
MODELOS DE ESCOLHA BINÁRIA

Wooldridge §7.5,17

0.1 Introdução

A teoria económica por vezes formaliza modelos que procuram explicar uma variável dependente que não é contínua e na qual pode existir truncagem no seu domínio. À custa de um conjunto de regressores, pode haver interesse em modelizar y_i , variável contínua ou discreta, na qual o espaço probabilístico (domínio em que a variável tem probabilidade não nula) é restrito ou truncado. Um exemplo é os gastos mensais na aquisição de viaturas em que se observa um conjunto de zeros para os indivíduos que não adquiriram um novo carro, $y_i \in \{0\} \cup (0, \infty)$. Em outros modelos, a variável é discreta (finita ou infinita) como por exemplo o número de carros que passam em n vias num dado período de tempo, $y_i \in \{0, 1, 2, \dots\}$. Noutros casos, procura-se modelizar uma variável escolha na qual a ordem é importante. Por exemplo, $y_i = 1$ quando o salário está entre $(0, 100]$, $y_i = 2$ para $(100, 200]$,..., e assim sucessivamente. De entre (muitas) outras tipologias salienta-se apenas mais uma: quando a variável dependente é de natureza qualitativa, não mensurável, e onde podem existir 2 ou mais atributos (ou escolhas) diferentes - Modelos multinomiais ou de escolha múltipla.

Nestes apontamentos, vamo-nos centrar na modelização de variáveis dependentes de natureza qualitativa, não mensurável, onde existem duas possibilidades e que é normalmente representada por uma variável Dummy, ou de escolha binária $\{0, 1\}$. Em vez de ser tratada como regressor, suponha-se que se procura modelizar uma variável do tipo Dummy.

Example 1 $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + u_i$, onde

$$y_i = \begin{cases} 1, & \text{se } i \text{ revela interesse em comprar/mudar casa} \\ 0, & \text{se } i \text{ não revela interesse em comprar/mudar casa} \end{cases};$$

x_{i2} é o rendimento anual; x_{i3} é o estado marital e x_{i4} é o número de filhos. Os outros fatores/características não observáveis estão no erro u_i .

Example 2 $y_i^* = x_i' \beta + u_i$, onde o que se observa é

$$y_i = \begin{cases} 1, & \text{se } i \text{ decide trabalhar (trabalho)} \\ 0, & \text{se } i \text{ decide não trabalhar (lazer)} \end{cases} = \begin{cases} 1, & \text{se a utilidade do trabalho } y_i^* \text{ é não negativa} \\ 0, & \text{se a utilidade do trabalho } y_i^* \text{ é negativa} \end{cases};$$

x_{i2} é o número de horas de trabalho oferecidas; x_{i3} é o salário oferecido ao indivíduo i ; Note-se que, à custa da função indicador $1(\cdot)$, y_i pode ser escrito em função da variável latente (que não é observada) y_i^* (des)utilidade do trabalho,

$$y_i = 1(y_i^* \geq 0) = 1(x_i' \beta + u_i \geq 0), \quad (1)$$

onde $y_i^* = Util_i^T - Util_i^L$, em que $Util_i^T$ é a utilidade que resulta da decisão de trabalhar e $Util_i^L$ de lazer. Desta forma, e supondo que $Util_i^T = x_i' \beta^T + u_i^T$ e $Util_i^L = x_i' \beta^L + u_i^L$, o modelo original $y_i^* = x_i' \beta + u_i$ resulta de $\beta = \beta^T - \beta^L$ e $u_i = u_i^T - u_i^L$.

