LISBOA SCHOOL OF ECONOMICS & MANAGEMENT

Econometria

Qualidade do ajustamento e Propriedades do estimador OLS

Luís Silveira Santos

lsantos@iseg.ulisboa.pt

Licenciatura em Economia,
Instituto Superior de Economia e Gestão – Universidade de Lisboa

21 de Abril de 2015

Programa desta aula

Qualidade do ajustamento

- Recorde-se que através Econometria conseguimos responder a duas questões-chave:
  - Previsão do valor da variável dependente (ou explicada), dados os valores das variáveis independentes (ou explicativas);
  - Quantificar a relação (ou relações) entre a variável dependente e as variáveis independentes
- Para tal, necessitamos de medir a qualidade do ajustamento, ou a "fiabilidade" do modelo estimado relativamente aos dados de que dispomos

Qualidade do ajustamento (cont.)

- Considere-se a expressão do modelo de regressão linear múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- Ao estimarmos o modelo acima vamos obter:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

sendo que $\hat{u}_i = y_i - \hat{y}_i$

- Ou seja, podemos facilmente escrever y_i como uma combinação linear do valor previsto, $\hat{y}_i$, e do resíduo, $\hat{u}_i$, do modelo estimado:

$$y_i = \hat{y}_i + \hat{u}_i$$

Qualidade do ajustamento (cont.)

- Consideremos agora a expressão da variação total de y_i:

$$\sum_{i=1}^{n}(y_i-\overline{y})^2$$

- Dado que $y_i=\hat{y}_i+\hat{u}_i$:

$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}\left[\hat{y}_i+\hat{u}_i-\left(\overline{\hat{y}}+\overline{\hat{u}}\right)\right]^2$$

- Como $\overline{\hat{y}}=\overline{y}$ e $\overline{\hat{u}}=0$:

$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}\left[\hat{y}_i+\hat{u}_i-\left(\overline{\hat{y}}+\overline{\hat{u}}\right)\right]^2=\sum_{i=1}^{n}[(\hat{y}_i-\overline{y})+\hat{u}_i]^2=$$

$$=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2-2\times\sum_{i=1}^{n}\hat{u}_i\left(\hat{y}_i-\overline{y}\right)+\sum_{i=1}^{n}\hat{u}_i^2$$

Qualidade do ajustamento (cont.)

- **PROBLEMA:**

$$\sum_{i=1}^{n} \hat{u}_i \left(\hat{y}_i - \overline{y}\right) = ?$$

- Sabemos que a covariância entre cada uma das variáveis independentes, $\{x_1, x_2, \ldots, x_k\}$, e os resíduos OLS, $\hat{u}_i$ é igual a zero

- Logo, como $\hat{y}_i$ é combinação linear destas variáveis independentes, também a sua covariância entre os resíduos OLS, $\hat{u}_i$, será zero.

- Assim sendo,

$$\sum_{i=1}^{n} \hat{u}_i \left(\hat{y}_i - \overline{y}\right) = 0$$

Qualidade do ajustamento (cont.)

- Por fim teremos que a variação total de y_i poderá ser reescrita da seguinte forma:

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}\hat{u}_i^2$$

- Ou analogamente,

$$\text{SQ Total} = \text{SQ Explicada} + \text{SQ dos Resíduos}$$

Qualidade do ajustamento (cont.)

- O coeficiente que mede a qualidade do ajustamento do modelo estimado é designado por R^2, sendo dado por:

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT}$$

- Ou seja, mede a proporção da variação total dos dados que é explicada pelo modelo estimado

- **PROPRIEDADES:**
  1. $0 \leq R^2 \leq 1$
  2. Aumenta sempre que adicionamos variáveis ao modelo (**!!!**)

Exemplos

1. Salário do CEO:

$$\widehat{\text{salário}}_i = 964 + 19\,\text{RCP}_i$$

$$n = 210,\ R^2 = 0.01$$

2. Média final de Licenciatura:

$$\widehat{\text{colGPA}}_i = 1.3 + 0.5\,\text{hsGPA}_i + 0.01\,\text{ACT}_i$$

$$n = 140,\ R^2 = 0.18$$

Hipóteses do MRLM

Hipótese MLR.1 – Linearidade nos parâmetros

O modelo na população pode ser escrito como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

onde $\beta_0, \beta_1, \ldots, \beta_k$ são parâmetros desconhecidos constantes de interesse e u é um erro aleatório não observado.

- Sob esta hipótese são admitidos os seguintes modelos:
  1. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + u$
  2. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \sqrt{x_2} + u$
- <u>Não</u> são admitidos os seguintes modelos:
  3. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3^2 x_3 + u$
  4. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \sqrt{\beta_3} x_3 + u$

Hipóteses do MRLM (cont.)

Hipótese MLR.2 – Amostragem aleatória

Temos uma amostra aleatória com n observações,

$$\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : i = 1, 2, \ldots, n\}$$

seguindo o modelo populacional da hipótese MLR.1

Hipóteses do MRLM (cont.)

