex, e da Fundação Instituto de Pesquisas Econômicas (Fipe). Agradece também os comentários recebidos a uma versão anterior deste artigo no Seminário IPE-USP, bem como os comentários e sugestões apresentados por pareceristas anônimos. Erros remanescentes são de responsabilidade do autor.

\*\* Professor de Economia da USP.

(aqueles que tipicamente dependem do serviço público de saúde) tem sido maior do que o dos segmentos mais prósperos. Esse grande grupo da população está envelhecendo, e está se aproximando o momento em que suas necessidades de assistência médica irão crescer rapidamente [Cutler e Meara (1998)]. Cogita-se, com certa preocupação, de que o sistema público de saúde brasileiro não irá atender a essa crescente demanda. Em particular, o já longo tempo de espera continuará a aumentar, implicando que muitos dos segmentos mais pobres da sociedade não irão receber tratamento algum. Esse mecanismo de exclusão social da população pobre e idosa crescerá na mesma proporção em que cresce esse segmento da população brasileira.

A cidade de São Paulo tem, aproximadamente, 10 milhões de habitantes. Ela pode ser vista como a representação, em menor escala, da maioria dos problemas de saúde que afetam a população do Brasil. Uma pesquisa recente coletou dados sobre os padrões de consumo de 2.200 domicílios na região, o que leva a um total de 2,5 milhões de moradores.<sup>1</sup> Dentre as variáveis coletadas, estão gasto com saúde, idade e educação de cada membro do domicílio, renda e despesas em bens e outros serviços. Os resultados dessa pesquisa fornecem uma base útil para a análise do impacto dessas variáveis sobre o gasto com saúde na região, potencialmente incluindo evidências a serem buscadas, mais adiante, em pesquisas específicas sobre o custo com saúde para diferentes estratos da população.

O uso de pesquisas domiciliares apresenta alguns problemas econométricos importantes que devem ser tratados para que se possa obter resultados estatisticamente relevantes. Tais problemas incluem censura e heterocedasticidade. Quando a análise combina os dois, o procedimento de estimação deve ser realizado de modo a levar em conta variáveis dependentes limitadas, devido à censura, e muito provavelmente heterocedasticidade, devido à natureza dos dados.

Este artigo está organizado em sete seções, incluindo esta introdução. A Seção 2 discute os dados coletados na pesquisa em São Paulo. A Seção 3 apresenta o modelo usado para explicar os gastos do domicílio com saúde. A Seção 4 discute os procedimentos de estimação, enfatizando as questões econométricas de censura e heterocedasticidade. As Seções 5 e 6 apresentam os resultados para os diferentes métodos de estimação. A Seção 7 apresenta as conclusões do estudo.

## 2 - Os dados

A pesquisa domiciliar de São Paulo cobre grande parte da área urbana do estado. A amostra engloba 10 milhões de pessoas, enquanto a população total do estado é de 30 milhões e a urbana, de 25 milhões. Os dados cobrem aproximadamente 400 mil itens de consumo. Serviços de saúde usados pelos moradores e medica-

---

<sup>1</sup> Pesquisa realizada pela Fipe para a construção dos pesos do Índice de Custo de Vida da cidade de São Paulo.

mentos comprados por eles em um mês específico de 1998 são reportados com bastante detalhes.

Gasto com saúde e características como renda do domicílio, despesa total em bens e serviços, e idade e educação de cada membro foram extraídos dos dados da pesquisa. Contudo, uma vez que a pesquisa não coletou dados sobre os preços dos serviços de saúde, fomos forçados a conduzir uma análise sobre o gasto em vez de uma típica análise de demanda. A Tabela 1 apresenta as médias para as variáveis usadas no estudo.