Example 3 $y_i^* = x_i' \beta + u_i$, onde o que se observa (para um dado banco) é

$$y_i = \begin{cases} y_i^*, & \text{se o indivíduo } i \text{ contraiu um empréstimo} \\ 0, & \text{se o indivíduo } i \text{ não contraiu um empréstimo} \end{cases};$$

y_i^* é o montante de empréstimo contraído pelo indivíduo i ; x_{i2} é uma medida da capacidade de financiamento junto do banco; x_{i3} é a duração do empréstimo; Como se observa, y_i nunca toma valores negativos: alguns são positivos e há uma série de zeros. Para os indivíduos que não contraíram empréstimo supõe-se que não há informação (y_i^* pode ser interpretado como não observado para estes casos). Para o outro grupo de indivíduos (observados), $y_i^* = y_i$. Por esta especificidade dos dados, este é um exemplo de censura (censoring) à esquerda de zero na amostra. Neste tipo de regressões, o modelo é de variável dependente limitada. É fácil de verificar que neste modelo,

$$\begin{cases} y_i^* = x_i' \beta + u_i \\ y_i = \max(0, y_i^*) = 1(y_i^* \geq 0) y_i^* \end{cases}, i = 1, \dots, n. \quad (2)$$

Estes três exemplos ajudam a compreender a especificação geral (de alguns) dos modelos de escolha binária e que serão alvo de estudo em seguida.

0.2 O Modelo Probabilístico Linear

Tendo como base o MRLM com termo independente, o modelo probabilístico linear (MPL) assume que

$$y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i = x_i' \beta + u_i, i = 1, \dots, n \Leftrightarrow y = X\beta + u, \quad (3)$$

onde $y_i \equiv D_i = \begin{cases} 1, & \text{se } i \in A \\ 0, & \text{se } i \notin A \end{cases}$, A é um certo atributo que é observável para a i -ésima observação. Sob a hipótese de exogeneidade,

$$E(y_i|x_i) = x_i'\beta, i = 1, \dots, n \Leftrightarrow E(y|X) = X\beta. \quad (4)$$

Porque a variável dependente y_i apenas assume dois valores diferentes com probabilidade positiva, $y_i \in \{0, 1\}$, o valor esperado condicional (recta de regressão) tem uma interpretação muito própria no MPL. Defina-se

$$P(y_i = 1|x_i) = p(x_i) \equiv p_i; P(y_i = 0|x_i) = 1 - p_i, \quad (5)$$

onde p depende dos regressores x , implicitamente, e a amostra é aleatória (independente) e obtida da população de interesse. Então,

$$E(y_i|x_i) = x_i'\beta = 1 \cdot P(y_i = 1|x_i) + 0 \cdot P(y_i = 0|x_i) = P(y_i = 1|x_i) = p_i, \quad (6)$$

isto é, no MPL (e nos modelos de escolha binária, em geral), o valor esperado condicional corresponde à probabilidade (condicional a x) da variável dependente Dummy ser um (atributo A). No exemplo 1, $E(y_i|x_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$ é a probabilidade (condicional a x_{i2}, x_{i3}, x_{i4}) do individuo i revelar interesse em comprar/mudar casa.

Como o MPL é na essência um MRLM, a análise de interpretação, estimação, inferência e previsão no MPL pode ser reportada à que foi feita quando se estudou o MRLM. Não pretendendo ser exaustiva, salientamos agora os aspectos mais importantes em termos destas questões. Devido a (6), β_1 é a probabilidade de i ter o atributo A quando (condicional a) todos os regressores têm o valor de zero. Por outro lado, $\beta_j, j = 2, \dots, k$, é a variação da probabilidade de i ter o atributo A quando o regressor x_{ij} aumenta numa unidade, mantendo os outros regressores fixos (*ceteribus paribus*). Em termos de previsão pontual, a análise resulta imediatamente de (6).

Em termos de estimação e inferência a análise também é extremamente simples. Para além de exogeneidade e independência na amostra (o que implica não autocorrelação dos erros) suponha-se que não existe multicolinearidade perfeita no modelo (4). Aparte da questão da distribuição dos erros (que é fundamental nos outros modelos nestes apontamentos) porque se assume que existem muitas observações (n grande - invocar os (F)CLT e (U)LLN), falta discutir a variância condicional dos erros (ou de y). De uma forma semelhante a (6), sabe-se que

$$V(y_i|x_i) = E(y_i^2|x) - [E(y_i|x)]^2 = p_i - p_i^2 = p_i(1 - p_i), \quad (7)$$

que depende de x . Portanto, o MPL tem erros heterocedásticos e, como vimos no capítulo de heterocedasticidade, o OLS para (4) é consistente mas a inferência é errónea e não é BLUE. Quer se use o OLS com a matriz robusta de White ou se use o WLS, necessitamos apenas de p_i se for conhecido ou de uma estimativa para p_i se este não for conhecido (ver a expressão (7)). Neste último caso, e como (6), pode-se usar

$$\hat{p}_i = \hat{y}_i = E(\widehat{y}_i|x_i) = x_i'\widehat{\beta}, i = 1, \dots, n \quad (8)$$

onde $\widehat{\beta}$ é o estimador OLS.

