

MODELOS DE REGRESSÃO COM VARIÁVEL DEPENDENTE TRUNCADA OU CENSURADA

Introdução

Sucedem por vezes que, numa amostra, a variável de interesse não pode ser observada em todo o seu domínio ou não é observada para alguns dos indivíduos que a integram. Por exemplo, dados quanto ao número de bilhetes vendidos para cada um dos jogos disputados num recinto desportivo ao longo de uma temporada não permitirão quantificar exactamente a procura de que são alvo os espectáculos sempre que a lotação for esgotada; nos jogos em que tal aconteça, apenas se poderá afirmar que a procura de bilhetes foi igual ou superior à capacidade do recinto. Aparelhos de medição do teor de glicose no sangue tipicamente permitem ler valores compreendidos entre 20 e 600 miligramas por decilitro; resultados inferiores a 20 mg/dl são codificados como *lo* e resultados superiores a 600 mg/dl são registados como *hi*. Em sistemas de tributação do rendimento que dispensam da obrigatoriedade de apresentação da declaração anual os agregados familiares com rendimento inferior a certo montante, uma amostra colhida com base nas declarações apresentadas não poderá incluir famílias de baixo nível de rendimento. Em estudos da duração da sobrevivência de indivíduos a quem foi diagnosticada determinada doença, é frequente que apenas se possa afirmar, relativamente aos que permanecem vivos à data da realização do estudo, que o seu período de sobrevivência é igual ou maior do que a duração do tempo transcorrido desde o diagnóstico. De modo análogo, é comum, em análises dos determinantes da duração do desemprego, obter de alguns inquiridos (os que continuam desempregados no momento do inquérito) informação quanto à extensão mínima do seu período forçado de inactividade.

Em todos os casos que se acaba de descrever, está-se perante restrições de amostragem que têm por consequência impedir a observação da variável de interesse numa parte do seu domínio ou relativamente a alguns indivíduos da população. Por uma via ou por outra, é quebrada a identidade entre a variável de interesse, por um lado, e a sua contrapartida observada, por outro.

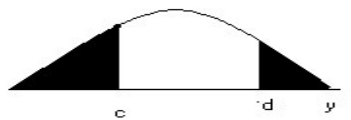
Para o que segue, é conveniente distinguir entre censura e truncamento. Diz-se que uma variável é censurada se não for possível observá-la para uma parte dos indivíduos de uma população; por extensão, diz-se censurada uma amostra em que a variável relevante é

conhecida para alguns indivíduos e desconhecida para os restantes. Uma distribuição truncada pode obter-se de qualquer distribuição de probabilidade pela dupla operação de restringir o domínio original da variável e redimensionar adequadamente a probabilidade sobre o novo domínio; diz-se truncada a variável aleatória assim definida. Uma amostra censurada é extraída da população considerada originalmente; ao invés, uma amostra truncada provém de uma população diferente da original, embora contida nela. Em termos da análise da regressão, uma amostra censurada é caracterizada pela observação das variáveis independentes em todos os elementos da amostra, mas pela observação da variável dependente numa fracção deles, apenas. Numa amostra truncada, todas as observações são completas (são conhecidos os valores das variáveis dependente e independentes para todos os indivíduos seleccionados), mas a amostra respeita somente a um segmento do universo de interesse.

Conquanto a censura e o truncamento possam revestir outras formas, as que se revestem de maior importância são as que resultam da supressão ou omissão de valores inferiores a uma certa quantidade, dita truncamento à esquerda, ou dos valores superiores, caso que corresponde ao truncamento à direita. Por vezes, fixados dois limites, a e b , $a < b$, e restringindo ao intervalo $[a, b]$ o domínio de uma variável, poderá ter interesse um duplo truncamento, à esquerda de a e à direita de b .

Distribuições de probabilidade truncadas

Considere-se a função de densidade de probabilidade de uma variável aleatória contínua, Y , representada na figura seguinte:



Designando por $F(y)$ a função de distribuição de Y , a função de distribuição truncada bilateralmente, à esquerda de c e à direita de d , deduz-se de

$$\text{Prob}(Y \leq y \mid c < Y < d) = \frac{\text{Prob}(c < Y \leq y)}{\text{Prob}(c < Y < d)} = \frac{F(y) - F(c)}{F(d) - F(c)},$$

para $c < y < d$. Por derivação, a função de densidade de probabilidade (f.d.p.) truncada é

$$\frac{d}{dy} \left(\frac{F(y) - F(c)}{F(d) - F(c)} \right) = \frac{f(y)}{F(d) - F(c)},$$

em que $f(y)$ denota a f.d.p. de Y e é $c < y < d$. Os momentos principais da distribuição truncada são

$$E(Y|c < Y < d) = \int_c^d y \frac{f(y)}{F(d) - F(c)} dy$$

e

$$\text{Var}(Y|c < Y < d) = \int_c^d [y - E(Y|c < Y < d)]^2 \frac{f(y)}{F(d) - F(c)} dy.$$

Nos casos particulares de truncamento unilateral, tem-se, para o truncamento à esquerda de c — caso em que $d = +\infty$ e $F(d) = 1$ —,

$$F(y|Y > c) = \frac{F(y) - F(c)}{1 - F(c)}$$

e

$$f(y|Y > c) = \frac{f(y)}{1 - F(c)},$$

e, para o truncamento à direita de d — caso em que $c = -\infty$ e $F(c) = 0$ —,

$$F(y|Y < d) = \frac{F(y)}{F(d)}$$

e

$$f(y|Y < d) = \frac{f(y)}{F(d)}.$$

Exemplo 1

Seja Y uma variável aleatória contínua com distribuição uniforme no intervalo $[0, 1]$. Então, é $f(y) = 1$, se $0 < y < 1$. Com $c = 1/3$, vem

$$f(y|Y > 1/3) = \frac{f(y)}{\text{Prob}(Y > 1/3)} = \frac{1}{2/3} = \frac{3}{2}, \quad 1/3 < y < 1,$$

com média

$$E(Y|Y > 1/3) = \int_{1/3}^1 \frac{3}{2} y dy = \frac{2}{3}$$

e variância

$$\text{Var}(Y|Y > 1/3) = \int_{1/3}^1 \frac{3}{2} \left(y - \frac{2}{3} \right)^2 dy = \frac{1}{27}.$$

Sem truncamento, é $\text{Var}(Y) = 1/12$.

