

Teste de Homocedasticidade.

Aula 14

Gujarati, 2000, Capítulo 11

Heij et al., Seção 5.4

Wooldridge, 2011 – Capítulo 8 (seções 8.1 a 8.4)

HETEROCEDASTICIDADE

Essa aula objetiva responder às seguintes perguntas:

- ✓ Qual é a natureza da heterocedasticidade?
- ✓ O que acontece com as propriedades dos estimadores de MQO quando a suposição de homocedasticidade dos erros é violada?
- ✓ Como podemos testar se há ou não validade da suposição de homocedasticidade dos erros.
- ✓ Caso exista, como levar em conta (ou corrigir) corrigir a heterocedasticidade?

HETEROCEDASTICIDADE

Leitura DETALHADA de Wooldridge (2011, Seções 8.1 a 8.4)

O que é Heterocedasticidade?

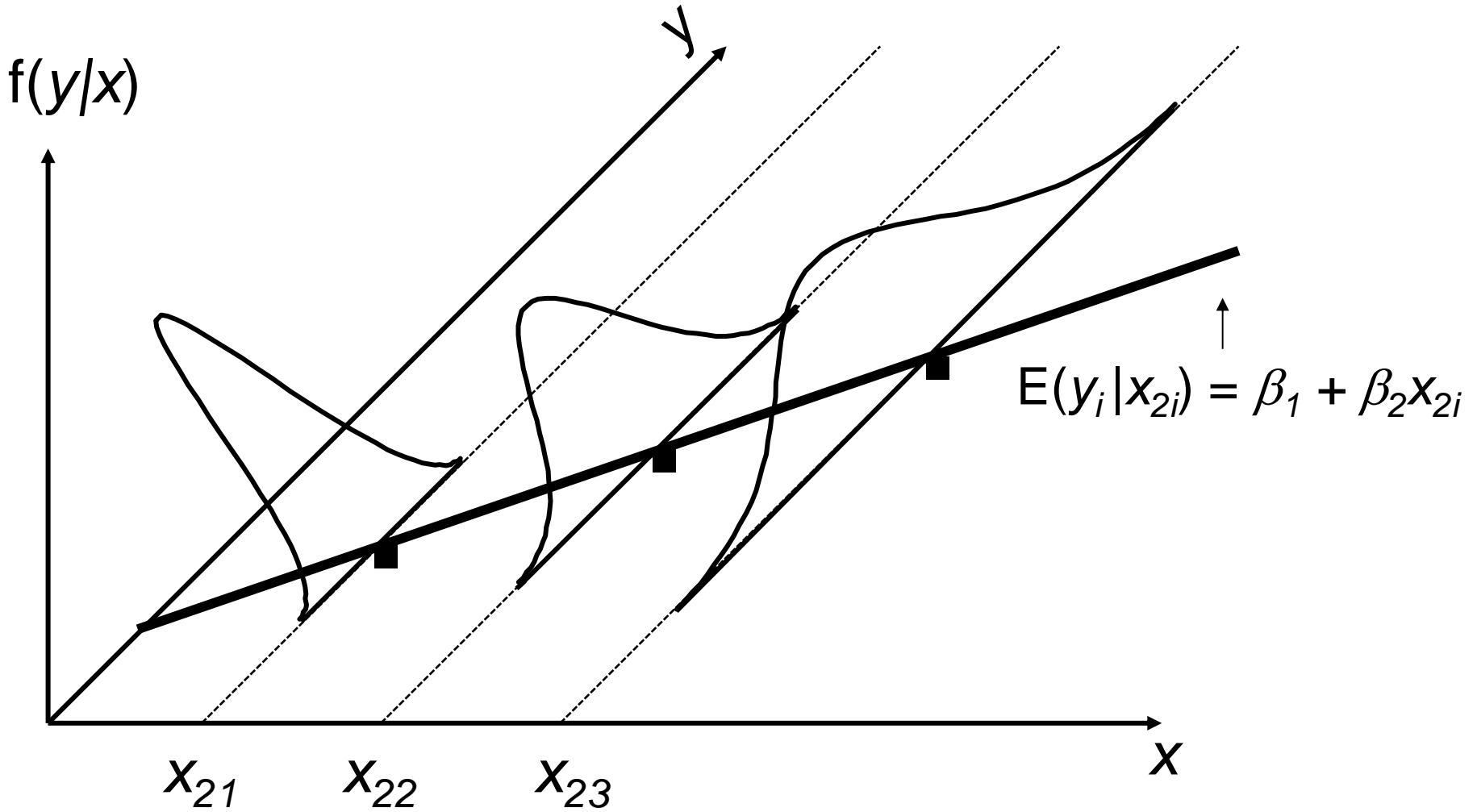
A suposição de homocedasticidade implica que, condicional às variáveis explicativas, a variância do erro é constante.