Apesar da sua simplicidade e de não impôr uma distribuição para os erros, o MPL tem inconvenientes em dois níveis. Primeiro, não é garantido que para todo x_i , $E(y_i|x_i) = P(y_i = 1|x_i) \in [0, 1]$. Como o modelo é linear e mesmo que os declives parciais sejam baixos, para x_i muito grande (positivo ou negativo) $P(y_i = 1|x_i)$ é negativa ou maior do que um! Por exemplo, a probabilidade do individuo i revelar interesse em comprar/mudar casa pode ser negativa para um rendimento anual nulo. Uma solução será truncar o espaço dos x_i 's no qual $P(y_i = 1|x_i) \notin [0, 1]$: Fixar como zero quando $P(y_i = 1|x_i) = E(y_i|x_i) < 0$ é negativo e como um quando $P(y_i = 1|x_i) = E(y_i|x_i) > 1$. Em segundo lugar, no MPL assume-se que os declives parciais (variações nas probabilidades) são os mesmos (constantes) para diferentes x_i 's: $\frac{\partial p_i}{\partial x_{ij}} = \frac{\partial E(y_i|x_{ij})}{\partial x_{ij}} = \beta_j$, para todo x_{ij} . Ora, é possível conceber um modelo na qual a variação da probabilidade de i ter o atributo A altera-se consoante o valor do regressor em causa. Por exemplo, a variação da probabilidade do individuo i revelar interesse em comprar/mudar casa pode ser maior para rendimentos altos do que para rendimentos baixos. Associado a estas duas questões, está o facto de que a variação das probabilidades condicionais p_i poder ser maior do que 1 ou de -1 !

Devido a estes problemas, tem-se procurado utilizar especificações alternativas, nomeadamente os modelos Probit e Logit, para os quais se assume uma dada lei probabilística para os erros. A ideia é a seguinte: Especificar $E(y_i|x_i) = x_i'\beta = P(y_i = 1|x_i)$ à custa de uma função que esteja limitada ao espaço $[0, 1]$ para todo x_i e que não seja necessariamente linear. A solução mais óbvia é definir

$$E(y_i|x_i) = P(y_i = 1|x_i) = F(x_i'\beta), \quad (9)$$

em que $F(\cdot)$ é uma função distribuição *cdf* (monótona não decrescente e $F : x_i'\beta \rightarrow [0, 1]$). A diferença entre os modelos Probit e Logit é na função $F(\cdot)$: No Probit, $F(\cdot)$ é a função distribuição da normal standard; no Logit é a logística standard.

0.3 O Modelo Probit

A função distribuição de uma variável aleatória normal standard $N(0, 1)$ é dada por

$$F(z) = \Phi(z) = \int_{-\infty}^z \phi(t) dt = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad (10)$$

em que $\Phi(z)$ é monótona crescente, declive não constante e

$$\Phi(-\infty) = 0, \Phi(0) = 0.5, \Phi(1.96) = 0.975, \Phi(+\infty) = 1. \quad (11)$$

Portanto, o modelo Probit em que

$$y_i = E(y_i|x_i) + u_i, i = 1, \dots, n, \quad (12)$$

onde a variável dependente y_i é Dummy de escolha binária e observável, satisfaz

$$E(y_i|x_i) = P(y_i = 1|x_i) = \Phi(x_i'\beta) = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \quad (13)$$