Prova-se que, se Y for uma variável com distribuição normal de média μ e variância σ^2 (doravante designada pela notação $Y \sim N(\mu, \sigma^2)$), é

$$E(Y|Y>c) = \mu + \sigma \frac{\phi\left(\frac{c-\mu}{\sigma}\right)}{1-\Phi\left(\frac{c-\mu}{\sigma}\right)}$$

(em que $\phi(\cdot)$ se refere à f.d.p. normal reduzida e $\Phi(\cdot)$ à função de distribuição correspondente) e

$$\text{Var}(Y|Y>c) = \sigma^2 \left\{ 1 - \frac{\phi\left(\frac{c-\mu}{\sigma}\right)}{1-\Phi\left(\frac{c-\mu}{\sigma}\right)} \left[\frac{\phi\left(\frac{c-\mu}{\sigma}\right)}{1-\Phi\left(\frac{c-\mu}{\sigma}\right)} - \frac{c-\mu}{\sigma} \right] \right\}.$$

Usando a notação

$$\alpha = \frac{c-\mu}{\sigma},$$

$$\lambda(\alpha) = \frac{f\left(\frac{c-\mu}{\sigma}\right)}{1-\Phi\left(\frac{c-\mu}{\sigma}\right)} = \frac{\phi(\alpha)}{1-\Phi(\alpha)},$$

pode dar-se um aspecto mais simples às expressões acima:

$$E(Y|Y>c) = \mu + \sigma \lambda(\alpha)$$

e

$$\text{Var}(Y|Y>c) = \sigma^2 \{1 - \lambda(\alpha)[\lambda(\alpha) - \alpha]\}.$$

A função $\lambda(\alpha)$ é o chamado *inverse Mills ratio* e corresponde à função *hazard* da distribuição normal. Trata-se de uma função convexa, que assume apenas valores positivos, que tende para 0 quando α tende para $-\infty$ e para α quando α tende para ∞ e cuja derivada,

$$\frac{d\lambda(\alpha)}{d\alpha} = \lambda(\alpha) [\lambda(\alpha) - \alpha],$$

apenas assume valores no intervalo $]0, 1[$.

Num caso de truncamento à direita, sendo $Y \sim N(\mu, \sigma^2)$, prova-se que é

$$E(Y|Y < c) = \mu - \sigma \frac{\phi\left(\frac{c - \mu}{\sigma}\right)}{\Phi\left(\frac{c - \mu}{\sigma}\right)}$$

e

$$\text{Var}(Y|Y < c) = \sigma^2 \left\{ 1 - \frac{\phi\left(\frac{c - \mu}{\sigma}\right)}{\Phi\left(\frac{c - \mu}{\sigma}\right)} \left[\frac{\phi\left(\frac{c - \mu}{\sigma}\right)}{\Phi\left(\frac{c - \mu}{\sigma}\right)} - \frac{c - \mu}{\sigma} \right] \right\}.$$

Como é patente nas expressões acima, a média de uma distribuição truncada à esquerda é maior do que a média da distribuição original e a média de uma distribuição truncada à direita é menor do que a média da distribuição original; em ambos os casos, a variância é comprimida. A tabela seguinte ilustra esse duplo efeito do truncamento para a distribuição normal reduzida, truncada à esquerda:

Média e variância da distribuição normal reduzida, truncada à esquerda

Ponto de truncamento	Média da dist. truncada	Variância da dist. truncada
c	$E(Y Y > c)$	$\text{Var}(Y Y > c)$
-2	0,0552	0,8865
-1	0,2876	0,6297
-0,5	0,5092	0,4862
0	0,7979	0,3634
0,5	1,1411	0,2685
1	1,5251	0,1991
2	2,3732	0,1143

Exemplo 2

Num recinto desportivo com capacidade para 10000 espectadores, realizaram-se ao longo de uma época vários jogos. Foram vendidos, em média, 8000 bilhetes por jogo e atingiu-se lotação completa em 70% dos casos.

Supondo que a variável Y^* , "procura de bilhetes para cada jogo", é uma variável aleatória normal com média, μ , e variância, σ^2 , desconhecidas, é

$$\text{Prob}(Y^* > 10000) = \text{Prob}\left(\frac{Y^* - \mu}{\sigma} > \frac{10000 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{10000 - \mu}{\sigma}\right).$$

Tomando 0,7 como estimativa dessa probabilidade e notando, nas tabelas da distribuição normal reduzida, que é $\Phi(-0,5244) = 0,3$, obtém-se a relação

$$\frac{10000 - \hat{\mu}}{\hat{\sigma}} = -0,5244.$$

A variável observada, Y , "número de bilhetes vendidos para cada jogo", é igual a Y^* se a procura for inferior a 10000 lugares; senão, é $Y = 10000$. O valor médio de Y é calculado por

$$E(Y) = 10000 \text{ Prob}(Y^* > 10000) + E(Y^* | Y^* \leq 10000) \text{ Prob}(Y^* \leq 10000).$$

As duas probabilidades citadas são estimadas em 0,7, a primeira, e, por consequência, 0,3, a segunda. Para apurar $E(Y^* | Y^* \leq 10000)$, um resultado evocado acima para o truncamento à direita indica

$$E(Y^* | Y^* \leq 10000) = \mu - \sigma \frac{\phi(-0,5244)}{\Phi(-0,5244)} = \mu - 1,1590 \sigma.$$

Usando 8000 como estimativa de $E(Y)$, vem

$$8000 = 10000 \times 0,7 + (\hat{\mu} - 1,1590 \hat{\sigma}) \times 0,3$$

ou

$$\hat{\mu} - 1,1590 \hat{\sigma} = 1000/0,3.$$

O sistema de duas equações em $\hat{\mu}$ e $\hat{\sigma}$ tem por solução aproximada $\hat{\mu} = 15509$ e $\hat{\sigma} = 10505$. Neste caso, a média de 8000 bilhetes vendidos por jogo parece subestimar grosseiramente a procura.

