

Econometria I

Aula 28

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

Licenciatura em Economia
2.º Semestre 2023-24

Teste para a Forma Funcional e Critérios de Seleção de Modelos

O Problema da Forma Funcional

- Vimos que o que define uma regressão linear é a **linearidade nos parâmetros**
- As **variáveis explicativas** podem, contudo, aparecer com **formas funcionais não lineares** (logarítmica, quadrática)
- Será possível saber se a forma funcional escolhida para no modelo é correta?

- Sim. Através do **teste RESET**

O Teste RESET

Princípios

- A **verdadeira forma funcional** $S(\cdot)$ de qualquer modelo econométrico é **desconhecida**

$$Y = S(X\beta) + u \Rightarrow \mathbb{E}[Y|X] = S(X\hat{\beta})$$

- Qualquer função matemática se pode aproximar por uma **expansão em série de Taylor**

$$\mathbb{E}[Y|X] = X\hat{\beta} + \sum_{i=1}^{\infty} \theta_j (X\hat{\beta})^{j+1}$$

O Teste RESET

Princípios

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u$$

- A forma mais comum do teste RESET acrescenta **dois polinômios** (2.^a e 3.^a ordem) ao modelo:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \theta_1 \hat{Y}^2 + \theta_2 \hat{Y}^3 + v$$

- Se $\theta_1 = \theta_2 = 0$ então $S(\mathbf{X}\hat{\beta}) = \mathbf{X}\hat{\beta}$

O Teste RESET

Procedimento

1. Estimar o modelo original: $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + u$
2. Obter as variáveis: \hat{Y} , \hat{Y}^2 e \hat{Y}^3
3. Estimar o modelo auxiliar: $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \theta_1 \hat{Y}^2 + \theta_2 \hat{Y}^3 + v$
4. Realizar um **teste F ou LM** para comparar os dois modelos:

$$H_0 : \theta_1 = \theta_2 = 0 \quad H_1 : \text{Não } H_0$$

5. Se **não rejeitamos** H_0 o modelo original está **bem especificado**

Critérios para a Seleção de Modelos

- Por vezes, podem existir **vários modelos que parecem adequados**.
Como escolher? 🤔
- Quando um dos modelos é um caso particular do outro (modelos encaixados) aplica-se um **teste F ou LM**
- Se tivermos modelos não encaixados, mas com o mesmo número de variáveis explicativas podemos **usar o R^2** para decidir qual é melhor (o que tiver R^2 maior)
- O problema são os modelos não encaixados com diferente número de regressores 😞

Critérios para a Seleção de Modelos

O R^2 Ajustado

- Para o caso de modelos não encaixados com diferentes números de regressores usa-se o **R^2 Ajustado**

$$\bar{R}^2 = 1 - (1 - R^2) \frac{N - 1}{N - p} = 1 - \frac{SQR / (N - k)}{SQT / (N - 1)}$$

- **Não tem qualquer interpretação útil** (pode ser negativo)
- O **melhor modelo** será o que apresentar um \bar{R}^2 maior
- Continuamos a **só poder comparar modelos** com a **mesma variável dependente** (incluindo a forma funcional com que aparece no modelo)

Exercícios

Exercício 6.3

Considere os dados de `PRECASA.DTA` e os modelos

$$(1) \quad \textit{preco} = \beta_0 + \beta_1 \textit{lote} + \beta_2 \textit{area} + \beta_3 \textit{quartos} + u$$

$$(2) \quad \textit{preco} = \beta_0 + \beta_1 \textit{lote} + v$$

$$(3) \quad \textit{preco} = \beta_0 + \beta_1 \ln(\textit{lote}) + \beta_2 \ln(\textit{area}) + \beta_3 \textit{quartos} + w$$

a. Através de testes e critérios apropriados, verifique qual dos modelos é mais adequado para explicar a variação no preço da habitação.