No modelo Probit, com variável latente não observável, em que se assume que os erros são distribuídos $N(0, 1)$,

$$\begin{cases} y_i^* = x_i'\beta + u_i \\ y_i = 1(y_i^* \geq 0) \end{cases}, u_i|x_i \sim i.i.d.N(0, 1); i = 1, \dots, n, \quad (14)$$

também temos que

$$E(y_i|x_i) = P(y_i = 1|x_i) = \Phi(x_i'\beta). \quad (15)$$

A explicação é que, sendo o modelo estimável dado por

$$y_i = 1(x_i'\beta + u_i \geq 0), u_i|x_i \sim i.i.d.N(0, 1); i = 1, \dots, n, \quad (16)$$

se conclui que

$$P(y_i = 1|x_i) = P(y_i^* \geq 0|x_i) = P(u_i \geq -x_i'\beta|x_i) = 1 - P(u_i \leq -x_i'\beta|x_i) \quad (17)$$

$$= 1 - \Phi(-x_i'\beta) = \Phi(x_i'\beta). \quad (18)$$

Ao contrário do MPL, no modelo Probit (e Logit) as variações das probabilidades não são iguais (constante β_j) para todos os níveis dos regressores. No modelo Probit (15),

$$\frac{\partial P(y_i = 1|x_{i1}, \dots, x_{ij}, \dots, x_{ik})}{\partial x_{ij}} = \frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \beta_j \phi(x_i'\beta); j = 2, \dots, k \quad (19)$$

não é constante (existe o termo ϕ) e tem o mesmo sinal de β_j . No modelo com termo independente, β_1 , a interpretação para o mesmo é trivial. Note-se que para que os parâmetros do modelo (15) sejam identificados¹ temos de impôr que a variância dos erros é conhecida e igual a um (e não um parâmetro adicional σ^2 para estimar). Isto porque se $V(u_i) = \sigma^2$, desconhecido, então (15) era $\Phi\left(\frac{x_i'\beta}{\sigma}\right)$ que é igual a $\Phi\left(\frac{x_i'c\beta}{c\sigma}\right)$, $c > 0$ ². Neste caso, o modelo era o mesmo para qualquer combinação $(c\beta, c\sigma)$, $c > 0$.

Claramente, o modelo (15) não é linear nos regressores e portanto não pode ser estimado por OLS/(F)GLS. O método NLLS (mínimos quadrados não linear) (ou GMM) pode ser utilizado mas não é tão simples quanto o método da máxima verosimilhança (MLE). A função

¹Identificação significa a possibilidade de estimar todos os parâmetros de uma forma independente uns dos outros.

²Por definição, $V(u_i) > 0$ e por isso c não pode ser negativo ou zero.

log-verosimilhança de (y_1, \dots, y_n) ³, condicional a (x_1, \dots, x_n) , no modelo Probit é

$$\begin{aligned} \ell_n(\tilde{\beta}) &= \log L_n(\tilde{\beta}) = \log f_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \tilde{\beta}) = \log \prod_{i=1}^n f_{Y_i}(y_i; \tilde{\beta}) = \sum_{i=1}^n \log f_{Y_i}(y_i; \tilde{\beta}) \\ &= \sum_{i=1}^n \log \left(P(y_i|x_i; \tilde{\beta})^{y_i} \left(P(y_i|x_i; \tilde{\beta}) \right)^{1-y_i} \right) = \sum_{i=1}^n \log \left(\Phi(x'_i \tilde{\beta})^{y_i} [1 - \Phi(x'_i \tilde{\beta})]^{1-y_i} \right) \\ &= \sum_{i=1}^n \left(y_i \log \Phi(x'_i \tilde{\beta}) + (1 - y_i) \log [1 - \Phi(x'_i \tilde{\beta})] \right). \end{aligned} \quad (20)$$

O estimador de máxima verosimilhança (MLE) é, por isso,

$$\hat{\beta}_{MLE} = \arg \max_{\tilde{\beta}} \ell_n(\tilde{\beta}) = \arg \max_{\tilde{\beta}} \sum_{i=1}^n \left(y_i \log \Phi(x'_i \tilde{\beta}) + (1 - y_i) \log [1 - \Phi(x'_i \tilde{\beta})] \right); \quad (21)$$

