Capítulo 5 – Modelos de escolha Binária

Introdução, modelos linear e não lineares

Luís Silveira Santos

lsantos@iseg.ulisboa.pt

Mestrado em Econometria Aplicada e Previsão,
Instituto Superior de Economia e Gestão – Universidade de Lisboa

7 de Abril de 2016

Programa desta aula

Variáveis dependentes binárias

- Pretendemos estudar variáveis que assumem apenas dois conjuntos de resultados possíveis (tem ou não um dado atributo).
- Exemplos:
  1. $y_i = \begin{cases} 1 & \text{se o indivíduo } i \text{ está empregado} \\ 0 & \text{o.c.} \end{cases}$
  2. $y_i = \begin{cases} 1 & \text{se a empresa } i \text{ tem fundo de pensões próprio} \\ 0 & \text{o.c.} \end{cases}$
  3. $y_i = \begin{cases} 1 & \text{se a Família } i \text{ tem seguro de saúde} \\ 0 & \text{o.c.} \end{cases}$
  4. ...

Quantidades de interesse

- Neste contexto, pretendemos explicar a probabilidade de y exibir o atributo, em função de um conjunto de regressores $\mathbf{X}$:

$$P\left(Y=1 \mid \mathbf{X}\right)=p\left(\mathbf{x}\right)$$

- Por sua vez, estamos também interessados em obter os efeitos sobre a probabilidade de y exibir o atributo, quando um determinado regressor x_j ($j=1,\ldots,K$) varia uma unidade:
  - Se x_j é variável contínua:

$$\frac{\partial P\left(Y=1 \mid \mathbf{X}\right)}{\partial x_j}=\frac{\partial p\left(\mathbf{x}\right)}{\partial x_j}, \quad j=1,\ldots,K$$

  - Se x_j é variável discreta:

$$p\left(x_1,x_2,\ldots,x_{j-1},c_j+1\right)-p\left(x_1,x_2,\ldots,x_{j-1},c_j\right)$$

    onde c_j é um atributo da variável x_j (ex.: determinar o efeito na probabilidade de "sucesso" quando se tem mais um filho)

Características das V.A.s com distribuição de Bernoulli

- Uma vez que $y \in \{0, 1\}$, sabemos que $Y \mid \mathbf{X} \sim Ber(\theta)$
- Neste caso em concreto teremos $\theta = p(\mathbf{x})$
- Os seus momentos são:
  - $E(Y \mid \mathbf{X}) = p(\mathbf{x})$
  - $Var(Y \mid \mathbf{X}) = p(\mathbf{x})[1 - p(\mathbf{x})]$
- Pela natureza de y, sabemos ainda que:
  - $P(Y = 0 \mid \mathbf{X}) = 1 - P(Y = 1 \mid \mathbf{X})$
  - Exibe heterocedasticidade natural, excepto no caso em que $p(\mathbf{x})$ não dependa de $\mathbf{X}$, i.e., $P(Y = 1 \mid \mathbf{X}) = P(Y = 1)$

Que modelo ajustar?

- A principal questão relativamente ao estudo das variáveis binárias é determinar qual o tipo de modelo a ajustar aos dados.
- Modelos lineares vs. modelos não lineares

Especificação

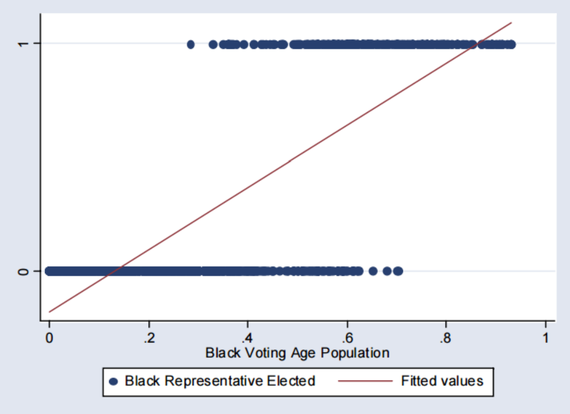
- O MPL estabelece que a probabilidade de "sucesso" é função linear de um conjunto de K regressores:
$$P(Y=1 \mid \mathbf{X}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K = \mathbf{X}\boldsymbol{\beta}$$
em que $\mathbf{X}$ é matriz $N \times K$.

