Capítulo 5 – Modelos de escolha Binária

Efeitos parciais, Testes de hipóteses, Qualidade do ajustamento

Luís Silveira Santos

lsantos@iseg.ulisboa.pt

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

Abril de 2017

Programa desta aula

1. Efeitos parciais
   - Efeito parcial na média
   - Efeito parcial médio
   - Cálculo dos erros-padrão
2. Testes de hipóteses
   - Testes de restrições de exclusão simples ou múltiplas
   - Testes de hipóteses não lineares
3. Interpretação dos resultados
   - Qualidade das previsões
   - R^2 de McFadden
   - Probit vs. Logit vs. MPL

Efeitos parciais

- Relativamente ao estudo dos efeitos parciais, recordemos que, num contexto de modelos não lineares, estes são dados por:

$$\frac{\partial P\left(Y=1 \mid \mathbf{X}\right)}{\partial x_j} = \frac{\partial p\left(\mathbf{x}\right)}{\partial x_j} = \beta_j \times g\left(\mathbf{X}\beta\right)$$

- No entanto não iremos obter apenas um valor para estes efeitos parciais, uma vez que a função $g\left(\mathbf{X}\beta\right)$ depende de todos os elementos da matriz $\mathbf{X}$

$$\Downarrow$$

De que forma podemos interpretar estes efeitos parciais?

Efeitos parciais (cont.)

- A solução passa por calcular dois tipos de efeitos parciais:
  1. Efeito parcial na média (PEA)
  2. Efeito parcial médio (APE)

Efeito parcial na média

- Neste contexto, vamos calcular a função $g(\mathbf{X}\beta)$ para valores "representativos" de cada elemento da matriz $\mathbf{X}$:
  1. Médias
  2. Máximos e Mínimos
  3. Quantis
- Desta forma podemos:
  1. Analisar os efeitos parciais no indivíduo "médio"
  2. Proceder a uma análise de sensibilidade dos efeitos parciais para certos valores de x_j, $j = 1, 2, \ldots, K$

Efeitos parciais na média (cont.)

- Se x_j é variável contínua:

$$\widehat{PEA}_j = \frac{\partial \widehat{P\left(Y = 1 \mid \mathbf{X} = \overline{\mathbf{X}}\right)}}{\partial x_j} = \hat{\beta}_j \times g\left(\overline{\mathbf{X}}\hat{\beta}\right)$$

onde $\overline{\mathbf{X}}$ é vector $1 \times K$ das médias de cada um dos regressores.

- Se x_j é variável discreta:

$$\widehat{PEA}_j = G\left[\hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \hat{\beta}_2\overline{x}_2 + \ldots + \hat{\beta}_{j-1}\overline{x}_{j-1} + \hat{\beta}_j\left(c_j + 1\right)\right] - \\ - G\left(\hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \hat{\beta}_2\overline{x}_2 + \ldots + \hat{\beta}_{j-1}\overline{x}_{j-1} + \hat{\beta}_j c_j\right)$$

- Quando x_j é variável binária, podemos utilizar o PEA_j anterior, fixando $c_j = 0$

Efeito parcial médio

- Uma alternativa passa por calcular a média da função $g\left(\mathbf{X}\boldsymbol{\beta}\right)$:

$$APE_j = \beta_j \times E\left[g\left(\mathbf{X}\boldsymbol{\beta}\right)\right]$$

- Se x_j é variável contínua, um estimador consistente será:

$$\widehat{APE}_j = \hat{\beta}_j \left[\frac{1}{N}\sum_{i=1}^{N} g\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right]$$

- Se x_j é variável discreta, um estimador consistente será:

$$\widehat{APE}_j = \frac{1}{N}\sum_{i=1}^{N}\left\{G\left[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_{j-1}x_{i,j-1} + \hat{\beta}_j\left(c_j+1\right)\right] - \right.$$
$$\left. -G\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_{j-1}x_{i,j-1} + \hat{\beta}_j c_j\right)\right\}$$

- Quando x_j é variável binária, fixamos $c_j = 0$ no APE_j anterior

Cálculo dos erros-padrão

- A maioria dos *softwares* econométricos permite-nos realizar inferência sobre as estimativas dos *PEA*s e dos *APE*s, através do cálculo dos seus erros-padrão
- No entanto, pela complexidade das fórmulas dos *PEA*s e dos *APE*s, os seus erros-padrão não têm uma forma funcional fechada, ao contrário do que é habitual
- Assim, estes erros-padrão são calculados numericamente através de duas metodologias:
  - Método Delta $\rightarrow$ aproximação via série de Taylor de 1ª ordem
  - *Bootstrap* $\rightarrow$ simulação de R amostras