o score para a i -ésima observação é

$$\frac{\partial \ell_{i,n}(\tilde{\beta})}{\partial \tilde{\beta}} = x_i \frac{y_i \phi(x'_i \tilde{\beta})}{\Phi(x'_i \tilde{\beta})} - x_i \frac{(1 - y_i) \phi(x'_i \tilde{\beta})}{1 - \Phi(x'_i \tilde{\beta})} = x_i \frac{\phi(x'_i \tilde{\beta}) [y_i - \Phi(x'_i \tilde{\beta})]}{\Phi(x'_i \tilde{\beta}) [1 - \Phi(x'_i \tilde{\beta})]}; \quad (22)$$

e a hessiana é

$$H_i = - \frac{\partial \ell_{i,n}(\tilde{\beta})'}{\partial \tilde{\beta}} \frac{\partial \ell_{i,n}(\tilde{\beta})}{\partial \tilde{\beta}} = -x_i x_i' \frac{\phi(x'_i \tilde{\beta})^2}{\Phi(x'_i \tilde{\beta}) [1 - \Phi(x'_i \tilde{\beta})]} \frac{[y_i - \Phi(x'_i \tilde{\beta})]^2}{\Phi(x'_i \tilde{\beta}) [1 - \Phi(x'_i \tilde{\beta})]} \quad (23)$$

$$= -x_i x_i' \frac{\phi(x'_i \tilde{\beta})^2}{\Phi(x'_i \tilde{\beta}) [1 - \Phi(x'_i \tilde{\beta})]} \frac{\tilde{u}_i^2}{V(y_i|x_i; \tilde{\beta})}. \quad (24)$$

Portanto, pode-se demonstrar que

$$\sqrt{n}(\hat{\beta}_{MLE} - \beta) \xrightarrow{d} N \left(0, \left[E \left(x_1 x_1' \frac{\phi_1^2}{\Phi_1 (1 - \Phi_1)} \right) \right]^{-1} \right), \quad (25)$$

onde $\phi_1 = \phi(x'_1 \beta)$, $\Phi_1 = \Phi(x'_1 \beta)$ e a onde a variância assintótica pode ser consistentemente estimada por

$$\left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \frac{\hat{\phi}_i^2}{\hat{\Phi}_i (1 - \hat{\Phi}_i)} \right)^{-1}, \quad (26)$$

onde $\hat{\phi}_i = \phi(x'_i \hat{\beta}_{MLE})$, $\hat{\Phi}_i = \Phi(x'_i \hat{\beta}_{MLE})$.

A inferência no método MLE baseia-se sobretudo (também se pode fazer rácios-t e testes F e Wald com base em (25)) na estatística LR (likelihood ratio) que resulta da comparação

³Com dados *i.i.d.*.

das verosimilhanças (20) entre dois modelos, o restrito (sob H_0) e o não restrito (modelo original). Para uma dada hipótese nula, com q restrições, a estatística de teste e a sua distribuição assintótica são

$$LR = 2 \left[\ell_n \left(\widehat{\beta}_{UR} \right) - \ell_n \left(\widehat{\beta}_R \right) \right] \xrightarrow{d} \chi_q^2. \quad (27)$$

O estimador MLE é naturalmente inconsistente se a distribuição dos erros não for a normal standard. O estudo de estimadores (semiparamétricos) robustos à distribuição no modelo Probit; o estudo quando erros e regressores não são independentes (versus endogeneidade); ou quando os erros são heterocedásticos não são discutidos nestes apontamentos.

0.4 O Modelo Logit

O modelo Logit assume uma expressão mais simples para a função distribuição dos erros. A *cdf* da logística é muito semelhante à da normal standard e tem a forma

$$F(z) = \Lambda(z) = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}; \quad (28)$$

$$\Lambda(-\infty) = 0, \Phi(0) = 0.5, \Phi(+\infty) = 1. \quad (29)$$

Tanto para o modelo Logit em que a variável dependente y_i é Dummy de escolha binária e observável como para o Logit com variável latente não observável, em que se assume que os erros são *i.i.d.* $(0, 1)$ standard logística, $\frac{\partial \Lambda(z)}{\partial z}$,

$$E(y_i | x_i) = P(y_i = 1 | x_i) = \Lambda(x_i' \beta) = \frac{1}{1 + e^{-x_i' \beta}}. \quad (30)$$