A proporção do gasto com saúde em relação à despesa total do domicílio foi de 7,4% para a amostra como um todo. Entretanto, essa proporção sobe para 11,1% quando as observações censuradas em zero para GASTO são excluídas. O gasto com saúde é relativamente alto como proporção da despesa total. Essa é uma informação importante porque implica que pequenas mudanças nos preços desses serviços afetam a renda real do domicílio em uma proporção significativa. A média de idade dos chefes das famílias é de 45,48 anos e a escolaridade média é de 7,37 anos. O quantil 0,25 é zero, devido ao fato de uma grande percentagem dos domicílios não apresentar nenhum valor para as despesas com saúde.<sup>2</sup>

TABELA 1

*Estatísticas das variáveis*

	GASTO	IDADE	CRIANÇA	IDOSO	TAMFAM	RENAFAM	EDUC
Média	108,640	45,48	0,374	0,239	3,37	1.838,74	7,37
Desvio-padrão	244,490	14,72	0,484	0,426	1,64	2.273,95	2,59
0,25 quantil	0	34	0	0	2	625,39	5
0,5 quantil	18,775	44	0	0	3	1.100	7
0,75 quantil	117,490	55,75	1	0	4	2.200	9

OBS.: GASTO = gasto com saúde; RENAFAM = renda da família; IDADE = idade do chefe da família; CRIANÇA = variável *dummy* indicando a presença de crianças menores de 14 anos que moram no domicílio; IDOSO = *dummy* que indica a presença de pessoas acima de 60 anos de idade no domicílio; TAMFAM = número de pessoas vivendo no domicílio; e EDUC = quantidade de educação do chefe da família, em anos de estudo.

### 3 - O modelo: dados censurados e solução de canto

A análise de regressão censurada é usualmente aplicada em problemas que são conceitualmente diferentes. Tais problemas podem ser colocados em duas categorias. O primeiro caso aparece quando a variável dependente, representando o que se quer explicar, é uma variável com significado quantitativo, mas ocorre um problema nos dados porque ela é censurada abaixo ou acima de algum valor, isto é, não é observa-

2 Esse fato acarreta vários problemas de estimação que serão discutidos no decorrer do artigo.

da para parte da população. Por exemplo, dados sobre renda do trabalho, em muitas pesquisas, normalmente são registrados até um valor-limite, mas acima desse valor é registrado apenas o fato de que a renda é maior do que esse limite.

Um segundo tipo de implicação dos modelos de regressão censurada se encaixa no problema que se quer estudar e no qual, infelizmente, o rótulo “regressão censurada” é menos apropriado. No caso aqui estudado, alguns domicílios não gastam dinheiro com saúde, ou porque não houve necessidade, porquanto os moradores são saudáveis, ou porque eles sempre terão um último recurso proporcionado pelo sistema público de saúde. Essa é uma solução na qual a observação do zero no gasto com saúde para alguns domicílios é o resultado esperado da solução de um problema de otimização. Gasto zero com saúde é a escolha ótima e é uma solução de canto. Esse tipo de variável de resposta é usualmente chamado de modelo de regressão censurada, embora seja bem diferente do problema de observação de dados relacionado com a classificação discutida no exemplo do parágrafo anterior.

No nosso caso, o problema não é de observação dos dados, mas que esses dados são resultantes da solução de um problema de otimização de parte dos indivíduos racionais e, assim, a solução será uma solução de canto.<sup>3</sup> Este estudo se caracteriza por uma solução de canto; estamos interessados na média da distribuição de  $y$ , gasto com saúde *per capita*, dado  $x$ ,  $E(y/x)$ . Quer dizer, estamos interessados em analisar o impacto das variáveis idade, educação, renda etc. no valor esperado do gasto *per capita* com saúde.

#### 4 - Procedimentos de estimação

Um caminho para se atingir esse objetivo é estimar o impacto de mudanças em  $X_k$  na resposta média.<sup>4</sup> Admite-se  $E(y/x)$

$$y_i = x_i \beta + u_i \quad (1)$$

levando a:

$$E(y_i / x) = x_i \beta \quad (2)$$

onde  $y_i$  é o logaritmo do gasto com saúde do domicílio  $i$  dividido pelo número de pessoas que vivem no domicílio<sup>5</sup> e  $x_i$  é o conjunto de variáveis explicativas usado

---

3 Observações censuradas não caracterizam corretamente a situação; entretanto, essa é a forma usual encontrada na literatura para o banco de dados resultante de uma solução de canto [ver Wooldridge (2001, p. 518)].

4  $k = 1, \dots, K$  representa o número de colunas da matriz  $x$ .