- Modelos (1) e (2): comparáveis por meio de um teste estatístico (F ou LM)
- Modelos (1) e (3): comparáveis através do R^2 (mesmo número de variáveis)
- Modelo (2) e (3): comparáveis através do R^2 ajustado (não encaixados)

Exercício 6.3

Alínea a) | 1.º passo: estimar os três modelos

Estimar modelo (1): `regress preco lote area quartos`

Call:

```
lm(formula = preco ~ lote + area + quartos, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-120.457	-38.372	-6.185	32.223	208.807

Exercício 6.3

Alínea a) | 1.º passo: estimar os três modelos

Estimar modelo (2): `regress preco lote`

Call:

```
lm(formula = preco ~ lote, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-268.76	-51.76	-22.80	35.73	354.41

Exercício 6.3

Alínea a) | 1.º passo: estimar os três modelos

- Criar variáveis logarítmicas: `gen llote = log(lote)` e `gen larea = log(area)`
- Estimar modelo (3): `regress preco llote larea quartos`

Call:

```
lm(formula = preco ~ log(lote) + log(area) + quartos, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-109.270	-38.209	-4.924	23.890	217.590

Exercício 6.3

Alínea a) | 2.º passo: testar modelos

Modelo (1) vs modelo (2)

$$H_0 : \beta_2 = \beta_3 = 0$$

$$H_1 : \text{Não } H_0$$

$$F = \frac{R^2 - R_*^2}{1 - R^2} \frac{N - p}{q} =$$
$$= 70.6988 > 3.1052 = F_{84}^2$$

$$p\text{-value} = 0 < 0.05 = \alpha$$

Rejeitamos H_0 , pelo que o modelo (1) é preferível ao modelo (2)

Modelo (1) vs modelo (3)

$$R_{(1)}^2 = 0.6722 < 0.6782 = R_{(3)}^2$$

O modelo (3) é preferível ao modelo (1)

Modelo (2) vs modelo (3)

$$\overline{R}_{(2)}^2 = 0.1103 < 0.6667 = \overline{R}_{(3)}^2$$

O modelo (3) é preferível ao modelo (2)

Exercício 6.3

Alínea b)

- b.** Verifique se a forma funcional do modelo que selecionou na Alínea anterior foi especificada corretamente, usando o teste RESET (versão F).

1.º passo: estimar regressão auxiliar

- Obter vetor de estimativas para a variável dependente: `predict yhat`
- Criar variáveis auxiliares: `gen yhat2 = yhat^2` e `gen yhat3 = yhat^3`
- Estimar modelo auxiliar: `regress preco llote larea quartos yhat2 yhat3`

Exercício 6.3

Alínea b) | 1.º passo: estimar regressão auxiliar

Call:

```
lm(formula = preco ~ log(lote) + log(area) + quartos + I(yhat_3^2) +  
    I(yhat_3^3), data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-121.49	-30.62	-6.50	26.29	192.50

Exercício 6.3

Alínea b) | 2.º passo: realizar o teste F

$$H_0 : \beta_4 = \beta_5 = 0$$

$$H_1 : \text{Não } H_0$$

$$\begin{aligned} F &= \frac{R^2 - R_*^2}{1 - R^2} \frac{N - p}{q} = \\ &= 15.9466 > 3.1079 = F_{82}^2 \end{aligned}$$

$$p\text{-value} = 0 < 0.05 = \alpha$$

Rejeitamos H_0 , pelo que a forma funcional do modelo (3) não é adequada

Exercício 6.3

Alínea c)

c. Repita a alínea anterior mas usando a versão LM do teste RESET.

1.º passo: estimar regressão auxiliar

- Obter vetor de estimativas para os resíduos do modelo (3): `predict uhat, resid`
- Estimar modelo auxiliar: `regress uhat llote larea quartos yhat2 yhat3`

Exercício 6.3

Alínea c) | 1.º passo: estimar regressão auxiliar

Call:

```
lm(formula = uhat_3 ~ log(lote) + log(area) + quartos + I(yhat_3^2) +  
    I(yhat_3^3), data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-121.49	-30.62	-6.50	26.29	192.50

Exercício 6.3

Alínea c) | 2.º passo: realizar o teste LM

H_0 : Forma funcional correta

H_1 : Forma funcional incorreta

$$\begin{aligned} LM &= N \times R^2 = \\ &= 24.6423 > 5.9915 = \chi_2^2 \end{aligned}$$

$$p\text{-value} = 0 < 0.05 = \alpha$$

Rejeitamos H_0 , pelo que confirmamos que a forma funcional do modelo (3) não é adequada