Ou seja,

$$\text{Var}(\varepsilon_i | x_{2i}, x_{3i}, \dots, x_{ki}) = \text{Var}(\varepsilon_i) = \sigma^2, \quad i = 1, 2, \dots, n$$

A homocedasticidade não se verifica sempre que a variância dos fatores não observáveis muda ao longo de diferentes segmentos da população, nos quais os segmentos são determinados pelos diferentes valores das variáveis explicativas.

Exemplo



O que é Heterocedasticidade?

Assim, do *slide* anterior, considerando mantidas as suposições 1, 2, 4, 5 e 6 (ou seja, a terceira suposição está sendo violada), teremos algo como, por exemplo

$$\text{Var}\left(\begin{array}{c} \boldsymbol{\varepsilon} \\ \sim \end{array} \mid \mathbf{X}\right) = E\left(\begin{array}{c} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \\ \sim \sim \end{array} \mid \mathbf{X}\right) = \sigma^2 \underset{\sim}{\boldsymbol{\Omega}} = \sigma^2 \begin{pmatrix} \omega_1 & & & & \\ & \omega_2 & & & \\ & & \omega_3 & & \\ & & & \dots & \\ & & & & \omega_n \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \dots & \\ & & & & \sigma_n^2 \end{pmatrix}$$

CONSEQUÊNCIAS PARA OS ESTIMADORES DE MQO

$$\hat{\beta}_{\sim}^{(MQO)} \stackrel{\text{X tem posto completo}}{=} \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' \underset{\sim}{y} \stackrel{\text{sup:6}}{=} \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' \left(\underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \right) = \underset{\sim}{\beta} + \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' \underset{\sim}{\varepsilon}$$

$$E \left(\hat{\beta}_{\sim}^{(MQO)} \right) \stackrel{\text{sup:1,6}}{=} E \left[\underset{\sim}{\beta} + \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' \underset{\sim}{\varepsilon} \right] \stackrel{\text{sup:5}}{=} \underset{\sim}{\beta} + \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' E \left[\underset{\sim}{\varepsilon} \right] \stackrel{\text{sup:2}}{=} \underset{\sim}{\beta}$$

$$\text{plim} \left(\hat{\beta}_{\sim}^{(MQO)} \right) \stackrel{\text{sup:6}}{=} \text{plim} \left(\underset{\sim}{\beta} + \left(\underset{\sim}{X}' \underset{\sim}{X} \right)^{-1} \underset{\sim}{X}' \underset{\sim}{\varepsilon} \right) \stackrel{\text{sup:1*,2,5}}{=} \underset{\sim}{\beta}$$

CONSEQUÊNCIAS PARA OS ESTIMADORES DE MQO

Note que a suposição de homocedasticidade (suposição 3) NÃO desempenha papel algum na demonstração de que o vetor de estimadores de mínimos quadrados ordinários do vetor de parâmetros do modelo de regressão linear múltipla é não viesado e consistente.

CONSEQUÊNCIAS PARA OS ESTIMADORES DE MQO

Todavia,

$$\begin{aligned} \text{Var}\left(\hat{\beta}_{\sim}^{(MQO)}\right) &= E\left[\left(\hat{\beta}_{\sim}^{(MQO)} - \beta_{\sim}\right)\left(\hat{\beta}_{\sim}^{(MQO)} - \beta_{\sim}\right)'\right] = E\left\{\left[\left(X'_{\sim} X_{\sim}\right)^{-1} X'_{\sim} \varepsilon_{\sim}\right]\left[\left(X'_{\sim} X_{\sim}\right)^{-1} X'_{\sim} \varepsilon_{\sim}\right]'\right\} = \\ &= E\left\{\left(X'_{\sim} X_{\sim}\right)^{-1} X'_{\sim} \varepsilon_{\sim} \varepsilon_{\sim}' X_{\sim} \left(X'_{\sim} X_{\sim}\right)^{-1}\right\} = \left(X'_{\sim} X_{\sim}\right)^{-1} X'_{\sim} E\left\{\varepsilon_{\sim} \varepsilon_{\sim}'\right\} X_{\sim} \left(X'_{\sim} X_{\sim}\right)^{-1} = \\ &= \sigma^2 \left(X'_{\sim} X_{\sim}\right)^{-1} X'_{\sim} \Omega_{\sim} X_{\sim} \left(X'_{\sim} X_{\sim}\right)^{-1} \end{aligned}$$

CONSEQUÊNCIAS PARA OS ESTIMADORES DE MQO

Aqui, percebemos que a **expressão usual** de cálculo da variância dos estimadores, quando a suposição de homocedasticidade é válida, dada por

$$\text{Var}\left(\hat{\beta}^{(MQO)}\right) = E\left[\left(\hat{\beta}^{(MQO)} - \beta\right)\left(\hat{\beta}^{(MQO)} - \beta\right)'\right] = \sigma^2\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}$$

não se aplica mais.