5 Devido à presença de valores iguais a zero para o gasto com saúde,  $y = \log(1 + \text{GASTO})$  e GASTO será sempre considerado em termos *per capita*.

nas regressões, como IDADE, RENDAFAM,<sup>6</sup> TAMFAM, CRIANÇA, IDOSO, EDUC, bem como as transformações e interações das variáveis para levar em conta possíveis não-linearidades e para aplicar Ordinary Least Square (OLS) para estimar o parâmetro  $\beta$ . Deixando de lado por um momento a censura, e admitindo as hipóteses da regressão linear-padrão, OLS pode ser usado para estimar o vetor de parâmetros  $\beta$  da equação (1), incluindo na matriz  $x_i$  a variável IDADE-ao-quadrado com vistas a capturar um possível comportamento não-linear da idade como variável explicativa. Também é admitido que  $u_i$  é normalmente distribuído com média zero e variância constante, hipóteses clássicas da regressão linear. Sob essas hipóteses, OLS apresenta as propriedades desejadas [ver Johnston e DiNardo (1997)]. Os resultados para o estimador de OLS estão apresentados na primeira coluna da Tabela 2.

TABELA 2

*Resultados dos modelos OLS e Tobit*

Y	OLS	Tobit	Impacto marginal – Tobit
IDADE	0,00490 (0,0187)	0,01970 (0,0231)	0,01440 (0,0168)
IDADE2	-0,00006 (0,0002)	-0,00007 (0,0002)	-0,00005 (0,0002)
EDUC	0,05800 (0,0183)**	0,06220 (0,0225)**	0,04540 (0,0164)**
LOG(RENDAFAM)	1,13280 (0,0503)**	1,38510 (0,0648)**	1,01080 (0,0473)**
CRIANÇA	0,25080 (0,1119)*	0,36970 (0,1376)**	0,27220 (0,1004)**
IDOSO	0,67940 (0,1530)**	0,80950 (0,1839)**	0,61400 (0,1342)**
TAMFAM	-0,12040 (0,0321)**	-0,01490 (0,0396)	-0,01090 (0,0289)
CONST.	-5,77780 (0,4763)**	-8,34640 (0,6104)**	-6,09080 (0,4454)**
ESCALA		2,27720 (0,0469)	
$R^2$	0,32310	0,31360	

\*Significativo ao nível de 5%.

\*\*Significativo ao nível de 1%.

6 Em todos os procedimentos de estimação RENDAFAM será definida como a renda do domicílio dividida pelo número de pessoas que nele vivem.

Os resultados de OLS indicam uma relação não-significativa entre o gasto com saúde e a idade do chefe da família. À medida que a renda do domicílio cresce, o gasto com saúde também aumenta. As estimativas de OLS indicam uma elasticidade de 1,13 entre o gasto *per capita* com saúde e a renda *per capita* da família. Os resultados de OLS também mostram alguma economia de escala no gasto com saúde: um membro adicional no domicílio leva a um decréscimo de 0,12% no gasto *per capita* com saúde. A presença de uma pessoa idosa e de crianças no domicílio aumenta, de modo significativo, a despesa *per capita* com saúde.

A estimativa de OLS fornece evidência sustentando que a idade em ambas as caudas da sua distribuição — crianças e pessoas idosas — leva a um maior gasto *per capita* com saúde para o domicílio. Escolaridade se correlaciona positivamente com o gasto com saúde: quanto maior a educação do chefe da família, maior é o gasto com saúde. Esses resultados podem mostrar que domicílios com maior renda e educação confiam mais nos serviços privados de saúde. Embora os resultados por OLS contenham uma história, eles devem ser considerados com certo cuidado devido à censura existente na amostra.<sup>7</sup>

Dados de uma amostra aleatória de uma população, na qual a solução de canto é uma possibilidade concreta, requerem um modelo diferente para explicar o gasto médio *per capita* com saúde. Isso é discutido na Seção 5.