Testes de hipóteses

- Um outro tópico de interesse prende-se com a realização de testes de hipóteses sobre os parâmetros do modelo estimado
- Neste contexto iremos dar enfoque a dois tipos de testes:
  1. Testes de restrições de exclusão simples ou múltiplas
  2. Testes de hipóteses não lineares
- As estatísticas de teste que seguidamente se irão apresentar são assimptoticamente equivalentes, porém algumas delas poderão ser computacionalmente difíceis de se obter

Testes de restrições de exclusão simples ou múltiplas

- Considere-se o seguinte modelo:

$$P\left(Y=1 \mid \mathbf{X}, \mathbf{Z}\right)=G\left(\mathbf{X}\beta+\mathbf{Z}\gamma\right)$$

  onde $\mathbf{Z}$ é matriz $N \times Q$ de regressores adicionais e γ é vector de dimensão $Q \times 1$

- Pretendemos testar a hipótese:

$$H_0 : \gamma = 0$$

- Existem três estatísticas de teste que podem ser consideradas:
  1. Estatística de Wald
  2. Rácio de verosimilhanças (ou teste LR)
  3. Teste de Score (ou teste LM)

Testes de restrições de exclusão simples ou múltiplas (cont.)

1. **ESTATÍSTICA DE WALD:**

$$W = (R\hat{\gamma} - r)' \left(R\hat{V}R' \right)^{-1} (R\hat{\gamma} - r) \xrightarrow{d} \chi^2_{(Q)}$$

onde $\hat{V}$ é a matriz de covariância assimptótica do modelo com $G(\mathbf{X}\beta + \mathbf{Z}\gamma)$

2. **RÁCIO DE VEROSIMILHANÇAS:**

$$LR = 2(\ell_{ur} - \ell_r) \xrightarrow{d} \chi^2_{(Q)}$$

onde ℓ_{ur} é a função log-verosimilhança para o modelo irrestrito, $G(\mathbf{X}\beta + \mathbf{Z}\gamma)$, e ℓ_r é a função log-verosimilhança para o modelo restrito, $G(\mathbf{X}\beta)$

Testes de restrições de exclusão simples ou múltiplas (cont.)

3. **TESTE DE SCORE:**

$$LM = N \times R_{\hat{u}}^2 \xrightarrow{d} \chi^2_{(Q)}$$

onde $R_{\hat{u}}^2$ é o R^2 não centrado da regressão auxiliar, estimada por OLS, de:

$$\frac{\hat{u}_i}{\sqrt{\hat{G}_i\left(1-\hat{G}_i\right)}} \text{ sobre } \frac{\hat{g}_i}{\sqrt{\hat{G}_i\left(1-\hat{G}_i\right)}}\mathbf{x}_i \text{ e } \frac{\hat{g}_i}{\sqrt{\hat{G}_i\left(1-\hat{G}_i\right)}}\mathbf{z}_i$$

com $\hat{u}_i = y_i - G\left(\mathbf{x}_i\hat{\beta}\right)$, $\hat{G} = G\left(\mathbf{x}_i\hat{\beta}\right)$ e $\hat{g} = g\left(\mathbf{x}_i\hat{\beta}\right)$

$\Rightarrow$ Note-se que se pondera por $\left[\hat{G}_i\left(1-\hat{G}_i\right)\right]^{-1/2}$ pois, sob H_0,

$$Var\left(u_i \mid \mathbf{X}, \mathbf{Z}\right) = G\left(\mathbf{x}_i\beta\right)\left[1 - G\left(\mathbf{x}_i\beta\right)\right]$$

Testes de hipóteses não lineares

- Relativamente aos testes de hipóteses não lineares, podemos utilizar as estatísticas anteriores. No entanto, existem problemas computacionais associados ao seu cálculo
- Em concreto, impor-se restrições não lineares na estimação dos modelos Probit e Logit eleva substancialmente o nível de complexidade computacional para o procedimento de maximização da função de verosimilhança, com implicações directas no cálculo das estatísticas de teste LM e LR
- Resta-nos, portanto, a estatística de Wald. No entanto, esta estatística não é invariante a reparameterizações (ao contrário das estatísticas de teste LM e LR), traduzindo-se em fracas propriedades em amostras finitas

Testes de hipóteses não lineares (cont.)

