iscte BUSINESS SCHOOL

Econometria I

Aula 27

Ricardo Gouveia-Mendes
ricardo.mendes@iscte-iul.pt

Licenciatura em Economia
2.º Semestre 2023-24

Efeitos da Má Especificação do Modelo

Estimadores OLS e Má Especificação

- Como pode um **modelo estar mal especificado?** 🤔
  - *Incluir* variáveis explicativas *irrelevantes*
  - *Excluir* variáveis explicativas *relevantes*
- A má especificação de um modelo pode ter **impacto nas propriedades dos estimadores OLS**
  - Centricidade
  - Precisão
- Mesmo que todos os pressupostos do método estejam cumpridos

Estimadores OLS e Má Especificação
Exemplo de Inclusão de Variável Irrelevante

- Suponhamos que estamos interessados no impacto da variável X_1 sobre a variável Y e estamos na dúvida entre estes dois modelos:

$$
\begin{aligned}
Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, && \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2 \\
Y &= \beta_0 + \beta_1 X_1 + v, && \tilde{\beta}_0, \tilde{\beta}_1
\end{aligned}
$$

- O que acontece **se X_2 for, de facto, irrelevante ($\beta_2 = 0$)?** 🤔
  - Se $\mathbb{E}[u|X_1, X_2] = 0$ e $\mathbb{E}[v|X_1] = 0$ *tanto $\hat{\beta}_1$ como $\tilde{\beta}_1$ são centrados*
  - Mas, quanto à variância, o *estimador OLS perde precisão*:
    - Se $\mathbf{Cov}(X_1, X_2) = 0$ então $\sigma^2_{\hat{\beta}_1} = \sigma^2_{\tilde{\beta}_1}$
    - Se $\mathbf{Cov}(X_1, X_2) \neq 0$ então $\sigma^2_{\hat{\beta}_1} > \sigma^2_{\tilde{\beta}_1}$

Estimadores OLS e Má Especificação
Exemplo de Exclusão de Variável Relevante

- Suponhamos, novamente, que estamos indecisos entre os mesmos dois modelos anteriores:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, \qquad && \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2 \\ Y &= \beta_0 + \beta_1 X_1 + v, && \tilde{\beta}_0, \tilde{\beta}_1 \end{aligned}$$

- O que acontece **se X_2 for, de facto, relevante ($\beta_2 \neq 0$)?** 🤔
  - Se $\mathbb{E}[u|X_1, X_2] = 0$ então $\hat{\beta}_1$ *é centrado*
  - Se $\mathbf{Cov}(X_1, X_2) = 0$ então $\mathbb{E}[v|X_1] = 0$ e $\tilde{\beta}_1$ *também é centrado*
  - Se $\mathbf{Cov}(X_1, X_2) \neq 0$ então $\mathbb{E}[v|X_1] \neq 0$ e $\tilde{\beta}_1$ *é enviesado*

Estimadores OLS e Má Especificação

Resumo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, \qquad \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$$

$$Y = \beta_0 + \beta_1 X_1 + v, \qquad \tilde{\beta}_0, \tilde{\beta}_1$$

- É possível demonstrar:

$$\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2 \tilde{\delta}_1$$

em que $\tilde{\delta}_1$ resulta da estimação de:

$$X_2 = \delta_0 + \delta_1 X_1 + w$$

- Portanto, $\mathbb{E}[\tilde{\beta}_1] = \beta_1$ apenas se:
  - $\beta_2 = 0$ (X_2 é irrelevante)
  - $\tilde{\delta}_1 = 0$ (X_2 é relevante, mas independente de X_1)

Exercícios

Exercício 6.1

A equação seguinte descreve o preço mediano da habitação em função do montante de poluição ($\boldsymbol{nox}$, óxido de nitrogénio) e do número médio de quartos nas casas de uma comunidade ($\boldsymbol{quartos}$):

$$\ln(preco) = \beta_0 + \beta_1 \ln(nox) + \beta_2 quartos + u$$

a. Quais os sinais prováveis de β_1 e β_2? Qual a interpretação de β_1? Explique.