## 5 - Modelo de regressão censurada

O exemplo clássico de censura é o estudo de Tobin (1958) sobre despesas domésticas, em que o autor apresenta o modelo de análise conhecido atualmente por modelo Tobit. O modelo censurado-padrão Tobit<sup>8</sup> é mais facilmente definido como um modelo de variável latente:

$$\dot{y} = x\beta + u, \quad u|x \sim N(0, \sigma^2) \quad (3)$$

$$y = \max(0, \dot{y}) \quad (4)$$

A variável latente  $\dot{y}$  satisfaz às suposições do modelo linear clássico; ela possui uma distribuição normal homocedástica com média condicional linear. A equação (4) requer que a variável observada,  $y$ , seja igual a  $\dot{y}$  quando  $\dot{y} \geq 0$ , mas

---

7 A presença de variável dependente limitada acarreta estimativas de OLS ineficientes. OLS também será viesado e inconsistente na presença de censura [ver Deaton (1997, p. 87) e Johnston e DiNardo (1997, p. 445)].

8 Também chamado de modelo Tobit tipo I por Amemiya (1985). O modelo censurado-padrão Tobit, descrito nas equações (1) e (2), é consistente com o típico problema de observação dos dados bem como com o modelo de solução de canto [ver Wooldridge (2001, p. 519)].

$y = 0$  quando  $\hat{y} < 0$ ;  $y$ , gasto com saúde, é zero para uma fração significativa da população. Como  $\hat{y}$  possui uma distribuição normal,  $y$  possui uma distribuição contínua para valores estritamente positivos. Em particular, a densidade de  $y$ , dado  $x$ , é a mesma densidade de  $\hat{y}$ , conhecido  $x$ , para valores positivos. Além disso,

$$\begin{aligned} P(y=0|x) &= P(\hat{y} < 0|x) = P(u < -x\beta) = \\ &= P(u/\sigma < -x\beta/\sigma) = \Phi(-x\beta/\sigma) = 1 - \Phi(x\beta/\sigma)^9 \end{aligned} \quad (5)$$

Considerando que a amostra é uma retirada aleatória da população, e dada a densidade de  $y_i$ , dado  $x_i$ , a função log-verossimilhança para cada observação pode ser obtida e as estimativas de máxima verossimilhança de  $\beta$  e  $\sigma$  são conseguidas por meio da maximização da função log-verossimilhança [ver Wooldridge (2000, p. 541)].

A contrapartida de (5) para a amostra é:

$$P(y_i = 0|x_i) = 1 - \Phi(x_i\beta) \quad (6)$$

onde  $(y_i, x_i)$  são as observações amostrais da variável dependente  $y$  e do vetor de variáveis explicativas  $x$ , sendo  $i = 1, \dots, n$ .<sup>10</sup>

O modelo Tobit é bastante diferente do modelo clássico de regressão linear. Ele usa toda a informação, incluindo a relacionada à censura, e fornece estimativa consistente dos parâmetros. É uma combinação de um modelo Probit com o de um modelo clássico de regressão linear.<sup>11</sup> Por meio da equação (3) vemos que  $\beta_k$  mede o efeito parcial de  $x_k$  em  $E(\hat{y}|x)$ , onde  $k$  representa as variáveis explicativas, utilizadas na explicação do gasto com saúde por domicílio na cidade de São Paulo. Desse modo, as estimativas de máxima verossimilhança do modelo Tobit, obtidas por meio da maximização da função de verossimilhança com respeito ao vetor de parâmetros  $\beta$ , estão apresentadas na segunda coluna da Tabela 2.

Os resultados da estimação do Tobit mostram novamente que EDUC, LOG (REDAFAM), CRIANÇA e IDOSO são estatisticamente significativos na explicação do gasto com saúde.<sup>12</sup> A variável EDUC possui um efeito positivo, embora nenhuma conclusão sobre a magnitude do efeito possa ser observada

9 Esse resultado é consequência de se admitir  $u/\sigma$  para se ter uma distribuição normal-padrão e  $\Phi$  é a distribuição normal acumulada [ver Wooldridge (2000, p. 540-541)].

10 A amostra de domicílios nesse estudo é de 2.200.

11 Os modelos estruturais para Tobit e Probit são os mesmos, mas a medida dos modelos difere. Para uma boa discussão acerca dessas similaridades, ver Long (1987, p. 199).

