TÉCNICAS APLICADAS DE INVESTIGAÇÃO EM GESTÃO

2. Regressão linear múltipla – Análise da especificação

- Variáveis em logaritmo
- Termos quadráticos e de interação
- Variáveis Dummy
- Testes de especificação
  - RESET e J
  - Heterocedasticidade

Variáveis em logaritmo

- **Forma semi-logarítmica**

(A) $\log(y) = \beta_0 + \beta_1 x + u$

- Efeito de uma variação unitária de x

Variação relativa de y

$$\beta_1 = \frac{d\log(y)}{dx} = \frac{1}{y}\frac{dy}{dx} = \frac{\frac{dy}{y}}{dx}$$

- Em geral:

$$\Delta x = 1 \Rightarrow \%\Delta y = \beta_1 \times 100\%$$
$$\Delta x \Rightarrow \%\Delta y = \beta_1 \Delta x \times 100\%$$

- O coeficiente β_1 é uma **semi-elasticidade**

(B) $y = \beta_0 + \beta_1 \log(x) + u$

$$\beta_1 = \frac{dy}{d\log(x)} = \frac{dy}{dx/x}$$

$$\boxed{\%\Delta x = 1 \ \ (\Delta\log(x) = 0.01) \Rightarrow \Delta y = \beta_1 \times 0.01}$$

Variação relativa de x

Variáveis em logaritmo

- **Forma log-logarítmica: modelo loglinear**

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

$$\beta_1 = \frac{d\log(y)}{d\log(x)} = \frac{dy/y}{dx/x}$$

Variação percentual de y se x aumentar 1% → **Elasticidade**

- O coeficiente β_1 é uma **elasticidade**

- Mdelo de **elasticidades constantes**

Termos quadráticos e interações

- **Termos quadráticos**

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

- Efeito de uma variação de x

$$\Delta y \approx \left(\frac{dy}{dx}\right)\Delta x \Leftrightarrow \Delta y \approx \Delta x\left(\beta_1 + 2\beta_2 x\right)$$

O efeito depende de *x*

  - A interpretação de β_1 não é a habitual:

$$x = 0, \Delta x = 1 \Rightarrow \Delta y \approx \beta_1 + 2\beta_2 \times 0 \Leftrightarrow \beta_1 \approx \Delta y \text{ quando } x = 0, \Delta x = 1$$

  - O efeito não é constante e depende de x

$$x = 3, \Delta x = 1 \Rightarrow \Delta y \approx \beta_1 + 2\beta_2 \times 3 = \beta_1 + 6\beta_2$$

$$x = 4, \Delta x = 1 \Rightarrow \Delta y \approx \beta_1 + 2\beta_2 \times 4 = \beta_1 + 8\beta_2$$

  - Ponto de inflexão do efeito parcial: valor de x que maximiza (minimiza) y

$$x^* = \frac{-\beta_1}{2\beta_2}$$

Termos quadráticos e interações

- **Termos de interação**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

- Efeito de uma variação de x_1

$$\Delta x_1 \Rightarrow \Delta y = \Delta x_1 (\beta_1 + \beta_3 x_2)$$

O efeito depende de x_2

- A interpretação de β_1 não é a habitual:

$$x_2 = 0, \Delta x_1 = 1 \Rightarrow \Delta y \approx \beta_1 + \beta_3 \times 0 \Leftrightarrow \beta_1 \approx \Delta y \text{ quando } x_2 = 0, \Delta x_1 = 1$$

- Reparametrização dos efeitos de interação

$$y = \beta_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u$$

Média de x_1 Média de x_2

Podem ser substituídas pelas médias observadas na amostra

- interpretação de δ_1

Efeito de uma variação unitária de x_1 *(ceteris paribus)* quando x_1 e x_2 são iguais às respetivas médias.