CONSEQUÊNCIAS PARA OS ESTIMADORES DE MQO

➤ A partir das expressões dos *slides* 5 e 8, pode-se demonstrar que os estimadores das variâncias dos estimadores dos parâmetros do modelo de regressão linear múltipla são viesados, se não for válida a suposição de homocedasticidade, o que afeta o erro-padrão dos estimadores de mínimos quadrados;

CONSEQUÊNCIAS PARA OS ESTIMADORES DE MQO

- Isso significa que os intervalos de confiança e os testes t , F e LM são prejudicados;
- Também, é sabido que o Teorema de Gauss-Markov, que afirma que os estimadores de MQO são os melhores estimadores lineares não viesados (BLUE), vale-se de forma crucial da suposição de homocedasticidade. Assim, na presença da heterocedasticidade, os estimadores de MQO não são mais BLUE e nem assintoticamente eficientes.

Observações

- (a)** A suposição de homocedasticidade é necessária para a determinação das distribuições das somas de quadrados e das estatísticas dos testes de hipóteses.
- (b)** Todavia, quando os erros são heterocedásticos, os estimadores de MQO dão mais peso para os resíduos associados às observações com maior variância, já que a soma de quadrados dos resíduos (SSR) associados com os termos de maior variância tende a ser maior que aquela associada aos termos de menor variância.

Observações

(c) A suposição de homocedasticidade entra fundamentalmente na derivação das distribuições das variáveis presentes nos testes. Logo, toda a análise neles baseada não é válida (a falha na suposição de homocedasticidade é mais grave que a falha na suposição de normalidade).

TESTES DE HOMOCEDASTICIDADE

Considere

$$H_0 : \text{Var}(\varepsilon_i | x_{2i}, x_{3i}, \dots, x_{ki}) = \text{Var}(\varepsilon_i) = \sigma^2, \quad i = 1, 2, \dots, n$$

Homocedasticidade $\Rightarrow H_0: \text{Var}(\varepsilon | \mathbf{x}_2, \dots, \mathbf{x}_k) =$

$$= E(\varepsilon^2 | \mathbf{x}_2, \dots, \mathbf{x}_k) - [E(\varepsilon | \mathbf{x}_2, \dots, \mathbf{x}_k)]^2 =$$
$$= E(\varepsilon^2 | \mathbf{x}_2, \dots, \mathbf{x}_k) = \sigma^2 .$$

Para H_0 ser rejeitada, precisamos encontrar relação entre o ε^2 e variáveis explicativas.

TESTES DE HOMOCEDASTICIDADE

Ou seja, do *slide* anterior, vem que:

$$\varepsilon^2 = \delta_1 + \delta_2 X_2 + \dots + \delta_k X_k + v,$$

em que

v – termo de erro estocástico da regressão auxiliar.

Assim,

H_0 (hipótese de homocedasticidade): $\delta_2 = \dots = \delta_k = 0$,

a qual pode ser testada utilizando um teste F ou LM.

TESTES DE HOMOCEDASTICIDADE

Todavia, como a variável resposta, no modelo anteriormente proposto, não é diretamente observada, teremos que estimá-la.

Assim, $\hat{\varepsilon}_i$ (resíduo do modelo de regressão de interesse), será a estimativa do erro ε_i .

Dessa maneira, a equação anterior fica dada por:

$$\hat{\varepsilon}_i^2 = \delta_1 + \delta_2 X_2 + \dots + \delta_k X_k + v.$$

TESTES DE HOMOCEDASTICIDADE

Estimando os parâmetros do modelo anterior, podemos calcular a estatística LM para verificar a relevância conjunta das variáveis x_2, \dots, x_k , como segue:

$$LM = n \cdot R_{\hat{e}^2}^2 \sim \chi_{(k-1)}^2$$

TESTES DE HOMOCEDASTICIDADE

- 1) Rejeitamos H_0 quando o valor observado da estatística de teste for superior ao crítico;
- 2) A versão LM deste teste é conhecida na literatura como **TESTE DE BREUSCH-PAGAN** (Teste BP).