12 A interpretação do impacto das variáveis explicativas no gasto com saúde difere da interpretação da estimação por OLS.

devido à diferente natureza da relação existente entre o coeficiente estimado e a variável dependente.<sup>13</sup> Com base nas medidas do *R*-quadrado, a função da média condicional de OLS se encaixa nos dados um tanto melhor, ainda que a diferença seja pequena.<sup>14</sup> O parâmetro mostra a relação entre o valor esperado da variável não-observada  $\dot{y}$  e o  $k^{\text{th}}$  vetor coluna matriz  $x$ .

$$\frac{\partial E(\dot{y}|x)}{\partial x_k} = \hat{\beta}_k \quad (7)$$

O valor esperado da variável observada  $y$  e da  $x_k$  variável é dado por:

$$\frac{\partial E(y|x)}{\partial x_k} = \Phi\left(\frac{x\hat{\beta}}{\sigma}\right)\beta_k \quad (8)$$

Desse modo, o coeficiente estimado indica o impacto de uma mudança na variável explicativa sobre a média condicional esperada da observação maior do que zero.<sup>15</sup> Os resultados da estimação, normalmente chamados de média não-condicional esperada,<sup>16</sup> embora condicionada às variáveis explicativas, são apresentados na Tabela 3.

O impacto da estimação do modelo Tobit pode ser visto a partir do seguinte exemplo. Os valores previstos para a variável latente, condicionados às variáveis explicativas, foram estimados, incluindo as observações censuradas. A média amostral da variável latente prevista foi calculada em R\$ 237,99. A Tabela 1, já apresentada, revela uma média amostral para o GASTO observado de R\$ 108,64. A comparação dos dois valores mostra que o aumento potencial no gasto com saúde é de 119%, devido a um aumento no gasto com saúde dos domicílios que reportavam um gasto igual a zero.

O modelo de regressão censurada é um aperfeiçoamento quando comparado com OLS. Entretanto, o estimador Tobit se baseia na variância homocedástica do termo de erro  $u$ . A Seção 6 discute um teste do modelo Tobit para a presença de heterocedasticidade.

13 Ver McDonald e Moffit (1980, p. 318-321), em que uma admirável interpretação da relação entre os parâmetros é apresentada.

14 *R*-quadrado para a estimação por Tobit é definido como o quadrado do coeficiente de correlação entre o valor previsto para  $y$  pelo Tobit e o valor observado de  $y$ . É importante lembrar que a estimação por Tobit não é feita de modo a maximizar o *R*-quadrado — ela maximiza a função de verossimilhança —, enquanto a estimação por OLS produz o maior *R*-quadrado devido à forma da função linear para a média condicional.

15 Para mais detalhes, ver McDonald e Moffit (1980).

16 Ela é normalmente chamada de não-condicional, pois não é condicional aos valores de  $y > 0$ , entretanto ela é condicional a  $x$ .



TABELA 3

*Resultados do estimador LAD*

<i>Y</i>	LAD	LAD (com desvios-padrão <i>bootstrapped</i> )
IDADE	0,0433 (0,0290)	0,0433 (0,0271)
IDADE2	-0,0003 (0,0003)	-0,0003 (0,0003)
EDUC	0,0570 (0,0276)*	0,0570 (0,0292)
LOG(RENDAFAM/TAMFAM)	1,5806 (0,0882)**	1,5806 (0,0922)**
CRIANÇA	0,3259 (0,1678)	0,3259 (0,1827)
IDOSO	0,9730 (0,2209)**	0,9730 (0,2320)**
TAMFAM	-0,0100 (0,0496)	-0,0100 (0,0514)
CONST.	-9,5972 (0,8042)**	-9,5972 (0,8417)**

\*Significativo ao nível de 5%.

\*\*Significativo ao nível de 1%.

## 6 - Regressão censurada e heterocedasticidade

OLS é ineficiente sob heterocedasticidade. Em uma regressão censurada, o modelo OLS é também viesado e inconsistente. O modelo Tobit é inconsistente sob heterocedasticidade [ver Deaton (1997, p. 85-87)]. Os coeficientes estimados são viesados mesmo quando o tamanho da amostra cresce para infinito. As previsões estão fora do intervalo mesmo para grandes amostras.