Variáveis Dummy

- **Informação qualitativa**
  - Exemplos:
    - Sexo de uma pessoa
    - Região onde vive o agregado familiar
    - Sector onde trabalha a empresa
    - Nível de satisfação
    - Classe de rendimento
  - Como introduzir no modelo de regressão

variáveis Dummy

Variáveis que assumem apenas o valor 0 ou 1

Variáveis Dummy

- **Apenas uma variável Dummy nas variáveis explicativas**

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + u$$

- Assumindo a hipótese DS.4, $E(u \mid female, educ) = 0$ então:

Regressão para as mulheres

$$E(sal \mid mulher = 1, educ) = \beta_0 + \delta_0 + \beta_1 educ$$

Regressão para os homens

$$E(sal \mid mulher = 0, educ) = \beta_0 + \beta_1 educ$$

- Interpretação do coeficiente da variável Dummy

$$\boxed{\delta_0 = E(sal \mid mulher = 1, educ) - E(sal \mid mulher = 0, educ)}$$

Diferença média ou esperada entre o salário de uma mulher e o salário de um homem com o mesmo nível de escolaridade (mantendo tudo o resto constante)

Variáveis Dummy

- **Dummy Trap**

$$sal = \beta_0 + \delta_0 mulher + \gamma_0 homem + \beta_1 educ + u$$

Existe multicolienaridade perfeita entre as variáveis Dummy e o termo independente

Não é possível estimar todos os coeficientes do modelo

- **O que fazer?**

- Só introduzir uma variável Dummy na equação

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + u$$

ou

$$sal = \beta_0 + \gamma_0 homem + \beta_1 educ + u$$

- Retirar o termo independente

Atenção: a interpretação dos coeficientes das Dummies é diferente

$$sal = \delta_0^* mulher + \gamma_0^* homem + \beta_1 educ + u$$

Inconveniente:
- Mais difícil de testar e avaliar diretamente a diferença entre os dois grupos
- A fórmula habitual do R^2 não é válida.

Variáveis Dummy

- **Mudar o grupo base**

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + u$$

$$\Downarrow$$

$$sal = \beta_0 + \gamma_0 homem + \beta_1 educ + u$$

então $\boxed{\gamma_0 = -\delta_0}$

- **Exemplo**

Dependent Variable: WAGE ($/hora)
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.319204	0.738825	-3.139042	0.0018
FEMALE	-2.114035	0.262550	-8.051927	0.0000
EDUC	0.556285	0.050287	11.06210	0.0000
EXPER	0.255128	0.034867	7.317144	0.0000
EXPER^2	-0.004440	0.000776	-5.719805	0.0000

Variáveis Dummy

- **Variável Dummy para explicar $\log(y)$**

$$\log(y) = \beta_0 + \delta_0 Dummy + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- **Variação exacta** $\%\Delta y = \left[\exp\left(\delta_0\right) - 1\right] \times 100\%$

- **Exemplo**

Dependent Variable: LOG(WAGE)
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.390483	0.102210	3.820413	0.0001
FEMALE	-0.337187	0.036321	-9.283424	0.0000
EDUC	0.084136	0.006957	12.09407	0.0000
EXPER	0.038910	0.004824	8.066683	0.0000
EXPER^2	-0.000686	0.000107	-6.388842	0.0000

Variáveis Dummy

- **Variáveis Dummy: v. qualitativas <u>com várias categorias</u>**

Introduzir uma variável dummy por categoria, deixando uma de fora que será o grupo base para leitura dos resultados

Dependent Variable: WAGE
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.655184	0.754921	-3.517167	0.0005
MARRFEM	-1.051078	0.427586	-2.458167	0.0143
MARRMALE	2.079997	0.407073	5.109641	0.0000
SINGFEM	-0.423726	0.424226	-0.998821	0.3183
EDUC	0.578108	0.050528	11.44132	0.0000
EXPER	0.053222	0.010823	4.917597	0.0000

Dependent Variable: WAGE
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic
C	-3.078910	0.725669	-4.242859
MARRFEM	-0.627352	0.388914	-1.613088
MARRMALE	2.503724	0.366868	6.824585
SINGMALE	0.423726	0.424226	0.998821
EDUC	0.578108	0.050528	11.44132
EXPER	0.053222	0.010823	4.917597

<u>Mudança do grupo base</u>

Variáveis Dummy

- **Variáveis Dummy: v. qualitativas <u>ordinais</u>**

- As variáveis qualitativas ordinais devem ser transformadas em variáveis dummy para conferir mais flexibilidade à estimação

$$escaleduc = \begin{cases} 1 \text{ se tiver completado 4 anos de escolaridade} \\ 2 \text{ se tiver completado 9 anos de escolaridade} \\ 3 \text{ se tiver completado 12 anos de escolaridade} \\ 4 \text{ se tiver completado 15 anos ou mais de escolaridade} \end{cases}$$