Observações

- a. Podemos considerar apenas um sub-conjunto das variáveis explicativas;
- b. Se H_0 for rejeitada, então, precisaremos recorrer a algum método de estimação que leve em conta a violação da suposição de homocedasticidade.

APLICAÇÃO

Considere o seguinte modelo de regressão linear múltipla

$$price = \beta_1 + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + \varepsilon$$

em que

price – preço da casa (em milhares de dólares);

lotsize – tamanho do terreno (em pés²);

sqrft – área construída (em pés²);

bdrms – número de quartos.

APLICAÇÃO (cont.)

Utilizando os dados do arquivo *HPRICE.XLS*:

- a) Encontre as estimativas de MQO dos parâmetros do modelo proposto e interprete-as.

- a) Conduza um teste BP para verificar se os erros do modelo proposto são homocedásticos. Comente.

a) Estimação, por MQO, dos parâmetros do modelo:

Dependent Variable: PRICE

Method: Least Squares

Date: 10/29/11 Time: 10:48

Sample: 1 88

Included observations: 88

PRICE=C(1)+C(2)*LOTSIZE+C(3)*SQRFT+C(4)*BDRMS

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-21.77031	29.47504	-0.738601	0.4622
C(2)	0.002068	0.000642	3.220096	0.0018
C(3)	0.122778	0.013237	9.275093	0.0000
C(4)	13.85252	9.010145	1.537436	0.1279
R-squared	0.672362	Mean dependent var		293.5460
Adjusted R-squared	0.660661	S.D. dependent var		102.7134
S.E. of regression	59.83348	Akaike info criterion		11.06540
Sum squared resid	300723.8	Schwarz criterion		11.17800
Log likelihood	-482.8775	Hannan-Quinn criter.		11.11076
F-statistic	57.46023	Durbin-Watson stat		2.109796
Prob(F-statistic)	0.000000			

b) Teste BP:

Dependent Variable: PRICE

Method: Least Squares

Date: 10/29/11 Time: 10:48

Sample: 1 88

Included observations: 88

PRICE=C(1)+C(2)*LOTSIZE+C(3)*SQRFT+C(4)*BDRMS

**Salve os resíduos ao
quadrado deste modelo
estimado (por exemplo, num
objeto chamado residsg)!!!**

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-21.77031	29.47504	-0.738601	0.4622
C(2)	0.002068	0.000642	3.220096	0.0018
C(3)	0.122778	0.013237	9.275093	0.0000
C(4)	13.85252	9.010145	1.537436	0.1279
R-squared	0.672362	Mean dependent var		293.5460
Adjusted R-squared	0.660661	S.D. dependent var		102.7134
S.E. of regression	59.83348	Akaike info criterion		11.06540
Sum squared resid	300723.8	Schwarz criterion		11.17800
Log likelihood	-482.8775	Hannan-Quinn criter.		11.11076
F-statistic	57.46023	Durbin-Watson stat		2.109796
Prob(F-statistic)	0.000000			

b) Teste BP: (cont.)

Dependent Variable: RESIDSQ

Method: Least Squares

Date: 10/29/11 Time: 10:52

Sample: 1 88

Included observations: 88

RESIDSQ=C(1)+C(2)*LOTSIZE+C(3)*SQRFT+C(4)*BDRMS

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-5522.795	3259.478	-1.694380	0.0939
C(2)	0.201521	0.071009	2.837961	0.0057
C(3)	1.691037	1.463850	1.155198	0.2513
C(4)	1041.760	996.3810	1.045544	0.2988
R-squared	0.160141	Mean dependent var		3417.316
Adjusted R-squared	0.130146	S.D. dependent var		7094.384
S.E. of regression	6616.646	Akaike info criterion		20.47695
Sum squared resid	3.68E+09	Schwarz criterion		20.58956
Log likelihood	-896.9860	Hannan-Quinn criter.		20.52232
F-statistic	5.338919	Durbin-Watson stat		2.351111
Prob(F-statistic)	0.002048			

b) Teste BP: (cont.)

$$\chi^2_{\text{obs}} = 88 * 0,160141 = 14,092$$

$$\chi^2(0,05; 3) = @qchisq(0.95, 3) = 7,81$$

$$\text{p-valor} = 1 - @cchisq(14.092, 3) = 0,0028$$

Solução via Eviews:

Na janela em que aparecem os resultados do modelo estimado (completo), clique na opção *View*, em seguida na opção *Residual Diagnostics*, em seguida na opção *Heteroskedasticity Tests*, finalmente, na opção *Breusch-Pagan-Godfrey*.