- Sabemos que $P(Y=1 \mid \mathbf{X}) = E(Y \mid \mathbf{X})$. Assim sendo, dentro do contexto das hipóteses clássicas (Wooldridge, 2013), as estimativas OLS do vector de parâmetros $\boldsymbol{\beta}$ são centradas e consistentes

- No contexto linear, os efeitos parciais são dados pela fórmula "tradicional":
$$\frac{\partial P(Y=1 \mid \mathbf{X})}{\partial x_j} = \beta_j, \quad j = 1, \ldots, K$$

Problemas

- Observemos graficamente a seguinte relação entre uma variável binária e uma variável contínua:



Problemas (cont.)

1. Os valores ajustados, $P\,(\widehat{Y = 1} \mid \mathbf{X}) = \mathbf{X}\hat{\beta} \notin [0,1]$
2. Variações sucessivas em x_j, *ceteris paribus*, implicam variações sucessivas de $P\,(Y = 1 \mid \mathbf{X})$, nos valores de β_j, conduzindo a situações em que $P\,(Y = 1 \mid \mathbf{X}) \notin [0,1]$
3. Heterocedasticidade natural,
$$Var\,(Y \mid \mathbf{X}) = P\,(Y = 1 \mid \mathbf{X})\,[1 - P\,(Y = 1 \mid \mathbf{X})] =$$
$$= \mathbf{X}\beta\,(1 - \mathbf{X}\beta)$$
4. Uma vez que hipótese de normalidade não se verifica, teremos que nos basear na teoria assimptótica para dedução das distribuições assimptóticas dos estimadores e das estatísticas de teste

Problemas (cont.)

- Na prática, apenas o problema 2 (sobre efeitos parciais) é verdadeiramente importante, no entanto este é consequência de estarmos a assumir um modelo linear
- O problema 3 (sobre heterocedasticidade natural) pode ser resolvido por duas vias:
  - Erros-padrão robustos de White
  - Estimação por GLS (uma vez que a forma teórica da matriz de covariância é conhecida)

Observações

- Se estimarmos o OLS sem corrigir os erros-padrão, existe apenas uma situação em que a estatística F é válida:
$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0$$
uma vez que, sob H_0, $Var\left(Y \mid \mathbf{X}\right) = \beta_0\left(1 - \beta_0\right)$

- Podemos estimar $P\left(Y = 1 \mid \mathbf{X}\right)$ através do WLS, utilizando como ponderador:
$$\left[\mathbf{x}_i\hat{\beta}\left(1 - \mathbf{x}_i\hat{\beta}\right)\right]^{-1/2}$$
Note-se que este ponderador é específico para cada indíviduo i

- O MPL pode ser interpretado como a melhor aproximação linear (em erro quadrático médio), à verdadeira probabilidade $p\left(\mathbf{x}\right)$

Especificação

- Pretende-se agora estudar modelos não lineares da probabilidade de "sucesso":
$$P(Y=1 \mid \mathbf{X}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K) \\ = G(\mathbf{X}\beta) \equiv p(\mathbf{x})$$
onde se assume que $G : \mathbb{R} \to [0,1]$

- Neste caso, os efeitos parciais já refletem a natureza não linear do modelo:
$$\frac{\partial p(\mathbf{x})}{\partial x_j} = \beta_j \times g(\mathbf{X}\beta)$$
onde $g(\cdot)$ é a primeira derivada da função $G(\cdot)$. Estas funções dependem de todos os valores observados de $\mathbf{X}$.