- Com variável ordinal

$$sal = \beta_0 + \beta_1 escaleduc + \text{outros fatores}$$

- Com Dummies

$$sal = \beta_0 + \delta_1 educ_1 + \delta_2 educ_2 + \delta_3 educ_3 + \text{outros fatores}$$

$$educ_j = \begin{cases} 1 \text{ se } escaleduc = j \\ 0 \text{ caso contrário} \end{cases}$$

Variáveis Dummy

- **Interações envolvendo variáveis Dummy**

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + \delta_1 mulher \times educ + u$$

- Assumindo a hipótese DS.4, $E(u \mid female, educ) = 0$ então:

Regressão para as mulheres

$$E(sal \mid mulher = 1, educ) = \beta_0 + \delta_0 + (\beta_1 + \delta_1) educ$$

Regressão para os homens

$$E(sal \mid mulher = 0, educ) = \beta_0 + \beta_1 educ$$

- Interpretação do coeficiente da variável Dummy

$$\boxed{\delta_0 = E(sal \mid mulher = 1, educ = 0) - E(sal \mid mulher = 0, educ = 0)}$$

Diferença média ou esperada entre o salário de uma mulher e o salário de um homem com nível de escolaridade=0 (mantendo tudo o resto constante)

$\boxed{\delta_1 = \text{ diferença no retorno do salário de uma mulher face a um homem}}$

Variáveis Dummy

- **Testar diferenças na forma funcional entre grupos da população**

Variável dummy: $D = \begin{cases} 1 \text{ se a observação pertence ao grupo 1} \\ 0 \text{ caso contrário} \end{cases}$

- Estimar o modelo pelo OLS

$$y = \beta_0 + \delta_0 D + \beta_1 x_1 + \delta_1 x_1 \times D + ... + \beta_k x_k + \delta_k x_k \times D + u$$

- Testar

$$H_0 : \delta_0 = \delta_1 = ... = \delta_k = 0 \qquad H_1 : H_0 \text{ falsa}$$

Teste F → Rejeitando H_0 há evidência de que a forma funcional é diferente nos dois grupos → Fazer uma regressão separada para cada grupo

- Por vezes tem interesse testar apena diferenças nos coeficientes de declive:

$$H_0 : \delta_1 = ... = \delta_k = 0 \qquad H_1 : H_0 \text{ falsa}$$

Testes da forma funcional

- **Regression specification error test – RESET**
  - Intuição:

  Se o modelo estiver bem especificado não tem variáveis omitidas relevantes, a forma funcional é correta;

  então polinómios das variáveis explicativas não podem ser relevantes para explicar a variável dependente.

  - Procedimento de teste:

1. OLS do modelo original: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + u \longrightarrow \hat{y}$
2. Estimação OLS da equação de teste:
   i. Variante 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + \gamma \hat{y}^2 + v_1$

      Testar $\gamma = 0$ contra $\gamma \neq 0$

   ii. Variante 2: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + \gamma_1 \hat{y}^2 + \gamma_2 \hat{y}^3 + v_2$

      Testar $\gamma_1 = 0, \gamma_2 = 0$ conjuntamente
3. Conclusão:

   Se se rejeitar H_0 então há evidência de que o modelo está mal especificado e necessita ser reformulado.

Testes da forma funcional

- **Teste J**
  - Serve para avaliar a forma funcional de modelos não encaixados para a mesma v. dependente $\Rightarrow$ permite escolher entre modelos não encaixados
  - Procedimento de teste:

1. Estimar pelo OLS cada um dos modelos

OLS do modelo A: $y = \beta_0 + \beta_1 x + \beta_2 w + u \longrightarrow \hat{y}_A$

OLS do modelo B: $y = \alpha_0 + \alpha_1 x + \alpha_2 z + v \longrightarrow \hat{y}_B$

2. Equações de teste:

OLS de $y = \beta_0 + \beta_1 x + \beta_2 w + \gamma_B \hat{y}_B + u_2 \longrightarrow$ testar $H_0 : \gamma_B = 0 \quad H_1 : \gamma_B \neq 0$