Solução via Eviews:

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Representations									
Estimation Output									
Actual, Fitted, Residual									
ARMA Structure...									
Gradients and Derivatives									
Covariance Matrix									
Coefficient Diagnostics									
Residual Diagnostics									
Stability Diagnostics									
Label									
S.E. of regression									
Sum squared resid									
Log likelihood									
F-statistic									
Prob(F-statistic)									
Std. Error									
t-Statistic									
Prob.									
29.47504									
-0.738601									
0.4622									
0.000642									
3.220096									
0.0018									
Correlogram - Q-statistics...									
Correlogram Squared Residuals...									
Histogram - Normality Test									
Serial Correlation LM Test...									
Heteroskedasticity Tests...									
Durbin-Watson stat									
2.109796									

Solução via Eviews:

The screenshot shows the EViews software interface. The main window displays the following information:

- Equation: UNTITLED Workfile: HPRICE1::Hprice1
- View Proc Object Print Name Freeze Estimate Forecast Stats Resids
- Dependent Variable: PRICE
- Method: Least Squares
- Date: 10/29/11 Time: 10:55
- Sample: 1 88
- Included observations: 88
- PRICE=C(1)+C(2)*LOTSIZE+C(3)*SQRFT+C(4)*BDRMS

The Heteroskedasticity Tests dialog box is open, showing the following options:

- Specification: Test type: Breusch-Pagan-Godfrey (highlighted with a red circle)
- Dependent variable: RESID^2
- The Breusch-Pagan-Godfrey Test regresses the squared residuals on the original regressors by default.
- Regressors: lotsize sqrft bdrms (highlighted with a red circle)
- Buttons: OK, Cancel

The background window shows a list of coefficients (C(1) to C(4)) and various statistical measures (R-squared, Adjusted R-squared, S.E. of regression, Sum squared resid, Log likelihood, F-statistic, Prob(F-statistic)).

Solução via Eviews:

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	5.338919	Prob. F(3,84)	0.0020
Obs*R-squared	14.09239	Prob. Chi-Square(3)	0.0028
Scaled explained SS	27.35542	Prob. Chi-Square(3)	0.0000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/29/11 Time: 10:59

Sample: 1 88

Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5522.795	3259.478	-1.694380	0.0939
LOTSIZE	0.201521	0.071009	2.837961	0.0057
SQRFT	1.691037	1.463850	1.155198	0.2513
BDRMS	1041.760	996.3810	1.045544	0.2988

R-squared	0.160141	Mean dependent var	3417.316
Adjusted R-squared	0.130146	S.D. dependent var	7094.384
S.E. of regression	6616.646	Akaike info criterion	20.47695
Sum squared resid	3.68E+09	Schwarz criterion	20.58956
Log likelihood	-896.9860	Hannan-Quinn criter.	20.52232
F-statistic	5.338919	Durbin-Watson stat	2.351111
Prob(F-statistic)	0.002048		

TESTES DE HOMOCEDASTICIDADE

TESTE DE WHITE

White (1980), segundo Wooldridge (2011), motivado pelo fato de que a suposição

$$\text{Var}(\varepsilon_i | x_{2i}, x_{3i}, \dots, x_{ki}) = \sigma^2, \quad i = 1, 2, \dots, n$$

podia ser substituída pela suposição mais fraca de que bastaria o erro ao quadrado, ε^2 , ser não correlacionado com todas as variáveis explicativas, com os quadrados das variáveis explicativas e com todos os produtos cruzados entre as variáveis explicativas, escreveu seu clássico artigo.

TESTES DE HOMOCEDASTICIDADE

TESTE DE WHITE

Assim, por exemplo, quando o modelo de interesse apresentar $k = 3$ variáveis explicativas, o teste de White ficará baseado nos resultados da estimação do seguinte modelo de regressão auxiliar

$$\begin{aligned}\hat{e}^2 = & \delta_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_4 + \\ & + \delta_5 x_2^2 + \delta_6 x_3^2 + \delta_7 x_4^2 + \\ & + \delta_8 x_2 x_3 + \delta_9 x_2 x_4 + \delta_{10} x_3 x_4 + v\end{aligned}$$

TESTES DE HOMOCEDASTICIDADE

TESTE DE WHITE

Neste caso, a hipótese nula de interesse seria

$$H_0: \delta_2 = \dots = \delta_{10} = 0,$$

(hipótese de homocedasticidade)

a qual, segundo White (1980), pode ser testada usando:

$$LM = n \cdot R_{\hat{e}^2}^2 \sim \chi_{(9)}^2$$

TESTES DE HOMOCEDASTICIDADE

TESTE DE WHITE

Comparado ao teste proposto por Breusch-Pagan, a equação auxiliar do *slide* 30 envolve 6 regressores a mais. Ou seja, dependendo do número de variáveis explicativas constantes do modelo original de interesse, a perda de graus de liberdade será, também, bastante grande. Dessa forma, é importante observar se o tamanho da amostra que estamos usando para a estimação é suficientemente grande para a aplicação de tal metodologia.