Portanto,

$$P(y_i = 0 | x_i) = 1 - P(y_i = 1 | x_i) = \frac{1}{1 + e^{x_i' \beta}} \quad (31)$$

o que implica que

$$\frac{P(y_i = 1 | x_i)}{P(y_i = 0 | x_i)} = e^{x_i' \beta} \Leftrightarrow x_i' \beta = \log \left(\frac{P(y_i = 1 | x_i)}{P(y_i = 0 | x_i)} \right) \quad (32)$$

é o logaritmo do rácio das probabilidades. Consequentemente,

$$\beta_1 = \log \left(\frac{P(y_i = 1 | x_{i2} = \dots = x_{ik} = 0)}{P(y_i = 0 | x_{i2} = \dots = x_{ik} = 0)} \right); \beta_j = \frac{\partial \log \left(\frac{P(y_i = 1 | x_i)}{P(y_i = 0 | x_i)} \right)}{\partial x_{ij}}, j = 2, \dots, k. \quad (33)$$

A análise e estimação do modelo Logit (30) é muito semelhante à do modelo Probit da secção anterior. Aliás, na prática, as diferenças entre os resultados obtidos com os dois modelos são relativamente pequenas. Salienta-se que, no modelo Logit,

$$\frac{\partial P(y_i = 1 | x_{i1}, \dots, x_{ij}, \dots, x_{ik})}{\partial x_{ij}} = \frac{\partial E(y_i | x_i)}{\partial x_{ij}} = \beta_j \frac{\partial \Lambda(z)}{\partial z}; j = 2, \dots, k \quad (34)$$

e que para o MLE,

$$\hat{\beta}_{MLE} = \arg \max_{\beta} \sum_{i=1}^n \left(y_i \log \Lambda \left(x_i' \tilde{\beta} \right) + (1 - y_i) \log \left[1 - \Lambda \left(x_i' \tilde{\beta} \right) \right] \right), \quad (35)$$

$$\sqrt{n} \left(\hat{\beta}_{MLE} - \beta \right) \xrightarrow{d} N \left(0, \left[E \left(x_1 x_1' \frac{\left(\frac{\partial \Lambda(z)}{\partial z} \right)^2}{\Lambda_1 (1 - \Lambda_1)} \right) \right]^{-1} \right). \quad (36)$$

0.5 O Modelo Tobit

Existem também modelos (ver exemplo 3) para os quais a variável a modelizar y_i^* não é necessariamente observável para todo o conjunto da amostra. Por exemplo, o domínio da variável rendimentos - y_i^* - pode não ser conhecido. Até um montante elevado (por exemplo, 100), o nível de rendimentos é conhecido e registado mas para valores superiores a 100 não existe controlo (informação) no seu registo e por isso todas essas observações são avaliadas como 100. Na amostra, num grupo de observações existe informação disponível mas para uma parte da amostra esta assume um único valor limite. Se, por exemplo, o valor limite é zero e a censura (*censoring*) na amostra está à esquerda de zero, então existe uma fracção da amostra na qual todas as observações são zero e a outra fracção são valores (distintos) maiores que zero.

O modelo Tobit, de variável dependente limitada e com variável latente y_i^* , tem a especificação

$$\begin{cases} y_i^* = x_i' \beta + u_i \\ y_i = \max(0, y_i^*) = 1(y_i^* \geq 0) y_i^* \end{cases}, i = 1, \dots, n. \quad (37)$$

Como se observa, a função densidade de y é mista: Em $y = 0$ é discreta e descontínua e para $y > 0$ é contínua. A estimação MLE de (37) requer a imposição de uma distribuição para os erros u_i . Supondo que estes são normalmente distribuídos, o modelo estimável é

$$y_i = 1(x_i' \beta + u_i \geq 0). \left(x_i' \beta + u_i \right), u_i | x_i \sim i.i.d. N(0, \sigma^2); i = 1, \dots, n, \quad (38)$$

onde, ao contrário do Probit e Logit, $E(y_i | x_i) \neq P(y_i = 1 | x_i)$ para todo x_i , e em que se pode demonstrar que