OLS de $y = \alpha_0 + \alpha_1 x + \alpha_2 w + \gamma_A \hat{y}_A + v_2 \longrightarrow$ testar $H_0 : \gamma_A = 0 \quad H_1 : \gamma_A \neq 0$

Testes da forma funcional

- Discussão dos resultados

  i. $\gamma_A \neq 0$ e $\gamma_B = 0 \longrightarrow$ **Modelo A** (modelo B mal especificado)

  A especificação de A acrescenta informação ao modelo B mas a especificação de B não acrescenta informação a A

  ii. $\gamma_A = 0$ e $\gamma_B \neq 0 \longrightarrow$ **Modelo B** (modelo A mal especificado)

  A especificação de B acrescenta informação ao modelo A mas a especificação de A não acrescenta informação a B

  iii. $\gamma_A = 0$ e $\gamma_B = 0 \longrightarrow$ **Modelo A ou Modelo B**

  A especificação de B não acrescenta informação ao modelo A e a especificação de A não acrescenta informação a B

  iv. $\gamma_A \neq 0$ e $\gamma_B \neq 0 \longrightarrow$ **Rejeitar o Modelo A e o Modelo B**

  A especificação de B acrescenta informação ao modelo A e a especificação de A acrescenta informação a B

HETEROCEDASTICIDADE: Motivação e consequências

- **Motivação**

- Hipótese DS.5: Homocedasticidade

$$Var(u_i \mid x_{i1}, x_{i2}, \ldots, x_{ik}) = \sigma^2$$

O valor das variáveis explicativas **não pode** conter qualquer informação sobre a variância do erro

A variância do erro **não pode** depender do valor das variáveis explicativas

- Com dados seccionais esta hipótese dificilmente se verifica
o mais comum é a própria variância do erro (ou da variável dependente) ser função das variáveis explicativas.

Exemplo:

$$cerveja = \beta_0 + \beta_1 rendimento + u$$

$$Var(cerveja \mid rendimento) = Var(u \mid rendimento) = h(rendimento) \neq \sigma^2$$

- **Definição:**

$$Var(u_i \mid x_{i1}, x_{i2}, \ldots, x_{ik}) = h(x_{i1}, x_{i2}, \ldots, x_{ik}) = \sigma_i^2 \neq \sigma^2$$

A variância do erro **não é constante!**

A variância do erro e depende do valor das variáveis explicativas

HETEROCEDASTICIDADE: Motivação e consequências

- **Consequências na estimação OLS**

- A heterocedasticidade não põe em causa as hipóteses DS.1 a DS.4, como tal, desde que estas se verifiquem então,

  - O estimador OLS para $\beta_j \;\; j = 0,1,\ldots,k$
    - É centrado
    - Não é eficiente

  - O estimador OLS para $Var(\hat{\beta}_j \mid x_1,\ldots,x_k) \;\; j = 0,1,\ldots,k$
    - É inconsistente!

      ↓

    toda a inferência habitual (I.C., testes *t* e *F*) é inválida

  - O estimador $\hat{\sigma}^2$ não é válido (não faz sentido)

  - O R^2 mantém a interpretação habitual

HETEROCEDASTICIDADE: Estimação consistente da matriz de covariâncias

- **Erros padrão de White (robustos à heterocedasticidade)**

$$se^*\left(\hat{\beta}_j\right) = \frac{\sqrt{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}}{SSR_j}$$

- $\hat{r}_{ij}$ é o resíduo da observação i na regressão de x_j sobre todas as outras variáveis do modelo
- $\hat{u}_i$ é o resíduo da observação i na regressão original
- SSR_j é a soma dos quadrados dos resíduos da regressão original
- O erro padrão de White é consistente para heterocedasticidade de forma desconhecida
- A inferência com $se^*\left(\hat{\beta}_j\right)$ é válida assintoticamente
- O cálculo da estatística F baseado na expressão com a soma dos quadrados dos resíduos ou com os R^2 não é válida (mesmo com erros padrão de White)