- Na maior parte das aplicações, assume-se que G é uma função de distribuição, cuja forma funcional depende do problema económico em questão

Modelo de variável latente

- Considere o seguinte modelo de variável latente:

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, N$$

onde,

  - y_i^* é variável dependente latente escalar
  - $\mathbf{x}_i$ é vector de regressores, $1 \times K$
  - $\boldsymbol{\beta}$ é vector de parâmetros, $K \times 1$
  - $\varepsilon_i \perp \mathbf{x}_i, \forall i$

Exemplo: compra de carro

- Vamos supor que pretendemos explicar a utilidade de um determinado indivíduo adquirir um veículo automóvel:

$$U_{\text{Comprar},i} = \mathbf{x}_i \gamma + \varepsilon_{1i}$$

e

$$U_{\text{Não comprar},i} = \mathbf{x}_i \boldsymbol{\delta} + \varepsilon_{2i}$$

Exemplo: compra de carro (cont.)

- Na prática não observamos **quantitativamente** a utilidade que um dado indivíduo, i, retira das suas decisões, mas sim o resultado da sua decisão:

$$
\begin{aligned}
y_i^* &= U_{\text{Comprar}} - U_{\text{Não comprar}} = \\
&= \mathbf{x}_i \underbrace{(\boldsymbol{\gamma} - \boldsymbol{\delta})}_{\boldsymbol{\beta}} + \underbrace{(\varepsilon_{1i} - \varepsilon_{2i})}_{\varepsilon_i}
\end{aligned}
$$

- Ou seja:

$$
y_i = \begin{cases} 1 & \text{se } y_i^* > 0 \\ 0 & \text{o.c.} \end{cases}
$$

Exemplo: compra de carro (cont.)

- Neste contexto, estamos interessados em calcular a probabilidade de um dado indivíduo adquirir o veículo automóvel, i.e.:

$$\begin{aligned} P(Y = 1 \mid \mathbf{X}) &= P(Y^* > 0 \mid \mathbf{X}) \\ &= P(\mathbf{X}\beta + \varepsilon > 0 \mid \mathbf{X}) \\ &= P(\varepsilon > -\mathbf{X}\beta \mid \mathbf{X}) \quad \text{(HIP.: distrib. de } \varepsilon \text{ é simétrica)} \\ &= P(\varepsilon < \mathbf{X}\beta \mid \mathbf{X}) \\ &= G(\mathbf{X}\beta) \end{aligned}$$

- A estimação desta probabilidade depende da distribuição assumida para o termo de erro (condicionado pelos regressores), $\varepsilon \mid \mathbf{X}$.

Exemplo: compra de carro (cont.)

- Habitualmente consideramos $\varepsilon \mid \mathbf{X} \sim N\left(0, \sigma_\varepsilon^2\right)$, no entanto esta decisão pode conduzir a problemas adicionais:

$$P\left(\varepsilon < \mathbf{X}\beta \mid \mathbf{X}\right) = P\left(\frac{\varepsilon}{\sigma_\varepsilon} < \frac{\mathbf{X}\beta}{\sigma_\varepsilon} \,\middle|\, \mathbf{X}\right)$$
$$= P\left(\frac{\varepsilon}{\sigma_\varepsilon} < \frac{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K}{\sigma_\varepsilon} \,\middle|\, \mathbf{X}\right)$$
$$= P\left(\frac{\varepsilon}{\sigma_\varepsilon} < \frac{\beta_0}{\sigma_\varepsilon} + \frac{\beta_1}{\sigma_\varepsilon} x_1 + \frac{\beta_2}{\sigma_\varepsilon} x_2 + \ldots + \frac{\beta_K}{\sigma_\varepsilon} x_K \,\middle|\, \mathbf{X}\right)$$

- O vector de parâmetros a ser estimado é:

$$\boldsymbol{\theta} = \left(\sigma_\varepsilon^2, \beta_0, \beta_1, \beta_2, \ldots, \beta_K\right)$$

Exemplo: compra de carro (cont.)