Exemplo 3

No exemplo anterior, suponha-se que 8000 é a média de bilhetes vendidos por jogo, mas apenas nos jogos em que não foi esgotada a lotação do recinto. Agora, 8000 é uma estimativa da média da distribuição truncada,

$$8000 = \hat{\mu} - 1,1590 \hat{\sigma},$$

e não da média da distribuição censurada. Conjugada com a condição

$$\frac{10000 - \hat{\mu}}{\hat{\sigma}} = -0,5244,$$

a nova solução do sistema é $\hat{\mu} = 11653$ e $\hat{\sigma} = 3152$.

O modelo *tobit*

Numa amostra com observações sobre a despesa das famílias em certos bens de consumo duradouro, é frequente encontrar um grupo significativo de observações que declaram despesa igual a 0, a par de outras que reportam valores positivos com grande dispersão. Num estudo de 1958, James Tobin, Prémio Nobel da Economia em 19??, propôs um modelo para acomodar situações como essa, em que há acumulação de observações num único ponto de massa na fronteira do domínio. Pelas semelhanças que apresentava com as formulações *probit* e *logit*, o modelo acabou por ficar conhecido na gíria por *tobit*.

Admita-se uma variável latente, Y_i^* , tal que

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

ou, mais concisamente,

$$Y_i^* = \mathbf{X}_i \boldsymbol{\beta} + u_i,$$

em que \mathbf{X}_i designa o o vector (1×k)

$$\mathbf{X}_i = [1 \quad X_{2i} \quad X_{3i} \quad \dots \quad X_{ki}]$$

e $\boldsymbol{\beta}$ o habitual vector (k×1) de coeficientes de regressão. A variável Y_i^* , contudo, não é observada em todo o seu domínio; observada é Y_i tal que

$$Y_i = \begin{cases} Y_i^*, & \text{se } Y_i^* > 0 \\ 0, & \text{se } Y_i^* \leq 0 \end{cases}.$$

Suponha-se, por fim, que $\{u_1, u_2, \dots, u_i, \dots\}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.),

$$u_i \sim N(0, \sigma^2).$$

A especificação do modelo *tobit* é, portanto,

$$Y_i^* = \mathbf{X}_i \boldsymbol{\beta} + u_i,$$

$$Y_i = \begin{cases} Y_i^*, & \text{se } Y_i^* > 0 \\ 0, & \text{se } Y_i^* \leq 0 \end{cases}.$$

$$u_i \sim \text{i.i.d.}, N(0, \sigma^2).$$

Várias questões suscitadas pela especificação do modelo e, entre elas, a da escolha da constante 0 como limiar de observação da variável dependente, serão discutidas mais adiante.

Numa amostra com censura da variável dependente, coexistem observações ditas completas, em que, para cada indivíduo, se dispõe de dados quer quanto às variáveis independentes quer quanto à variável dependente, e observações incompletas, em que apenas se observam as primeiras. A probabilidade de uma observação incompleta num modelo tobit é dada por

$$\text{Prob}(Y_i^* \leq 0) = \text{Prob}(\mathbf{X}_i \boldsymbol{\beta} + u_i \leq 0) = \text{Prob}\left(\frac{u_i}{\sigma} \leq -\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) = \Phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)$$

(continuando a usar a notação $\Phi(\cdot)$ para a função de distribuição normal reduzida) ou, atendendo à simetria relativamente à origem,

$$\text{Prob}(Y_i^* \leq 0) = 1 - \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right).$$

De acordo com a especificação do modelo, é ainda

$$E(Y_i^* | \mathbf{X}_i) = \mathbf{X}_i \boldsymbol{\beta};$$

mas, para as observações completas, é

$$E(Y_i^* | \mathbf{X}_i, Y_i^* > 0) = \mathbf{X}_i \boldsymbol{\beta} + \sigma \frac{\phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)},$$

recorrendo à média da distribuição truncada que se apresentou atrás. Supondo $Y_i = 0$ para as observações incompletas, tem-se para a média de Y , que resulta da ponderação de um grupo de observações com $Y_i = 0$ e probabilidade $1 - \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)$ e de outro grupo com

$Y_i > 0$ e probabilidade $\Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)$,

$$E(Y_i | \mathbf{X}_i) = 0 \times \left[1 - \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)\right] + [E(Y_i | \mathbf{X}_i, Y_i^* > 0)] \times \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right),$$

ou seja

$$E(Y_i | \mathbf{X}_i) = \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) \left[\mathbf{X}_i \boldsymbol{\beta} + \sigma \frac{\phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)} \right].$$

Os efeitos marginais de variações *coeteris paribus* nas variáveis explicativas num modelo *tobit* são também complexos. Claro que, se Y^* for uma função linear de X_j , continua a ser

$$\frac{\partial E(Y_i^* | \mathbf{X}_i)}{\partial X_{ji}} = \beta_j,$$

uma constante. Ao invés, o efeito marginal sobre a variável observada é variável conforme o valor de X_j e conforme o valor das demais variáveis explicativas, mesmo para dois indivíduos com igual valor observado para X_j . De facto, prova-se (ver, por exemplo, Greene (2003), p. 766) que é

$$\frac{\partial E(Y_i | \mathbf{X}_i)}{\partial X_{ji}} = \beta_j \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right).$$

Variações em X_{ji} têm efeito não só sobre a média da variável quando observada, mas também sobre a probabilidade de a variável ser observada. Uma vez que a função de distribuição $\Phi(\cdot)$, para valores finitos do argumento, apenas assume valores no intervalo $]0; 1[$, pode afirmar-se que o efeito marginal sobre a variável observada tem o mesmo sinal, mas é menor, em valor absoluto, do que o efeito sobre a variável latente. O exemplo seguinte ilustra alguns dos problemas que se colocam na interpretação dos efeitos marginais das variáveis.