Hipótese MLR.3 – Ausência de colinearidade perfeita

Na amostra (tal como na população), nenhuma das variáveis independentes é constante e não existem relações lineares exactas entre as variáveis independentes

- Esta hipótese admite que exista correlação entre as diferentes variáveis independentes, ou regressores
- Porém, existem dois casos em que esta hipótese é violada:
  1. $|Cov(x_p, x_q)| = 1, \forall p \neq q$
  2. Se $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$, com $x_1 = x_2 + x_3$

Hipóteses do MRLM (cont.)

Hipótese MLR.4 – Média condicional igual a zero

O erro u tem valor esperado igual a zero, dados quaisquer valores das variáveis independentes:

$$E\left(u \mid x_1, x_2, \ldots, x_k\right) = 0$$

- Esta hipótese será violada se omitirmos propositadamente do modelo uma variável relevante para explicar y <u>e</u> que esteja correlacionada com alguma das variáveis independentes, x_k
- **EXEMPLO** – vamos omitir propositadamente x_3 do modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \quad u = \delta_3 x_3 + v$$

onde v é o novo termo de erro.

Assim, $E\left(u \mid x_1, x_2, \ldots, x_k\right) = \delta_3 x_3 + \underbrace{E\left(v \mid x_1, x_2, \ldots, x_k\right)}_{=0,\ \text{pela hipótese MLR.4}} \neq 0$

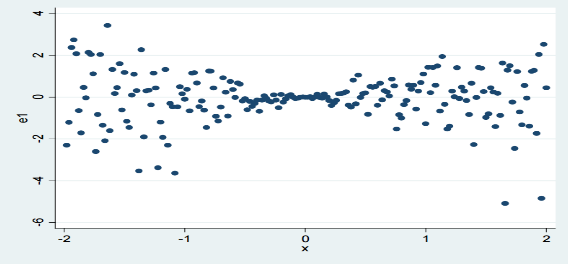
Hipóteses do MRLM (cont.)

Hipótese MLR.5 – Homocedasticidade

O erro u apresenta a mesma variância, dados quaisquer valores das variáveis independentes:

$$Var\left(u \mid x_1, x_2, \ldots, x_k\right) = \sigma^2$$

- Exemplo:



Representação matricial do MRLM

- Por forma a deduzir algumas propriedades do estimador OLS é substancialmente mais fácil utilizar a notação vectorial e matricial. O modelo inicial,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

poderá ser reescrito como,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

Representação matricial do MRLM (cont.)

- Notação matricial:

$$\underset{n\times 1}{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \underset{n\times 1}{\mathbf{u}} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$\underset{n\times(k+1)}{\mathbf{X}} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \qquad \underset{(k+1)\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

Representação matricial do MRLM (cont.)

- As estimativas para os elementos do vector β são aquelas que minimizam a Soma Quadrática dos Resíduos (SQR):

$$\text{SQR} = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)'$$

- A condição de primeira ordem será dada por:

$$\frac{\partial \text{SQR}}{\partial \hat{\beta}} \equiv \mathbf{0} \Leftrightarrow \mathbf{X}'\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right) \equiv \mathbf{0} \Leftrightarrow$$

$$\Leftrightarrow \mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y} \Leftrightarrow \hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

se $\mathbf{X}'\mathbf{X}$ for não singular.

Observações

- Note-se que $\mathbf{X}$ não é matriz quadrada, logo
$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} \neq \mathbf{X}^{-1}\left(\mathbf{X}'\right)^{-1}$$
- O vector de valores ajustados é dado por:
$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$$
- Pela condição de primeira ordem, podemos verificar que matricialmente os resíduos também não deverão estar correlacionados com os regressores, i.e.
$$\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$$
- Todas as hipóteses do MRLM já vistas podem rapidamente ser generalizadas para o caso matricial. Atente-se apenas à hipótese MLR.3, implicando que característica da matriz $\mathbf{X}$ seja igual a $k+1$

O estimador OLS é centrado

Teorema 1

Sob as hipóteses MLR.1 a MLR.4,

$$E\left(\hat{\beta}\right) = \beta$$

para quaisquer valores do parâmetro β_j, na população. Isto é, o estimador OLS é centrado nos parâmetros populacionais.

- **DEMONSTRAÇÃO:**
  Sob as hipóteses MLR.1 a MLR.3 sabemos que

  $$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\beta + \mathbf{u}\right) = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}$$

  O valor esperado virá

  $$E\left(\hat{\beta} \mid \mathbf{X}\right) = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left(\mathbf{u} \mid \mathbf{X}\right) = \beta$$

  pois sob MLR.4, $E\left(\mathbf{u} \mid \mathbf{X}\right) = \mathbf{0}$

O estimador OLS é centrado (cont.)