TESTES DE HOMOCEDASTICIDADE

TESTE DE WHITE

OBSERVAÇÕES

- 1) A equação auxiliar envolvida no teste pode ser útil na identificação da forma funcional da heterocedasticidade ou de erro de especificação, ou de ambos.
- 2) Para minimizar a perda de graus de liberdade, discutida no slide anterior, o teste pode ser conduzido utilizando a seguinte regressão auxiliar:

$$\hat{e}^2 = \gamma_1 + \gamma_2 \hat{y} + \gamma_3 \hat{y}^2 + \xi$$

VOLTANDO À APLICAÇÃO

Considere o seguinte modelo de regressão linear múltipla

$$price = \beta_1 + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + \varepsilon$$

Ainda, utilizando a base de dados *HPRICE.xls*, conduza o teste de White para verificar se os erros são homocedásticos.

Solução via Eviews:

Na janela em que aparecem os resultados do modelo estimado (completo), clique na opção *View*, em seguida na opção *Residual Diagnostics*, em seguida na opção *Heteroskedasticity Tests*, finalmente, na opção *White*.

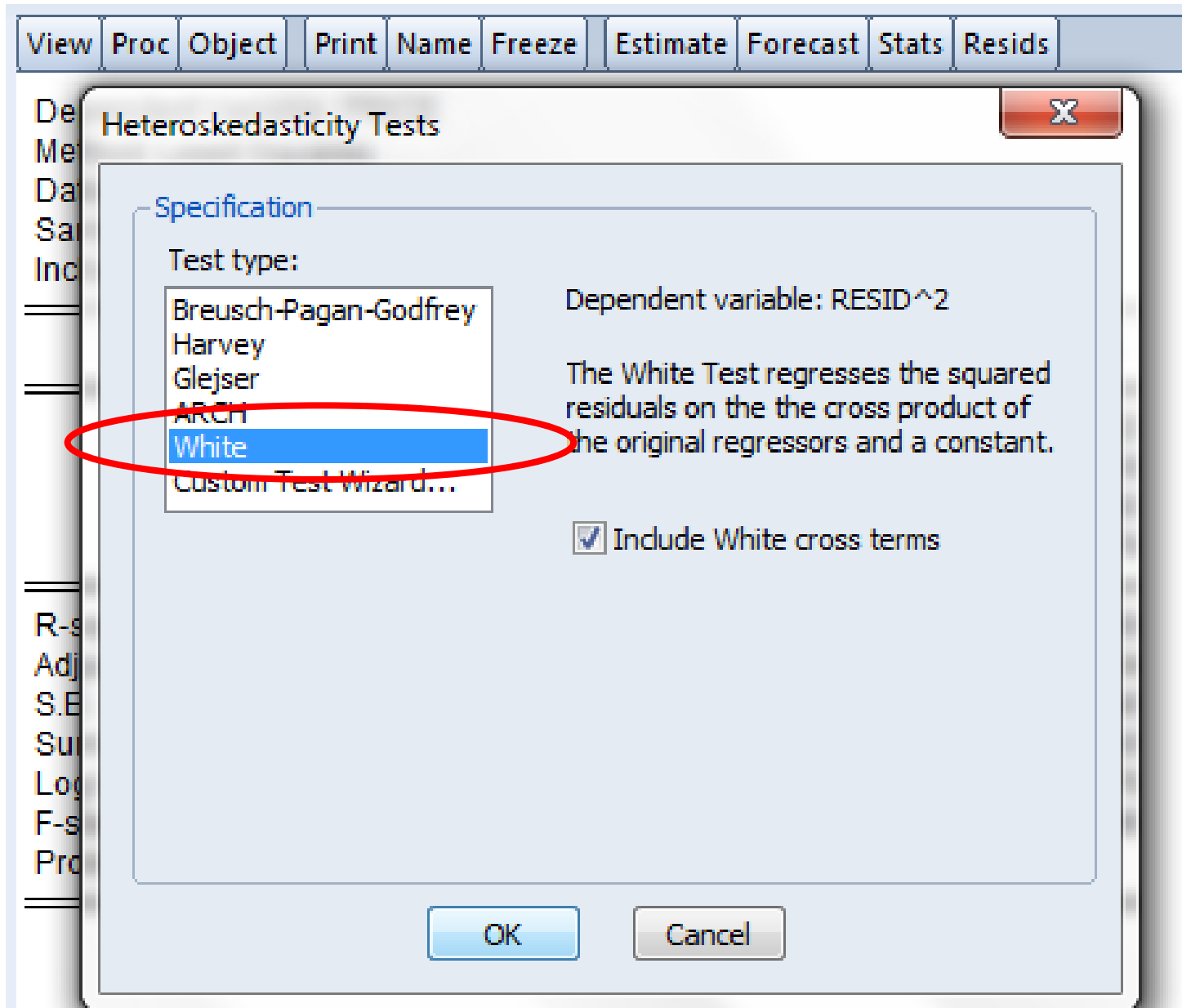
Teste de White

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Representations									
Estimation Output									
Actual, Fitted, Residual ▶									
ARMA Structure...									
Gradients and Derivatives ▶									
Covariance Matrix									
Coefficient Diagnostics ▶									
Residual Diagnostics ▶									
Stability Diagnostics ▶									
Label									
S.E. of regression				39.63346					
Sum squared resid				300723.8					
Log likelihood				-482.8775					
F-statistic				57.46023		Durbin-Watson stat		2.109796	
Prob(F-statistic)				0.000000					

	Std. Error	t-Statistic	Prob.
	29.47504	-0.738601	0.4622
	0.000642	3.220096	0.0018

Correlogram - Q-statistics...
Correlogram Squared Residuals...
Histogram - Normality Test
Serial Correlation LM Test...
Heteroskedasticity Tests...

Teste de White



Teste de White

Heteroskedasticity Test: White

F-statistic	5.386953	Prob. F(9,78)	0.0000
Obs*R-squared	33.73166	Prob. Chi-Square(9)	0.0001