Exemplo 4

Suponha-se que num recinto desportivo com capacidade para 10000 espectadores se realizaram num certo período 10 jogos, relativamente aos quais foi possível quantificar o número de bilhetes procurados, ainda que a procura não tenha sido integralmente satisfeita. Suponha-se, adicionalmente, que se sabe que determinado aumento do preço dos bilhetes teria por efeito diminuir uniformemente a procura de bilhetes em 500 lugares. A tabela seguinte descreve a situação antes (2.^a e 3.^a colunas) e depois (4.^a e 5.^a colunas) do aumento do preço:

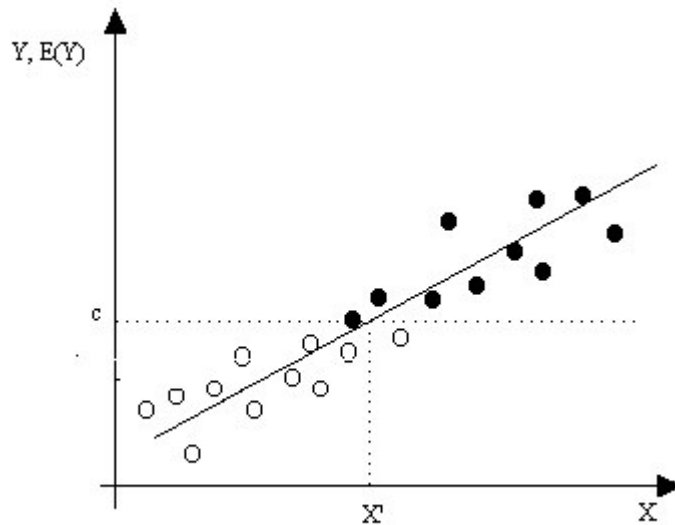
Jogo	Antes		Depois	
	N.º de bilhetes desejados	N.º de bilhetes vendidos	N.º de bilhetes desejados	N.º de bilhetes vendidos
1	25700	10000	25200	10000
2	23000	10000	22500	10000
3	15800	10000	15300	10000
4	15000	10000	14500	10000
5	15000	10000	14500	10000
6	10300	10000	9800	9800
7	10200	10000	9700	9700
8	5000	5000	4500	4500
9	3000	3000	2500	2500
10	2000	2000	1500	1500
Totais	125000	80000	120000	78000

Note-se que, como seria de esperar, o valor médio da variável latente, número de lugares pretendidos, diminuiu por 500. O efeito do aumento do preço dos bilhetes sobre a variável observada, número de ingressos vendidos, foi mais modesto: menos 200, em média. Note-se ainda que, a julgar apenas pela variável truncada "número de bilhetes vendidos para os jogos em que a lotação não foi esgotada", o efeito marginal do aumento do preço dos bilhetes poderia mesmo parecer positivo: enquanto, antes do aumento, se vendiam, em média, cerca de 3333 lugares para esses jogos, depois do aumento a média passou a ser de 5600...

Estimação do modelo *tobit*

Fosse a variável explicada observável em todo o seu domínio, não haveria dúvida em utilizar directamente a equação $Y_i^* = \mathbf{X}_i \boldsymbol{\beta} + u_i$, como base para a estimação do vector $\boldsymbol{\beta}$. Perante uma amostra censurada, porém, há razão para alguma hesitação quanto à estratégia de estimação a adoptar.

A estratégia mais simples seria a de ignorar as observações para as quais não se dispõe de um valor de Y^* e proceder à estimação por OLS da relação, usando somente as observações completas. O problema é que os estimadores OLS, nessas circunstâncias, não são consistentes, por razões que a figura da página seguinte sugere.



Na figura, círculos a negro correspondem a observações completas, enquanto os círculos em branco se referem a indivíduos para quem Y não atinge o limiar de observação c e, por conseguinte, não é observada, a despeito de X o ser. A recta inserida no diagrama representa a função de regressão na população e tem por equação $E(Y) = \beta_1 + \beta_2 X$.

Note-se que, para observações completas com $X \leq X'$, terá de ser necessariamente positiva a perturbação aleatória associada. Conquanto seja $E(u|X') = 0$ na população, em qualquer amostra truncada à esquerda de c será $E(u|X', Y \geq c) > 0$. Tal é incompatível com a hipótese clássica de que $E(u|X') = 0$, se forem sistematicamente ignoradas as observações incompletas.

Tomando outra perspectiva sobre a questão, atente-se também em que, numa amostra truncada, é

$$E(Y_i | \mathbf{X}_i, Y_i^* > 0) = \mathbf{X}_i \boldsymbol{\beta} + \sigma \lambda_i,$$

em que se designou por λ_i a variável

$$\lambda_i = \frac{\phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)}.$$

A estimação de $\boldsymbol{\beta}$ conduzida com base numa amostra truncada e na equação de regressão

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + u_i$$

corresponde à omissão de um regressor, λ_i , com a penalidade bem conhecida de inconsistência dos estimadores OLS, a menos que a variável explicativa omitida e as incluídas estivessem não correlacionadas. Por outro lado, a variável λ_i não é observada e depende ela própria do vector desconhecido β .