### 6.1 - Modelo paramétrico heterocedástico

Usando o valor estimado da função de verossimilhança obtida, testamos o modelo Tobit para heterocedasticidade por meio da maximização de uma função de verossimilhança restrita,<sup>17</sup> na qual o padrão de heterocedasticidade é dado por:

$$\sigma_i^2 = \sigma^2 \exp(x_i \gamma) \quad (9)$$

17 O modelo é estimado usando um estimador ML restrito, em que a função de verossimilhança é maximizada sob a restrição  $S_i^2 > 0$ . Para uma discussão do método de estimação da máxima verossimilhança restrita, ver Schoenberg (1995).

Os resultados da estimação do modelo são apresentados na Tabela 4.

Os resultados apresentados na Tabela 4, intitulados Equação 2, indicam um padrão de heterocedasticidade nos dados. O teste razão de verossimilhança também foi usado para confirmar os resultados.<sup>18</sup> Esse teste apresenta como resultado rejeição à hipótese de homocedasticidade.

Os resultados do teste apresentados na Tabela 4 mostram que a hipótese nula não é sustentada pelos dados. A presença de heterocedasticidade acarreta estimadores inconsistentes dos parâmetros para o estimador de máxima verossimilhança do modelo Tobit.

TABELA 4

*Máxima verossimilhança restrita do modelo paramétrico heterocedástico*

Variáveis	Equação 1 Variável dependente: $y$	Equação 2 Variável dependente: $\sigma^2$
IDADE	0,0140 (0,0231)	0,0222 (0,0205)
IDADE2	-0,00001 (0,0002)	-0,0003 (0,0002)
EDUC	0,0644 (0,0225)**	-0,0224 (0,0203)
LOG(RENDAFAM)	1,4066 (0,0649)**	-0,0932 (0,0547)
CRIANÇA	0,0375 (0,1394)**	-0,1018 (0,1240)
IDOSO	0,7063 (0,1882)**	0,1150 (0,1673)
TAMFAM	0,0132 (0,0352)	-0,1230 (0,0331)**
CONST.	-8,49590 (0,6315)**	3,0858 (0,5192)**

\*Significativo ao nível de 5%.

\*\*Significativo ao nível de 1%.

<sup>18</sup> Para uma excelente discussão sobre um teste de heterocedasticidade para o modelo linear censurado normal, ver Ghesher e Irish (1987, p. 33-61).

## 6.2 - Estimador LAD

Um estimador semiparamétrico proposto por Powell é uma alternativa ao modelo Tobit quando há a presença de heterocedasticidade.<sup>19</sup> Considere-se novamente o modelo de variável latente, mas no qual a *mediana* de  $u$ , dado  $x$ , é zero:

$$\dot{y} = x\beta + u \quad \text{Med}(u|x) = 0 \quad (10)$$

Essa equação implica que  $\text{Med}(\dot{y}|x) = x\beta$ , tal que a mediana de  $\dot{y}$  é linear em  $x$ . O estimador de Powell foi proposto como um estimador semiparamétrico robusto que não depende de nenhuma outra hipótese sobre distribuição. O estimador de Powell é  $\sqrt{N}$  consistente para o típico problema de dados censurados, mas também pode ser aplicado em problemas de solução de canto [ver Wooldridge (2001, p. 536)]; entretanto, a diferença entre a mediana condicional de  $y$  e a mediana condicional desse estimador torna-se crucial. A  $\text{Med}(y|x)$  não depende da distribuição de  $u$ , dado  $x$ , da qual  $E(y|x)$  e  $E(y|x, y > 0)$  dependem [ver Deaton (1997, p. 85-87)].

O procedimento de Powell é o chamado estimador Least Absolute Deviation (LAD) tornado operacional por um algoritmo iterativo proposto por Buchinsky.<sup>20</sup> O estimador de Powell consiste em minimizar a seguinte soma de resíduos absolutos:

$$\min_{\beta} \sum_{i=1}^{\infty} |y - \max(0, x\beta)| \quad (11)$$

Powell mostra que a consistência desse estimador não requer conhecimento da distribuição dos  $u$ s e tampouco admite que a distribuição é homocedástica, isto é, considera apenas que a mediana é zero [ver Deaton (1997, p. 89), Powell (1984) e Buchinsky (1994)].

Os resultados do procedimento LAD, usando para a estimação o algoritmo proposto por Buchinsky, são apresentados na coluna LAD, na Tabela 3.