HETEROCEDASTICIDADE: Estimação consistente da matriz de covariâncias

Exemplo

Dependent Variable: LOG(WAGE)
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.390483	0.102210	3.820413	0.0001
FEMALE	-0.337187	0.036321	-9.283424	0.0000
EDUC	0.084136	0.006957	12.09407	0.0000
EXPER	0.038910	0.004824	8.066683	0.0000
EXPER^2	-0.000686	0.000107	-6.388842	0.0000

R-squared	0.399590	Mean dependent var	1.623268
Adjusted R-squared	0.394981	S.D. dependent var	0.531538
S.E. of regression	0.413446	Akaike info criterion	1.080882
Sum squared resid	89.05862	Schwarz criterion	1.121427
Log likelihood	-279.2720	Hannan-Quinn criter.	1.096757
F-statistic	86.68521	Durbin-Watson stat	1.775544
Prob(F-statistic)	0.000000		

Dependent Variable: LOG(WAGE)
Method: Least Squares
Included observations: 526
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.390483	0.108598	3.595658	0.0004
FEMALE	-0.337187	0.036184	-9.318715	0.0000
EDUC	0.084136	0.007690	10.94104	0.0000
EXPER	0.038910	0.004675	8.322568	0.0000
EXPER^2	-0.000686	0.000100	-6.828754	0.0000

R-squared	0.399590	Mean dependent var	1.623268
Adjusted R-squared	0.394981	S.D. dependent var	0.531538
S.E. of regression	0.413446	Akaike info criterion	1.080882
Sum squared resid	89.05862	Schwarz criterion	1.121427
Log likelihood	-279.2720	Hannan-Quinn criter.	1.096757
F-statistic	86.68521	Durbin-Watson stat	1.775544
Prob(F-statistic)	0.000000	Wald F-statistic	81.96798
Prob(Wald F-statistic)	0.000000		

HETEROCEDASTICIDADE: Testes

- **Motivação**

$$Var(u \mid x_1,\ldots,x_k) = E(u^2 \mid x_1,\ldots,x_k) = h(x_1, x_2,\ldots,x_k)$$

$$\Downarrow$$

$$E(u^2 \mid x_1,\ldots,x_k) = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k$$

$$\Downarrow$$

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + \varepsilon$$

- **Teste de Breusch-Pagan**

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

- Objetivo:

$$H_0 : Var(u \mid x_1,\ldots,x_k) = \sigma^2 \qquad H_1 : Var(u \mid x_1,\ldots,x_k) = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k$$

HETEROCEDASTICIDADE: Testes

- **Procedimento de teste**

1. Estimação OLS do modelo original ➡ $\hat{u} = y - \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$
2. Regressão auxiliar de teste:

   OLS de $\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + v$
3. Testar

   $H_0 : \delta_1 = \delta_2 = \ldots = \delta_k = 0 \quad H_1 : H_0$ falsa

   Teste *F* de significância global da regressão auxiliar de teste
4. Conclusão

   Rejeitar H_0 ➡ evidência de heterocedasticidade no erro do modelo inicial

   Não rejeitar H_0 ➡ não há evidência de heterocedasticidade.

HETEROCEDASTICIDADE: Testes

- **Exemplo**

$\hat{u}^2$ Com $\hat{u}$ o resíduo da regressão para log(wage)

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.037480	0.068360	0.548268	0.5837
FEMALE	-0.013118	0.024292	-0.540000	0.5894
EDUC	0.005509	0.004653	1.183964	0.2370
EXPER	0.008761	0.003226	2.715742	0.0068
EXPER^2	-0.000169	7.18E-05	-2.358137	0.0187

R-squared	0.018981	Mean dependent var	0.169313
Adjusted R-squared	0.011449	S.D. dependent var	0.278118
S.E. of regression	0.276521	Akaike info criterion	0.276402
Sum squared resid	39.83772	Schwarz criterion	0.316946
Log likelihood	-67.69362	Hannan-Quinn criter.	0.292277
F-statistic	2.520076	Durbin-Watson stat	1.967219
Prob(F-statistic)	0.040375		

Estatística de teste

Evidência de heterocedasticidade

5.3. Testes para deteção de heterocedasticidade

- **Teste de White**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- Teste de hipóteses

$$H_0 : Var(u \mid x_1, x_2, x_3) = \sigma^2 \qquad H_1 : Var(u \mid x_1, x_2, x_3) = h(x_1, x_2, x_3)$$