Há, contudo, a possibilidade de levar a cabo a estimação pelo método NLS do modelo não linear

$$Y_i = f(\mathbf{X}_i, \beta, \sigma) + v_i,$$

em que v_i designa uma perturbação aleatória de média nula e $f(\mathbf{X}_i, \beta, \sigma)$ é a função

$$f(\mathbf{X}_i, \beta, \sigma) = \mathbf{X}_i \beta + \sigma \frac{\phi\left(-\frac{\mathbf{X}_i \beta}{\sigma}\right)}{1 - \Phi\left(-\frac{\mathbf{X}_i \beta}{\sigma}\right)}.$$

Apesar da forte não linearidade da relação (que, recorde-se, envolve um integral no denominador do rácio que define λ_i), os modernos meios de cálculo automático viabilizam, em regra, o emprego da técnica. As razões da crítica de inconsistência que acima se invocaram contra o estimador OLS não prevalecem contra o estimador NLS. Subsistirá, contudo, o problema de ineficiência. É que as observações incompletas, excluídas numa amostra truncada, contêm potencialmente informação valiosa sobre os parâmetros de interesse; um estimador que ignore essa informação está condenado à perda de eficiência.

Uma segunda estratégia de estimação possível consistiria no uso da análise *probit* ou *logit* para estimação dos parâmetros em β , usando apenas a informação $Y = 1$, se se dispõe do valor observado para a variável dependente, ou $Y = 0$ em caso contrário. Uma penalidade em que se incorreria com essa estratégia é, novamente, a da ineficiência dos estimadores, originada pelo não emprego de toda a informação relevante. Conquanto satisfatória para estimar a probabilidade (condicional a \mathbf{X}_i) de ser observada a variável de interesse, essa estratégia não tira partido da informação quantitativa disponível sobre Y_i^* para estimar com maior precisão o impacto de cada variável explicativa sobre o valor médio dessa variável. Por outro lado, a abordagem não permitiria a estimação de σ .

Assente que a melhor estratégia de estimação passa pelo emprego da amostra censurada, pese embora a falta da informação desejável quanto à variável dependente para uma parte dos seus elementos, resta a questão do método de estimação a empregar. Por razões análogas às que atrás se apontaram, estimadores OLS, que fariam uso de $Y_i = 0$ para uma parte da amostra, não seriam consistentes. A função de regressão correcta a especificar seria

$$Y_i = f(\mathbf{X}_i, \beta, \sigma) + v_i,$$

agora com

$$f(\mathbf{X}_i, \boldsymbol{\beta}, \sigma) = \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) \left[\mathbf{X}_i \boldsymbol{\beta} + \sigma \frac{\phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)} \right],$$

ou, de forma equivalente,

$$f(\mathbf{X}_i, \boldsymbol{\beta}, \sigma) = \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) \mathbf{X}_i \boldsymbol{\beta} + \sigma \phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right).$$

Mais uma vez, tratando-se de uma função não linear em $\boldsymbol{\beta}$ e σ , haveria que recorrer ao estimador NLS.

O método de eleição para estimação de um modelo *tobit* é o método da máxima verosimilhança. Relativamente às observações censuradas, em que Y é não observada ou é $Y = 0$, tudo o que se sabe é que essas observações ocorrem com probabilidade

$$\text{Prob}(Y_i^* \leq 0 | \mathbf{X}_i) = 1 - \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right);$$

esta função mede o contributo de cada uma das observações incompletas para a função de verosimilhança.

Para as observações completas, já é possível usar a densidade de probabilidade:

$$\text{Prob}(Y_i^* > 0 | \mathbf{X}_i) f(y_i^* | Y_i^* > 0, \mathbf{X}_i) = \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) \frac{f(y_i^*)}{\Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)} = f(y_i^*),$$

designando por $f(y_i^*)$ a f.d.p. da variável Y_i^* , que, conforme resulta das hipóteses do modelo *tobit*, tem distribuição normal de média $\mathbf{X}_i \boldsymbol{\beta}$ e variância σ^2 ,

$$f(y_i^*) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i^* - \mathbf{X}_i \boldsymbol{\beta})^2} = \frac{1}{\sigma} \phi\left(\frac{y_i^* - \mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right).$$

A função de verosimilhança para estimação dos parâmetros do modelo *tobit* é

$$L(\boldsymbol{\beta}, \sigma; \mathbf{Y}, \mathbf{X}) = \prod_0 \left[1 - \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) \right] \prod_1 \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i^* - \mathbf{X}_i \boldsymbol{\beta})^2} \right],$$

expressão onde o primeiro produto se deve entender iterado sobre i tal que $Y_i = 0$ e o segundo produto iterado sobre i tal que $Y_i = 1$. Usando notação análoga e supondo haver na amostra n_0 observações com $Y_i = 0$ e n_1 observações completas, $n = n_0 + n_1$, a função logarítmica de verosimilhança pode escrever-se

$$\ln L(\boldsymbol{\beta}, \sigma; \mathbf{Y}, \mathbf{X}) = \sum_0 \ln \left[1 - \Phi \left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma} \right) \right] - \frac{n_1}{2} \ln(2\pi) - \frac{n_1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_1 (y_i^* - \mathbf{X}_i \boldsymbol{\beta})^2 .$$

A condição de 1.ª ordem para maximização da função logarítmica de verosimilhança requer que seja um vector nulo o vector ($k \times 1$) de derivadas parciais

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}}$$

e que seja também nula a derivada parcial $\frac{\partial \ln L}{\partial \sigma}$. No caso, vem