Entretanto, o estimador dos erros-padrão para os coeficientes avaliados pela regressão mediana não está correto [ver Rogers (1992, p. 133-137)]. Logo, usamos um procedimento de *bootstrap* para obter uma estimação consistente dos erros-padrão. Os resultados estão apresentados na Tabela 3.

---

19 Powell (1984, p. 303-325) recomenda o uso de um estimador robusto para avaliar o modelo de regressão censurada.

20 Para maiores detalhes, ver Buchinsky (1994, p. 405-458) e Alves (1999).

Os erros-padrão obtidos pelo procedimento de *bootstrap* mudam o significado do estimador da regressão mediana para a educação do chefe da família. A idade do chefe da família e a idade ao quadrado continuam não-significativas. Entretanto, a variável *dummy* para a presença de idosos no domicílio é significativa na estimação por LAD, bem como quando os erros-padrão *bootstrapped* são usados. Os resultados do estimador LAD não alteraram alguns dos resultados do estimador Tobit. A idade do chefe do domicílio não produziu um coeficiente significativo, confirmando a conclusão obtida por meio da estimação por Tobit. A variável *dummy* para a presença de idosos no domicílio continuou positiva e altamente significativa. Em todas as idades está, então, inquestionavelmente, um fator que aumenta o gasto com saúde no domicílio (Tabela 5).

TABELA 5

*Impacto marginal das variáveis explicativas sobre o gasto com saúde*

	OLS	LAD	Tobit <sup>a</sup>
IDADE	0,0049	0,0433	0,0144
EDUC	0,0580**	0,0570	0,0454**
LOG(RENDAFAM)	1,1328**	1,5806**	1,0108**
CRIANÇA	0,2508*	0,3259	0,2722**
IDOSO	0,6794**	0,9730**	0,6140**
TAMFAM	-0,1204**	-0,0100	-0,0109
CONST.	-5,7778**	-9,5972**	-6,0908**

<sup>a</sup> No modelo Tobit o efeito marginal consiste na avaliação de  $\frac{\partial E(y/x)}{\partial x_k}$  nos valores médios amostrais das variáveis explicativas usando os coeficientes estimados por Tobit apresentados na Tabela 2.

\*Significativo ao nível de 5%.

\*\*Significativo ao nível de 1%.

## 7 - Conclusão

Algumas conclusões interessantes podem ser derivadas a partir da análise dos resultados das estimações. A primeira delas está relacionada à estimação da regressão censurada. Existem duas explicações possíveis para a censura nos dados: a) a família não realizou gasto algum com saúde; e b) os problemas de saúde foram tratados pelo sistema público de saúde. Ambas as explicações são consistentes com a solução de canto como um possível resultado do comportamento maximizador do consumidor sob o preço zero do serviço público de saúde. O resultado pode mostrar que uma proporção substancial da população conta com o setor público para a provisão de serviços de saúde. Uma advertência precisa ser feita aqui: devido à amostra usada, é difícil separar as pessoas saudáveis das doentes que foram atendidas pelo setor público de saúde.

A idade é um fator importante para explicar os gastos com saúde. A presença de crianças no domicílio como um fator que aumenta os gastos com saúde é sustentada pelos resultados de OLS e Tobit. A presença de idosos no domicílio leva a um gasto maior com saúde. Esse resultado prevalece em todos os procedimentos de estimação utilizados na análise. À medida que a estrutura etária muda, a presença de idosos cresce, devido ao aumento da expectativa de vida, colocando pressão sobre a renda familiar e sobre o setor público como último recurso para a provisão de serviços de saúde. Por outro lado, famílias de baixa renda têm um número grande de filhos. Portanto, famílias pobres com um elevado número de membros, seja em virtude da presença de crianças, seja devido à presença de idosos, podem ser as que mais sofrem com o aumento dos custos com tratamento médico.

Em suma, este estudo é um primeiro passo na análise dos serviços de saúde e do seu impacto nas decisões de alocação da renda domiciliar. Inúmeras outras questões podem ser tratadas em estudos futuros, como, por exemplo, o uso de uma estrutura de demanda por gasto para estimar a elasticidade-substituição entre os serviços de saúde e outros itens de consumo alternativos. Adicionalmente, pode ser estimado o impacto dos custos com assistência médica sobre a renda e discutidas suas conseqüências sobre a distribuição de renda.