Scaled explained SS	65.47818	Prob. Chi-Square(9)	0.0000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 10/29/11 Time: 11:07

Sample: 1 88

Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	15626.24	11369.41	1.374411	0.1733
LOTSIZE	-1.859507	0.637097	-2.918719	0.0046
LOTSIZE^2	-4.98E-07	4.63E-06	-0.107498	0.9147
LOTSIZE*SQRFT	0.000457	0.000277	1.649673	0.1030
LOTSIZE*BDRMS	0.314647	0.252094	1.248135	0.2157
SQRFT	-2.673918	8.662183	-0.308689	0.7584
SQRFT^2	0.000352	0.001840	0.191484	0.8486
SQRFT*BDRMS	-1.020860	1.667154	-0.612337	0.5421
BDRMS	-1982.841	5438.483	-0.364595	0.7164
BDRMS^2	289.7541	758.8303	0.381843	0.7036

R-squared	0.383314	Mean dependent var	3417.316
Adjusted R-squared	0.312158	S.D. dependent var	7094.384
S.E. of regression	5883.814	Akaike info criterion	20.30444
Sum squared resid	2.70E+09	Schwarz criterion	20.58596
Log likelihood	-883.3955	Hannan-Quinn criter.	20.41786
F-statistic	5.386953	Durbin-Watson stat	2.052712
Prob(F-statistic)	0.000010		

TESTES DE HETEROCEDASTICIDADE

Exercício

Para minimizar a perda de graus de liberdade, repita a aplicação anterior utilizando o seguinte modelo de regressão auxiliar

$$\hat{e}^2 = \gamma_1 + \gamma_2 \hat{y} + \gamma_3 \hat{y}^2 + \xi$$

para testar

$$H_0: \gamma_2 = \gamma_3 = 0 \text{ (o erro é homocedástico).}$$

Compare todos os resultados obtidos.

ESTIMADORES ROBUSTOS

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

Na prática é muito difícil conhecer a verdadeira forma como a heterocedasticidade se apresenta. Assim, precisamos buscar alguma metodologia que nos forneça resultados válidos **na presença de heterocedasticidade** cuja forma é desconhecida.

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

Recentemente, muito se tem desenvolvido com relação ao ajuste de erros padrões e estatísticas de testes para que os mesmos se tornem válidos na presença de heterocedasticidade.

Estes procedimentos são conhecidos como **ROBUSTOS** pois são válidos, pelo menos com amostras grandes, sendo ou não a variância do erro constante.