- $\hat{\beta}_1 < 0$: a poluição desvaloriza as casas
- $\hat{\beta}_2 > 0$: quanto maior a casa, maior o preço
- $\hat{\beta}_1$ é a uma estimativa de um **elasticidade** — qual a variação percental do preço resultante de um aumento de 1% do nível de poluição

Exercício 6.1

b. Porque é que nox (mais precisamente, o seu logaritmo) e $quartos$ poderão estar negativamente correlacionadas? Se assim for, a regressão simples de $\ln(preco)$ sobre $\ln(nox)$ produzirá um estimador de β_1 enviesado para cima ou para baixo?

- Poluição → pobreza → casas mais pequenas
- A regressão simples referida, produziria:

$$\ln(preco) = \tilde{\beta}_0 + \tilde{\beta}_1 \ln(nox)$$

sabendo que $\mathbb{E}(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1$

- $\tilde{\delta}_1$ vem de:

$$\ln(nox) = \tilde{\delta}_0 + \tilde{\delta}_1 quartos$$

- Se $\beta_2 > 0$ e $\tilde{\delta}_1 < 0$ então $\mathbb{E}(\tilde{\beta}_1) < \beta_1$
- Enviesamento para baixo poderá sobrestimar efeito da poluição

Exercício 6.1

c. Após recolher os dados necessários foram estimadas as seguintes equações:

$$\widehat{\ln(preco)} = 11.71 - 1.043\ln(nox), \qquad N = 506, \quad R^2 = 0.264,$$
$$\widehat{\ln(preco)} = 9.23 - 0.718\ln(nox) + 0.306 quartos, \qquad N = 506, \quad R^2 = 0.514.$$

A relação entre as estimativas da elasticidade do preço em relação a $\boldsymbol{nox}$ na regressão simples e na regressão múltipla é aquela que teria previsto, dada a sua resposta em b)? Isto quer dizer que (-0.718) está certamente mais próxima da verdadeira elasticidade que (–1.043)?

- $\tilde{\beta}_1 = -1.043 < -0.718 = \hat{\beta}_1$ tal como esperado
- Portanto, $\hat{\beta}_1 = -0.718$ está mais próximo da verdadeira elasticidade

Exercício 6.2

Para explicar o salário dos executivos (sal) usou-se o modelo

$$\ln(sal) = \beta_0 + \beta_1 \ln(vendas) + \beta_2 valmerc + \beta_3 mglucro + \beta_4 ase + \beta_5 ast + u$$

onde $valmerc$ é o valor de mercado da empresa, $mglucro$ é o lucro como percentagem das vendas, ase os anos de serviço como executivo na empresa atual e ast os anos de serviço totais na empresa atual. A estimação deste modelo pelo método dos mínimos quadrados, com $N = 177$, forneceu um $R^2 = 0.353$. Quando ase^2 e ast^2 são adicionadas, obtém-se $R^2 = 0.375$. Há evidência de má especificação da forma funcional do modelo apresentado?

Exercício 6.2

$$\ln(sal) = \beta_0 + \beta_1 \ln(vendas) + \beta_2 valmerc + \beta_3 mglucro + \\ + \beta_4 ase + \beta_5 ast + \beta_6 ase^2 + \beta_7 ast^2 + u$$

$$H_0 : \beta_6 = \beta_7 = 0$$
$$H_1 : \text{Não } H_0$$

$$F = \frac{R^2 - R_*^2}{1 - R^2} \frac{N - p}{q} = \\ = 2.9744 < 3.0495 = F_{169}^2$$

$$p\text{-value} = 0.0538 > 0.05 = \alpha$$

- **Não rejeitamos** a hipótese nula, mas por muito pouco
- Indicativo de que há alguma evidência, embora fraca, de má especificação do modelo
- O efeito de $\boldsymbol{ase}$ e $\boldsymbol{ast}$ não deverá ser linear