- **Procedimento de teste**

1. Estimação OLS do modelo original $\Rightarrow \hat{u} = y - \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$
2. Regressão auxiliar de teste: OLS de
$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 + \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + v$$
3. Testar
$$H_0 : \delta_1 = \delta_2 = \ldots = \delta_9 = 0 \qquad H_1 : H_0 \text{ falsa}$$
Teste *F* de significância global da regressão auxiliar de teste
4. Conclusão
Rejeitar H_0 $\Rightarrow$ evidência de heterocedasticidade no erro do modelo inicial

HETEROCEDASTICIDADE: Testes

- **Exemplo**

Test Equation:
Dependent Variable: RESID^2 → $\hat{u}^2$ Com $\hat{u}$ o resíduo da regressão para log(wage)
Method: Least Squares
Included observations: 526
Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.558137	0.268879	2.075795	0.0384
FEMALE	-0.125279	0.140393	-0.892343	0.3726
FEMALE*EDUC	0.012647	0.010043	1.259261	0.2085
FEMALE*EXPER	-0.003751	0.006530	-0.574430	0.5659
FEMALE*EXPER^2	4.52E-05	0.000146	0.310811	0.7561
EDUC^2	0.002544	0.001125	2.260940	0.0242
EDUC*EXPER	-2.69E-05	0.001256	-0.021411	0.9829
EDUC*EXPER^2	1.40E-05	2.73E-05	0.511937	0.6089
EDUC	-0.066880	0.033663	-1.986760	0.0475
EXPER^2	0.000813	0.001351	0.601774	0.5476
EXPER*EXPER^2	-2.99E-05	4.44E-05	-0.673824	0.5007
EXPER	-0.005241	0.021205	-0.247149	0.8049
EXPER^2^2	2.50E-07	4.70E-07	0.531774	0.5951

R-squared	0.037890	Mean dependent var	0.169313
Adjusted R-squared	0.015385	S.D. dependent var	0.278118
S.E. of regression	0.275970	Akaike info criterion	0.287356
Sum squared resid	39.06984	Schwarz criterion	0.392772
Log likelihood	-62.57470	Hannan-Quinn criter.	0.328631
F-statistic	1.683600	Durbin-Watson stat	1.970279
Prob(F-statistic)	0.066942		

Estatística de teste

não há evidência de heterocedasticidade a 5%

HETEROCEDASTICIDADE: Testes

- **Teste simplificado de White**

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

- Teste de hipóteses

$$H_0 : Var(u \mid x_1,\ldots,x_k) = \sigma^2 \qquad H_1 : Var(u \mid x_1,\ldots,x_k) = h(x_1, x_2,\ldots,x_k)$$

- **Procedimento de teste**

1. Estimação OLS do modelo original ➡ $\hat{u},\ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$
2. Regressão auxiliar de teste: OLS de

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v$$

3. Testar

$$H_0 : \delta_1 = \delta_2 = 0 \qquad H_1 : H_0 \text{ falsa}$$

$\boxed{\text{Teste } F \text{ de significância global da regressão auxiliar de teste}}$

4. Conclusão

Rejeitar H_0 ➡ evidência de heterocedasticidade no erro do modelo inicial

HETEROCEDASTICIDADE: Testes

- **Exemplo**

$\hat{u}^2$ Com $\hat{u}$ o resíduo da regressão para log(wage)

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.233483	0.209660	1.113627	0.2660
$\hat{y}$ FIT	-0.187119	0.263787	-0.709356	0.4784
$\hat{y}^2$ FIT^2	0.087191	0.081226	1.073446	0.2836

R-squared	0.014904	Mean dependent var	0.169313
Adjusted R-squared	0.011137	S.D. dependent var	0.278118
S.E. of regression	0.276565	Akaike info criterion	0.272944
Sum squared resid	40.00326	Schwarz criterion	0.297271
Log likelihood	-68.78420	Hannan-Quinn criter.	0.282469
F-statistic	3.956441	Durbin-Watson stat	1.943744
Prob(F-statistic)	0.019706		

Estatística de teste

Evidência de heterocedasticidade