- No entanto, pela especificação do modelo e da distribuição condicionada $\varepsilon \mid \mathbf{X}$, não é possível estimar separadamente σ_ε^2 dos restantes βs

- Como em geral se desconhece σ_ε^2, porque não observamos ε, o que estamos a estimar?

$$(\beta_0, \beta_1, \beta_2, \ldots, \beta_K) \text{ OU } \left(\frac{\beta_0}{\sigma_\varepsilon}, \frac{\beta_1}{\sigma_\varepsilon}, \frac{\beta_2}{\sigma_\varepsilon}, \ldots, \frac{\beta_K}{\sigma_\varepsilon}\right)$$

$$\Downarrow$$

PROBLEMA DE IDENTIFICAÇÃO

- A única quantidade que sabemos <u>de certeza</u> são as *odds*:

$$\frac{\partial p(\mathbf{x}) / \partial x_j}{\partial p(\mathbf{x}) / \partial x_h} = \frac{\beta_j \phi(\cdot)}{\beta_h \phi(\cdot)} = \frac{\beta_j}{\beta_h}, \quad h \neq j; \; h, j = 1, 2, \ldots, K$$

Formas da função $G(\cdot)$

- A especificação da função G depende da hipótese assumida sobre a distribuição do termo de erro
- No entanto, ao contrário dos modelos lineares, não basta assumir hipóteses genéricas sobre os parâmetros dessa distribuição, uma vez que podem conduzir a problemas de identificação
- Na literatura surgiram duas especificações para a distribuição $\varepsilon \mid \mathbf{X}$, que devido à sua simplicidade e por acomodar a maioria dos problemas económicos são bastante utilizadas

Formas da função $G(\cdot)$ (cont.)

- $\varepsilon \mid \mathbf{X} \sim N(0,1)$:

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(z)\,\mathrm{d}z$$

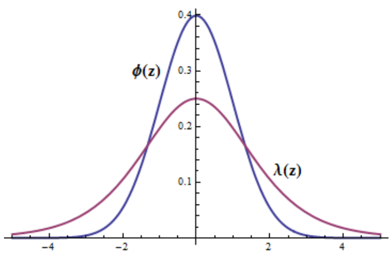
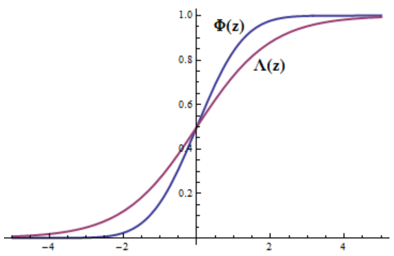
onde $\phi(z) = (2\pi)^{-1/2} \exp\left(-z^2/2\right)$, a função densidade da distribuição Normal estandardizada, $Var(\varepsilon \mid \mathbf{X}) = 1$

- $\varepsilon \mid \mathbf{X} \sim Logistic(0,1)$:

$$G(z) = \Lambda(z) = \frac{\exp(z)}{1 - \exp(z)}$$

onde $\Lambda(z)$ é a função densidade da distribuição Logística estandardizada, $Var(\varepsilon \mid \mathbf{X}) = \pi^2/3$

Formas da função $G(\cdot)$ (cont.)

PDF	*CDF*
	

Observações

- Atente-se ao facto de não estarmos interessados nos efeitos parciais de x_j sobre y_i^* mas sim nos efeitos parciais de x_j sobre y_i, dado que:
  - Não observamos y_i^* mas sim y_i
  - Mesmo que observássemos y_i^*, a sua unidade de medida seria pouco clara, ou até inexistente (em especial, o exemplo da utilidade)

- Adicionalmente, destaca-se a utilidade da estimação por via de um modelo linear, na medida em que a **direcção** dos efeitos parciais é igual à versão dos efeitos parciais num modelo não linear:

$$\frac{\partial p(\mathbf{x})}{\partial x_j} = \beta_j \quad \text{vs.} \quad \frac{\partial p(\mathbf{x})}{\partial x_j} = \beta_j \times \underbrace{g(\mathbf{X}\beta)}_{>0}$$