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma} \sum_0 \lambda_i \mathbf{X}_i' + \frac{1}{\sigma^2} \sum_1 (y_i^* - \mathbf{X}_i \boldsymbol{\beta}) \mathbf{X}_i'$$

e

$$\frac{\partial \ln L}{\partial \sigma} = \frac{1}{\sigma^2} \sum_0 \lambda_i \mathbf{X}_i \boldsymbol{\beta} - \frac{n_1}{\sigma} + \frac{1}{\sigma^3} \sum_1 (y_i^* - \mathbf{X}_i \boldsymbol{\beta})^2 ,$$

em que se designou por λ_i o *inverse Mills ratio* definido acima. Notando que a condição de anulamento do vector de derivadas parciais em ordem a $\boldsymbol{\beta}$ requer

$$\frac{1}{\sigma} \sum_1 (y_i^* - \mathbf{X}_i \boldsymbol{\beta}) \mathbf{X}_i' = \sum_0 \lambda_i \mathbf{X}_i'$$

e que, por outro lado, é

$$\sum_0 \lambda_i \mathbf{X}_i \boldsymbol{\beta} = \boldsymbol{\beta}' \sum_0 \lambda_i \mathbf{X}_i' ,$$

é possível escrever a condição de anulamento da derivada parcial em ordem a σ na forma

(----- TEXTO INCOMPLETO ! -----)

Um exemplo

Para exemplificar os diversos métodos de estimação discutidos acima, criou-se uma amostra artificial de 1000 observações das variáveis Y^* , X_2 e X_3 que obedecem à estrutura

$$Y^* = 10 - 0,5 X_2 + 0,2 X_3 + u,$$

em que u é uma variável aleatória normal, com média 0 e desvio padrão 10, $u \sim N(0,10^2)$.

Usando o programa informático *Econometric Views (EViews)*, obteve-se, por OLS, na regressão de Y^* num termo constante, em X_2 e em X_3 , com todas as 1000 observações disponíveis:

Quadro 1 – Resultados da estimação por OLS, se Y^* fosse observada

Dependent Variable: YSTAR				
Method: Least Squares				
Sample: 1 1000				
Included observations: 1000				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.32630	0.662989	15.57537	0.0000
X2	-0.510061	0.010973	-46.48351	0.0000
X3	0.194529	0.009331	20.84676	0.0000
R-squared	0.718220	Mean dependent var	-10.14445	
Adjusted R-squared	0.717655	S.D. dependent var	19.20308	
S.E. of regression	10.20378	Akaike info criterion	7.486389	
Sum squared resid	103804.8	Schwarz criterion	7.501112	
Log likelihood	-3740.195	F-statistic	1270.611	
Durbin-Watson stat	2.015871	Prob(F-statistic)	0.000000	

Obviamente, a regressão não pode ser replicada com dados de uma amostra real, em que a variável dependente não seria observada em todo o domínio. O propósito da sua inclusão aqui é, meramente, o de servir de termo de comparação dos resultados obtidos pelos diversos métodos de estimação com os que seriam "ideais".

A partir de Y^* , criou-se uma variável Y segundo a regra

$$Y_i = \begin{cases} Y_i^*, & \text{se } Y_i^* > 0 \\ 0, & \text{se } Y_i^* \leq 0 \end{cases}.$$

Na amostra, resultou que Y^* é observada em 305 casos; para as 695 observações restantes, fez-se $Y = 0$.

Uma segunda estimação foi conduzida com base na equação

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i,$$

usando Y (em vez de Y*) como variável dependente:

Quadro 2 – Resultados da estimação por OLS com todas as observações disponíveis

Dependent Variable: YOBS				
Method: Least Squares				
Sample: 1 1000				
Included observations: 1000				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.447564	0.380851	22.18079	0.0000
X2	-0.134291	0.006303	-21.30467	0.0000
X3	0.081646	0.005360	15.23149	0.0000
R-squared	0.401065	Mean dependent var		3.787605
Adjusted R-squared	0.399864	S.D. dependent var		7.566313
S.E. of regression	5.861506	Akaike info criterion		6.377686
Sum squared resid	34254.18	Schwarz criterion		6.392409
Log likelihood	-3185.843	F-statistic		333.8110
Durbin-Watson stat	2.058769	Prob(F-statistic)		0.000000

Saliente-se que o método empregue, OLS, não é consistente; as estimativas produzidas revelam, sem surpresa, uma deterioração significativa por referência ao padrão do Quadro 1.

Constam do Quadro 3 os resultados da estimação por OLS da mesma equação, mas agora com a amostra restrita às 305 observações em que o valor da variável dependente de interesse é, de facto, observado:

Quadro 3 – Resultados da estimação por OLS com as observações completas, apenas

Dependent Variable: YOBS				
Method: Least Squares				
Sample(adjusted): 4 1000 IF YPOS=1				
Included observations: 305 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.75702	0.730945	18.82086	0.0000
X2	-0.220681	0.028210	-7.822663	0.0000
X3	0.091027	0.010615	8.575172	0.0000
R-squared	0.236582	Mean dependent var		12.41838
Adjusted R-squared	0.231526	S.D. dependent var		8.977654
S.E. of regression	7.870049	Akaike info criterion		6.973793
Sum squared resid	18705.18	Schwarz criterion		7.010386
Log likelihood	-1060.503	F-statistic		46.79462
Durbin-Watson stat	2.042491	Prob(F-statistic)		0.000000

Insista-se, o método não é consistente; contudo, os resultados são, na generalidade, melhores do que os alcançados com a inclusão nos cálculos de um conjunto considerável de zeros que "mascaram" o valor real de Y^* .