- Ainda assim, iremos optar por calcular a estatística de Wald, uma vez que os resultados assimptóticos permanecem válidos

- Tendo por base o modelo com $G(\mathbf{X}\beta)$, vamos considerar que pretendemos testar uma restrição não linear genérica, i.e.,

$$H_0 : \mathbf{c}(\beta) = 0$$

onde $\mathbf{c}(\beta)$ é vector de dimensão $K_1 \times 1$, com $K_1 \leq K$

- Neste caso, a estatística de Wald é dada por:

$$W = \mathbf{c}\left(\hat{\beta}\right)' \left(\nabla_\beta \mathbf{c}\left(\hat{\beta}\right) \hat{V} \; \nabla_\beta \mathbf{c}\left(\hat{\beta}\right)'\right)^{-1} \mathbf{c}\left(\hat{\beta}\right) \xrightarrow{d} \chi^2_{(K_1)}$$

onde $\nabla_\beta \mathbf{c}\left(\hat{\beta}\right)$ é o jacobiano do vector $\mathbf{c}(\beta)$ avaliado em $\hat{\beta}$

Qualidade do ajustamento

- A avaliação do potencial do modelo estimado é também um aspecto a ter em conta na nossa análise
- Para tal, temos duas medidas da qualidade do ajustamento:
  1. Qualidade das previsões
  2. R^2 de McFadden (ou pseudo-R^2)
- Note-se, porém, que não devemos esperar valores convincentes destas medidas

Qualidade das previsões

- Sabe-se que:

$$\hat{y}_i = \begin{cases} 1 & \text{se } G\left(\mathbf{x}_i\hat{\beta}\right) \geq 0.5 \\ 0 & \text{se } G\left(\mathbf{x}_i\hat{\beta}\right) < 0.5 \end{cases}$$

- Vamos então definir:
  - $N_0 = \#\left(y_0\right)$, número de "insucessos" na amostra
  - $N_1 = \#\left(y_1\right)$, número de "sucessos" na amostra
  - $N_{00} = \#\left(\hat{y}_0\right)$, número de "insucessos" bem previstos
  - $N_{11} = \#\left(\hat{y}_1\right)$, número de "sucessos" bem previstos

Qualidade das previsões (cont.)

- Uma medida da qualidade das previsões envolve a relação entre a proporção dos "sucessos" e dos "insucessos" bem previstos:

$$q_0 = \frac{N_{00}}{N_0} \text{ e } q_1 = \frac{N_{11}}{N_1}$$

- Quando uma das proporções acima apresenta um valor demasiado baixo, devemos ajustar o *threshold* das probabilidades estimadas: em vez de 0.5 podemos passar a considerar a média de y, $\bar{y}$, uma vez que esta é um estimador consistente para a probabilidade de sucesso não condicionada

- A proporção total de previsões correctas é dada por:

$$q = \frac{N_{00} + N_{11}}{N} = \frac{N_0}{N} q_0 + \frac{N_1}{N} q_1$$

R^2 de McFadden

- O R^2 de McFadden é dado por

$$R^2_{\text{Pseudo}} = 1 - \frac{\ell_{ur}}{\ell_0}$$

onde ℓ_{ur} é a função log-verosimilhança para o modelo estimado e ℓ_0 é a função log-verosimilhança para o modelo apenas com a constante

Probit vs. Logit vs. MPL

- A escolha entre o Probit e o Logit é, em geral, difícil, quando o verdadeiro modelo não é conhecido à partida
- Uma vez que ambos têm implícita uma função de distribuição, a decisão recai sobre a caracterização dos valores extremos dos dados
- Nos casos em que exista uma grande densidade de valores extremos, devemos optar pelo Logit, uma vez que a distribuição Logística estandardizada tem caudas mais pesadas

Probit vs. Logit vs. MPL (cont.)

- No entanto, é possível comparar as magnitudes das estimativas entre os três modelos através de factores de escala:
  - MPL $\rightarrow g(0) = 1$
  - Probit $\rightarrow g(0) = \dfrac{1}{\sqrt{2\pi}} \approx 0.4$
  - Logit $\rightarrow g(0) = \dfrac{\exp(0)}{1+\exp(0)} \times \left[1 - \dfrac{\exp(0)}{1+\exp(0)}\right] = 0.25$
- **EXEMPLO:** dividir as estimativas Logit por $0.4/0.25 = 1.6$ por forma a torná-las comparáveis com as estimativas Probit