## Abstract

*This paper addresses the question: what happens to health expenditures when a developing country population gets older? One implication of the analysis is that the older the population the higher the demand for the public health system. The analysis is based on a household consumption survey for the city of São Paulo, the largest Brazilian city. The estimation procedures consisted of parametric methods, such as Ordinary Least Square and Tobit Censored Regression, and the Least Absolute Deviation, a semiparametric method. Contrary to Ordinary Least Square and Tobit Censored Regression the Least Absolute Deviation does not require the hypothesis of a positive impact of age on health expenditure is sustained by the data with an estimator presenting large sample properties.*

## Bibliografia

- ALVES, D. Modelling health patterns: a censored regression analysis. *Brazilian Econometric Society Annals*. Belém, Pará, Brasil, 1999.
- AMEMIYA, T. *Advanced econometrics*. Cambridge, Mass.: Harvard University Press, 1985.
- BUCHINSKY, M. Changes in the US wage structure 1963-1987: application of quantile regression. *Econometrica*, v. 2, p. 405-458, Mar. 1994.

- CUTLER, D. M., MEARA, E. The medical costs of the young and old: a forty-year perspective. In: WISE, D. A. (ed.). *Frontiers in the economics of aging*. Chicago: Chicago University Press, 1998.
- DEATON, A. *The analysis of household surveys: a microeconomic approach to development policy*. Johns Hopkins, Md, 1997.
- EATON, J., TAMURA, A. Bilateralism and regionalism in Japanese and US trade and direct foreign investment. *Journal of Japanese and International Economics*, v. 8, p. 478-510, 1994.
- FAIR, R. A theory of extramarital affairs. *Journal of Political Economy*, v. 86, p. 45-61, 1978.
- FROSTIN, P., HOLTSMANN, G. The determinants of residential property damage caused by hurricane Andrew. *Southern Economic Journal*, v. 61, p. 387-397, 1994.
- GHESHER, A., IRISH, M. Residual analysis in the grouped and censored normal linear model. *Journal of Econometrics*, v. 25, p. 303-325, 1987.
- JOHNSTON, J., DI NARDO, J. *Econometric Methods*. New York: McGraw Hill, 1997.
- LECLERE, M. J. The decomposition of coefficients in censored regression models: understanding the effect of independent variables in taxpayer behavior. *National Tax Journal*, v. 47, p. 837-845, 1987.
- LONG, J. S. *Regression models for categorical and limited dependent variables: advanced quantitative technics in the social sciences*, 7. London, England: Sage Publications, 1987.
- MCDONALD, J., MOFFIT, R. The uses of Tobit Analysis. *Review of Economics and Statistics*, v. 62, p. 318-321, 1980.
- POWELL, J. Least absolute deviations estimation for censored regression model. *Journal of Econometrics*, v. 32, p. 143-155, 1984.
- QUESTER, E., GREENE, W. Divorce risk and wife's labor supply behavior. *Social Science Quarterly*, v. 63, p. 17-27, 1982.
- ROGERS, W. H. Quantile regression standard errors. *Stata Technical Bulletin*, College Station, TX, v. 9, p. 16-19, 1992.
- SCHOENBERG, R. *CML users guide*. Maple Valey, WA, USA: Aptech Systems, Inc., 1995.
- STEPHAN, P. E., LEVIN, S. G. *Striking the mother load: the importance of age, place and time*. New York: Oxford University Press, 1992.
- TOBIN, J. Estimation of relationships for limited dependent variables. *Econometrica*, v. 26, p. 24-36, 1958.

WOOLDRIDGE, J. M. *Introductory econometrics: a modern approach*. Cincinnati, OH, South-West, 2000.

———. *Econometric analysis of cross section and panel data*. Cambridge, Mass.: MIT Press, 2001.

WORLD BANK. *Averting the old age crisis: policies to protect the old and promote growth*. Oxford: Oxford University Press, 1994.

———. The Brazilian health system. *Impact Evaluation Report*. Operations Evaluation Department, 1998.

———. *Entering the 21st century: World Bank development report 1999/2000*. Oxford: Oxford University Press, 2000.

*(Originais recebidos em novembro de 2000. Revistos em dezembro de 2001.)*