O quadro seguinte documenta as estimativas obtidas pelo método da máxima verosimilhança, por meio de uma rotina criada especificamente para determinação do máximo da função

$$\ln L(\boldsymbol{\beta}, \sigma; \mathbf{Y}, \mathbf{X}) = \sum_0 \ln \left[1 - \Phi \left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma} \right) \right] - \frac{n_1}{2} \ln(2\pi) - \frac{n_1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_1 (y_i^* - \mathbf{X}_i \boldsymbol{\beta})^2 .$$

Quadro 4 – Resultados da estimação por ML com o comando *logl*

LogL: TOBIT_ML				
Method: Maximum Likelihood (Marquardt)				
Sample: 1 1000				
Included observations: 1000				
Evaluation order: By observation				
Initial Values: C(1)=13.7570, C(2)=-0.22068, C(3)=0.09103, C(4)=5.00000				
Convergence achieved after 12 iterations				
	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	10.25078	0.849662	12.06454	0.0000
C(2)	-0.524352	0.030636	-17.11564	0.0000
C(3)	0.195038	0.013155	14.82596	0.0000
C(4)	10.69815	0.480644	22.25796	0.0000
Log likelihood	-1320.957	Akaike info criterion		2.649914
Avg. log likelihood	-1.320957	Schwarz criterion		2.669545
Number of Coefs.	4	Hannan-Quinn criter.		2.657375

Os coeficientes designados no *output* por C(1), C(2), C(3) e C(4) correspondem, por essa ordem, a β_1 , β_2 , β_3 e σ . Conforme se lê no quadro, escolheram-se, para ponto de partida das iterações, as estimativas dos coeficientes de regressão do Quadro 3 e, no que toca a σ , um valor arbitrário, 5. O *software* que se empregou permite também a estimação por máxima verosimilhança do modelo *tobit*, sem necessidade de programação pelo utilizador. Invocando simplesmente o comando respectivo, os resultados são os que o Quadro 5 exhibe (ver página seguinte). As estimativas obtidas pelos dois procedimentos são praticamente coincidentes e concorrem na conclusão de que o máximo da função logarítmica de verosimilhança é igual a -1320,957. É também patente que o método (que, ao contrário dos citados anteriormente, goza da propriedade de consistência) produz, no exemplo, estimativas de qualidade muito satisfatória.

Quadro 5 – Resultados da estimação por ML com o comando *censored*

Dependent Variable: YOBS				
Method: ML - Censored Normal (TOBIT)				
Sample: 1 1000				
Included observations: 1000				
Left censoring (value) at zero				
Convergence achieved after 8 iterations				
Covariance matrix computed using second derivatives				
	Coefficient	Std. Error	z-Statistic	Prob.
C	10.25078	0.844483	12.13853	0.0000
X2	-0.524352	0.028145	-18.63026	0.0000
X3	0.195038	0.012707	15.34826	0.0000
Error Distribution				
SCALE:C(4)	10.69816	0.461430	23.18478	0.0000
R-squared	0.505504	Mean dependent var	3.787605	
Adjusted R-squared	0.504015	S.D. dependent var	7.566313	
S.E. of regression	5.328667	Akaike info criterion	2.649914	
Sum squared resid	28281.12	Schwarz criterion	2.669545	
Log likelihood	-1320.957	Hannan-Quinn criter.	2.657375	
Avg. log likelihood	-1.320957			
Left censored obs	695	Right censored obs	0	
Uncensored obs	305	Total obs	1000	

Como se referiu, outra opção quanto à estimação de modelos *tobit* é o método NLS, aplicado a funções de regressão definidas de modo a acomodar os efeitos de censura ou truncamento. No primeiro caso, a equação de regressão é

$$Y_i = \Phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) \mathbf{X}_i \boldsymbol{\beta} + \sigma \phi\left(\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right) + v_i$$

e obtiveram-se, por NLS, os resultados documentados no Quadro 6 (ver página seguinte). O estimador é, em geral, consistente e essa propriedade transparece no confronto com as estimativas obtidas por OLS. No entanto, a perturbação v_i não é homoscedástica; ao ignorar esse facto, o estimador NLS perde eficiência, por comparação com o estimador ML, e a lacuna é visível na qualidade das estimativas, claramente aquém das reproduzidas nos Quadros 4 e 5. A deficiência é particularmente notória no que se refere à estimativa de σ e, no exemplo em apreço, é agravada pela inclusão na amostra de quase 70% de observações completamente inúteis para estimação desse parâmetro.

A estimação por NLS da equação de regressão para a amostra truncada,

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + \sigma \frac{\phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(-\frac{\mathbf{X}_i \boldsymbol{\beta}}{\sigma}\right)} + v_i,$$

conduz aos resultados do Quadro 7, abaixo.

Quadro 6 – Resultados da estimação por NLS com todas as observações disponíveis

Dependent Variable: YOBS				
Method: Least Squares				
Sample: 1 1000				
Included observations: 1000				
Convergence achieved after 10 iterations				
YOBS=(@CNORM((C(1)+C(2)*X2+C(3)*X3)/C(4)))*(C(1)+C(2)*X2+C(3)*X3+C(4))*(@DNORM((-C(1)-C(2)*X2-C(3)*X3)/C(4))/(1-@CNORM((-C(1)-C(2)*X2-C(3)*X3)/C(4))))				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	11.08555	0.459218	24.14008	0.0000
C(2)	-0.348160	0.032902	-10.58165	0.0000
C(3)	0.140714	0.010147	13.86793	0.0000
C(4)	4.841266	1.764487	2.743724	0.0062
R-squared	0.513908	Mean dependent var		3.787605
Adjusted R-squared	0.512444	S.D. dependent var		7.566313
S.E. of regression	5.283195	Akaike info criterion		6.170931
Sum squared resid	27800.50	Schwarz criterion		6.190562
Log likelihood	-3081.465	Durbin-Watson stat		2.055474