- Note-se que o teorema 1 também pode ser visto como:

$$E\left(\hat{\boldsymbol{\beta}}\right) = E\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{bmatrix} = \begin{bmatrix}E\left(\hat{\beta}_0\right)\\ E\left(\hat{\beta}_1\right)\\ E\left(\hat{\beta}_2\right)\\ \vdots\\ E\left(\hat{\beta}_k\right)\end{bmatrix} = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \vdots\\ \beta_k\end{bmatrix} = \boldsymbol{\beta}$$

Variância do estimador OLS

- No contexto matricial vamos utilizar a hipótese MLR.5 alargada:

Hipótese MLR.5 alargada – Homocedasticidade e ausência de autocorrelação

i) $Var\left(u_i \mid x_1, x_2, \ldots, x_k\right) = \sigma^2$

ii) $Cov\left(u_p, u_q \mid x_1, x_2, \ldots, x_k\right) = 0, \forall p \neq q$ (para diferentes momentos de tempo, a correlação entre os erros é nula)

Matricialmente teremos,

$$Var\left(\mathbf{u} \mid \mathbf{X}\right) = \sigma^2 \mathbf{I}_n$$

onde $\mathbf{I}_n$ é a matriz identidade de dimensão $n \times n$

Variância do estimador OLS (cont.)

- Note-se que a hipótese MLR.5 alargada também pode ser vista como:

$$Var(\mathbf{u} \mid \mathbf{X}) = \begin{bmatrix} Var(u_1) & Cov(u_1, u_2) & \dots & Cov(u_1, u_n) \\ Cov(u_2, u_1) & Var(u_2) & \dots & Cov(u_2, u_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(u_n, u_1) & Cov(u_n, u_2) & \dots & Var(u_n) \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Variância do estimador OLS (cont.)

Teorema 2

Sob as hipóteses MLR.1 a MLR.5 alargada,

$$Var\left(\hat{\beta} \mid \mathbf{X}\right) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

- **DEMONSTRAÇÃO:**

$$\hat{\beta} = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{u}$$

A variância virá

$$Var\left(\hat{\beta} \mid \mathbf{X}\right) = Var\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{u} \mid \mathbf{X}\right] =$$

$$= \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' Var\left(\mathbf{u} \mid \mathbf{X}\right) \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

pois sob MLR.5 alargada, $Var\left(\mathbf{u} \mid \mathbf{X}\right) = \sigma^2 \mathbf{I}_n$

Teorema de Gauss-Markov

Teorema de Gauss-Markov

Sob as hipóteses MLR.1 a MLR.5, $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ são os melhores estimadores não enviesados, dentro da classe de todos os estimadores lineares, para os parâmetros $\beta_0, \beta_1, \ldots, \beta_k$. Por outras palavras, o estimador OLS é "**BLUE**", **B**est **L**inear **U**nbiased **E**stimator

- Sob as hipóteses MLR.1 a MLR.4, qualquer estimador linear será centrado
- São os melhores estimadores, uma vez que, dentro da classe dos estimadores lineares e centrados, o estimador OLS é aquele que apresenta a variância mais baixa
- Naturalmente o estimador OLS deixará de ser o melhor, em termos de variância, no caso em que MLR.5 não se verifique

Observações

- Em termos escalares, o teorema 1 resume-se a
$$E\left(\hat{\beta}_j\right) = \beta_j$$

- Quanto ao teorema 2, este virá
$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\mathrm{SQT}_j\left(1 - R_j^2\right)}$$
onde $\mathrm{SQT}_j = \sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2$ e R_j^2 é o R^2 da regressão entre x_j e todas as outras variáveis independentes (incluindo a constante).

- É importante discutir, no contexto do teorema 2, os efeitos do mau ajustamento do modelo aos dados ($\uparrow \sigma^2$) e ainda o efeito da multicolinearidade ($R_j^2 \to 1$)

Estimação da variância do estimador OLS

- Como sabemos, a variância do estimador OLS é dada por:

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\mathrm{SQT}_j\left(1 - R_j^2\right)}$$

- Porém desconhecemos σ^2 uma vez que não observamos o erro aleatório u. Um estimador "natural" seria:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2$$

uma vez que,

$$E\left(\hat{\sigma}^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} u_i^2\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(u_i^2\right) = \sigma^2$$

recordemos que $E\left(u_i\right) = 0$

Estimação da variância do estimador OLS (cont.)

- Como não observamos u mas sim $\hat{u}$, vamos obter o seguinte resultado:

$$E\left(\sum_{i=1}^{n}\hat{u}_i^2 \mid x_1, x_2, \ldots, x_k\right) = (n-k-1)\,\sigma^2$$

- Intuitivamente, este resultado decorre do facto de termos estimado $k+1$ coeficientes, $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$

- Assim sendo, um "candidato" a estimador de σ^2, utilizando $\hat{u}$ em vez de u, será

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2$$

Estimação da variância do estimador OLS (cont.)

- Resta-nos referir um último teorema, que nos garante o não enviesamento de $\hat{\sigma}^2$:

Teorema 3

Sob as hipóteses MLR.1 a MLR.5, $E\left(\hat{\sigma}^2\right) = \sigma^2$

- A partir deste momento estamos em condições de estimar a variância do estimador OLS, substituindo σ^2 pela sua estimativa $\hat{\sigma}^2$:

$$\widehat{Var\left(\hat{\beta}_j\right)} = \frac{\hat{\sigma}^2}{\mathrm{SQT}_j\left(1 - R_j^2\right)}$$