$$E(y_i | x_i) = E \left(\left(x_i' \beta + u_i \right) \cdot 1(u_i \geq -x_i' \beta) | x_i \right) = x_i' \beta \Phi \left(\frac{x_i' \beta}{\sigma} \right) + \sigma \phi \left(\frac{x_i' \beta}{\sigma} \right) \geq x_i' \beta, \quad (39)$$

o que implica que o OLS a $y_i = x_i' \beta + u_i$ é inconsistente. A estimação MLE (ou NLLS) é semelhante à apresentada para os modelos Probit e Logit mas a derivação das fórmulas é mais complicada, e por isso omitida nestes apontamentos. Endogeneidade e heterocedasticidade também não serão discutidas nestas páginas (tal como nos Probit e Logit). A estimação MLE de (37) baseia-se na função log-verosimilhança

$$\ell_n \left(\tilde{\beta}, \tilde{\sigma} \right) = \sum_{i=1}^n \left(1(y_i = 0) \log \left[1 - \Phi \left(\frac{x_i' \tilde{\beta}}{\tilde{\sigma}} \right) \right] + 1(y_i > 0) \log \left[\frac{1}{\tilde{\sigma}} \phi \left(\frac{y_i - x_i' \tilde{\beta}}{\tilde{\sigma}} \right) \right] \right) \quad (40)$$

porque⁴

$$P(y_i = 0|x_i) = 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right); P(y_i > 0|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right); \quad (41)$$

$$f_{y_i|x_i}(y_i|y_i > 0, x_i) = \phi(u_i) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right). \quad (42)$$

Como $E(y_i|x_i) \neq P(y_i = 1|x_i)$, não se pode interpretar e comparar com os Probit e Logit o valor de $\frac{\partial P(y_i=1|x_{i1}, \dots, x_{ij}, \dots, x_{ik})}{\partial x_{ij}} \neq \frac{\partial E(y_i|x_i)}{\partial x_{ij}}$! Sabe-se, no entanto, que para $j = 2, \dots, k$,

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \beta_j \left(\Phi\left(\frac{x_i'\beta}{\sigma}\right) + x_i'\beta \frac{1}{\sigma} \phi\left(\frac{x_i'\beta}{\sigma}\right) + \frac{\partial \phi(z)}{\partial z} \right) = \beta_j \Phi\left(\frac{x_i'\beta}{\sigma}\right); \quad (43)$$

$$\frac{\partial P(y_i = 0|x_{i1}, \dots, x_{ij}, \dots, x_{ik})}{\partial x_{ij}} = -\beta_j \frac{1}{\sigma} \phi\left(\frac{x_i'\beta}{\sigma}\right); \quad (44)$$

$$\frac{\partial P(y_i > 0|x_{i1}, \dots, x_{ij}, \dots, x_{ik})}{\partial x_{ij}} = \beta_j \frac{1}{\sigma} \phi\left(\frac{x_i'\beta}{\sigma}\right); \quad (45)$$

$$\frac{\partial f_{y_i|x_i}(y_i|y_i > 0, x_{i1}, \dots, x_{ij}, \dots, x_{ik})}{\partial x_{ij}} = \beta_j \frac{1}{\sigma^2} \frac{\partial \phi(z)}{\partial z}. \quad (46)$$

Note-se que β_j é o efeito de x_{ij} em y_i^* e não em y_i . A estimação consistente destes termos pode ser feita à custa de $\hat{\beta}_{MLE}, \hat{\sigma}_{MLE}$ desde que estes sejam estimados consistentemente.

0.6 Truncagem na Amostra

(por completar!) Missing observations: Por exemplo $E(y_i|y_i > 0, x_i)$. Sample selection.

...

Aplicações

(...)

Exercícios

1. Prove que se $V(u_i) = \sigma^2$, desconhecido, então o modelo Probit é $E(y_i|x_i) = P(y_i = 1|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right)$.

⁴ $f_{y_i|x_i}(y_i|y_i > 0, x_i) = \frac{f_{y_i|x_i}(y_i|x_i)}{P(y_i > 0|x_i)}$

2. Que modelo é mais adequado para explicar o tempo que um indivíduo demora até encontrar um emprego. Defina a amostra para este caso. Nota: Para algumas observações, o período da amostra termina antes de se encontrar um novo emprego!
3. ...