Quadro 7 – Resultados da estimação por NLS com as observações completas, apenas

Dependent Variable: YOBS				
Method: Least Squares				
Sample(adjusted): 4 1000 IF YPOS=1				
Included observations: 305 after adjusting endpoints				
Convergence achieved after 13 iterations				
YOBS=C(1)+C(2)*X2+C(3)*X3+C(4)*(@DNORM((C(1)+C(2)*X2+C(3)*X3)/C(4)))/(1-@CNORM((-C(1)-C(2)*X2-C(3)*X3)/C(4))))				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.526555	12.21457	0.124978	0.9006
C(2)	-0.666460	0.284831	-2.339844	0.0199
C(3)	0.215294	0.078332	2.748477	0.0063
C(4)	16.64464	5.867459	2.836772	0.0049
R-squared	0.252688	Mean dependent var		12.41838
Adjusted R-squared	0.245239	S.D. dependent var		8.977654
S.E. of regression	7.799513	Akaike info criterion		6.959028
Sum squared resid	18310.56	Schwarz criterion		7.007819
Log likelihood	-1057.252	Durbin-Watson stat		1.953952

Por último, ensaiou-se o método proposto por Heckman. Como se referiu, nesse método o primeiro passo é a estimação de um modelo *probit* com que se obtêm estimativas $\left(\hat{\beta}/\sigma\right)$, a partir das quais se constrói, por sua vez, uma estimativa da variável a que se atribuiu o símbolo λ_i :

$$\hat{\lambda}_i = \frac{\phi\left[-\mathbf{X}_i\left(\hat{\beta}/\sigma\right)\right]}{1 - \Phi\left[-\mathbf{X}_i\left(\hat{\beta}/\sigma\right)\right]} = \frac{\phi\left[\mathbf{X}_i\left(\hat{\beta}/\sigma\right)\right]}{\Phi\left[\mathbf{X}_i\left(\hat{\beta}/\sigma\right)\right]}.$$

É essa variável que, no *output* (ver Quadro 8), aparece designada por LAMBDA_EST. O segundo passo é a simples estimação por OLS de um modelo em que, com a amostra truncada, se adicionou $\hat{\lambda}_i$ ao conjunto de regressores:

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + \sigma \hat{\lambda}_i + v_i.$$

Quadro 8 – Resultados da estimação pelo método de Heckman

Dependent Variable: YOBS				
Method: Least Squares				
Sample(adjusted): 4 1000 IF YPOS=1				
Included observations: 305 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.23191	1.015313	12.04742	0.0000
X2	-0.336421	0.060686	-5.543600	0.0000
X3	0.129103	0.020611	6.263739	0.0000
LAMBDA_EST	4.757209	2.212076	2.150563	0.0323
R-squared	0.248134	Mean dependent var	12.41838	
Adjusted R-squared	0.240641	S.D. dependent var	8.977654	
S.E. of regression	7.823238	Akaike info criterion	6.965102	
Sum squared resid	18422.12	Schwarz criterion	7.013893	
Log likelihood	-1058.178	F-statistic	33.11251	
Durbin-Watson stat	1.942008	Prob(F-statistic)	0.000000	

Confrontados os resultados no Quadro 8 com os do Quadro 3, em que o modelo fora estimado com exclusão da variável independente que "corrige" os efeitos do truncamento, é patente que a qualidade das estimativas dos coeficientes de regressão é superior. Contudo, a estimativa do desvio-padrão σ (que, no Quadro 8, é a estimativa do coeficiente de LAMBDA_EST) subestima grosseiramente o parâmetro.

Finalmente, reuniram-se no Quadro 9 (página seguinte) os principais resultados obtidos com os diversos métodos de estimação referidos no exemplo em causa.

Quadro 9 – Sumário de resultados da estimação

Parâmetro	Valor	Estimativas						
		OLS, com Y^*	OLS, com amostra censurada	OLS, com amostra truncada	ML	NLS, com amostra censurada	NLS, com amostra truncada	Heckman
β_1	10	10.326	8.448	13.757	10.251	11.086	1.527	12.232
β_2	-0,5	-0.510	-0.134	-0.221	-0.524	-0.348	-0.666	-0.336
β_3	0,2	0.195	0.082	0.091	0.195	0.141	0.215	0.129
σ	10	10.204	5.862	7.870	10.698	4.841	16.645	4.757
Dimensão da amostra		1000	1000	305	1000	1000	305	305

O aspecto mais saliente do quadro é, sem dúvida, o de que os melhores resultados foram conseguidos com o método da máxima verosimilhança (ML), que reproduz, de forma muito satisfatória, os resultados a que se teria chegado se existisse a possibilidade de levar a cabo a estimação sem as distorções da censura ou do truncamento. Utilize-se a amostra censurada ou a versão truncada, o estimador ordinário de mínimos quadrados (OLS) não é consistente e as consequências transparecem na qualidade das estimativas contidas na quarta e quinta colunas da tabela. Neste exemplo, os estimadores não lineares (NLS) e o estimador de Heckman produziram resultados, quanto aos coeficientes das variáveis explicativas, que, se bem que não atinjam a qualidade da estimação ML, evidenciam uma melhoria por respeito aos obtidos por OLS.

Note-se, por último, que uma debilidade do método da máxima verosimilhança, por referência aos fundados no paradigma dos mínimos quadrados, é, frequentemente, a sua maior sensibilidade à não verificação da hipótese da normalidade das perturbações. No exemplo que se discutiu, todavia, essa debilidade não poderia manifestar-se: a estimação por NLS teve por base as expressões que resultariam de uma distribuição normal truncada e o método de Heckman, quer na adopção de uma especificação *probit* para o primeiro passo, quer na construção de uma estimativa para o *inverse Mills ratio*, implicitamente assumiram o mesmo pressuposto quanto à forma da distribuição das perturbações.