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

Apenas para ilustração, suponha

$$\text{Var}\left(\begin{array}{c} \boldsymbol{\varepsilon} \\ \sim \\ \sim \end{array} \mid \mathbf{X}\right) = E\left(\begin{array}{c} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \\ \sim \sim \\ \sim \end{array} \mid \mathbf{X}\right) = \sigma^2 \underset{\sim}{\boldsymbol{\Omega}} = \sigma^2 \begin{pmatrix} \omega_1 & & & & \\ & \omega_2 & & & \\ & & \omega_3 & & \\ & & & \dots & \\ & & & & \omega_n \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \dots & \\ & & & & \sigma_n^2 \end{pmatrix}$$

em que, σ_i^2 , $i = 1, 2, \dots, n$, são parâmetros desconhecidos.

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

Assim, do *slide* anterior, considerando mantidas as suposições 1, 2, 4, 5 e 6 (ou seja, somente a terceira suposição está sendo violada), teremos

$$\begin{aligned} \text{Var}\left(\underset{\sim}{\hat{\beta}}^{(MQO)}\right) &= E\left[\left(\underset{\sim}{\hat{\beta}}^{(MQO)} - \underset{\sim}{\beta}\right)\left(\underset{\sim}{\hat{\beta}}^{(MQO)} - \underset{\sim}{\beta}\right)'\right] = E\left\{\left[\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'\underset{\sim}{\varepsilon}\right]\left[\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'\underset{\sim}{\varepsilon}\right]'\right\} = \\ &= E\left\{\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}'\underset{\sim}{X}\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\right\} = \left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'E\left\{\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}'\right\}\underset{\sim}{X}\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1} = \\ &= \sigma^2\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'\underset{\sim}{\Omega}\underset{\sim}{X}\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1} \end{aligned}$$

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

Usando o resultado do *slide* 43 na expressão do *slide* 44, temos que

$$\text{Var}\left(\begin{array}{c} \hat{\beta}^{(MQO)} \\ \sim \end{array}\right) = \left(\begin{array}{c} X' \\ \sim \end{array} \begin{array}{c} X \\ \sim \end{array}\right)^{-1} \left(\sum_{i=1}^n \sigma_i^2 \begin{array}{c} x_i \\ \sim \end{array} \begin{array}{c} x_i' \\ \sim \end{array}\right) \left(\begin{array}{c} X' \\ \sim \end{array} \begin{array}{c} X \\ \sim \end{array}\right)^{-1}$$

White (1980) demonstrou que uma estimativa bem simples das quantidades desconhecidas pode ser obtida a partir do cálculo de $\hat{\epsilon}_i^2$ (quadrado do resíduo de MQO).

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

Assim,

$$\widehat{\text{Var}}\left(\hat{\beta}^{(MQO)}\right) = \left(\underset{\sim}{X}' \underset{\sim}{X}\right)^{-1} \left(\sum_{i=1}^n \hat{e}_i^2 \underset{\sim}{x}_i \underset{\sim}{x}_i'\right) \left(\underset{\sim}{X}' \underset{\sim}{X}\right)^{-1}$$

Se tomarmos a raiz quadrada dos elementos da diagonal principal desta matriz teremos o que usualmente costuma se chamar de erro-padrão devido a White (ou erro padrão robusto).

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

Observações:

1. O erro-padrão robusto pode ser maior ou menor do que o erro-padrão não robusto (não sabemos se o viés é para cima ou para baixo);
2. Usando um método de estimação robusto, a estatística de teste t também será robusta.

EXEMPLO

Nesta questão foi utilizado o arquivo de dados ***HPRICE1.wf1***. Os resultados seguem :

$$\hat{price} = -21,77031 + 0,00207 \cdot lotsize + 0,12277 \cdot sqrft + 13,85252 \cdot bdrms$$

(29,47504)	(0,00064)	(0,01324)	(9,01014)
[37,13821]	[0,00125]	[0,01772]	[8,47862]

$$n = 88 \quad R^2 = 0,6724$$

Observação: entre colchetes encontram-se os erros padrões robustos.

Comente as principais diferenças encontradas ao utilizarmos o erro padrão robusto ao invés do usual.

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

**Como encontrar o erro padrão robusto
usando o Eviews?**

VOLTANDO À APLICAÇÃO

Considere o seguinte modelo de regressão linear múltipla

$$price = \beta_1 + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + \varepsilon$$

Ainda, utilizando a base de dados *HPRICE.xls*, estime os parâmetros do modelo de regressão de interesse, baseando-se em algum procedimento robusto.

Estimação via procedimento robusto:

Equation: UNTITLED Workfile: HPRICE1::Hprice1\

View Proc Object Print Name Freeze **Estimate** Forecast Stats Resids

Dependent Variable: PRICE

Method: Least Squares

Date: 10/29/11 Time: 11:33

Sample: 1 88

Included observations: 88

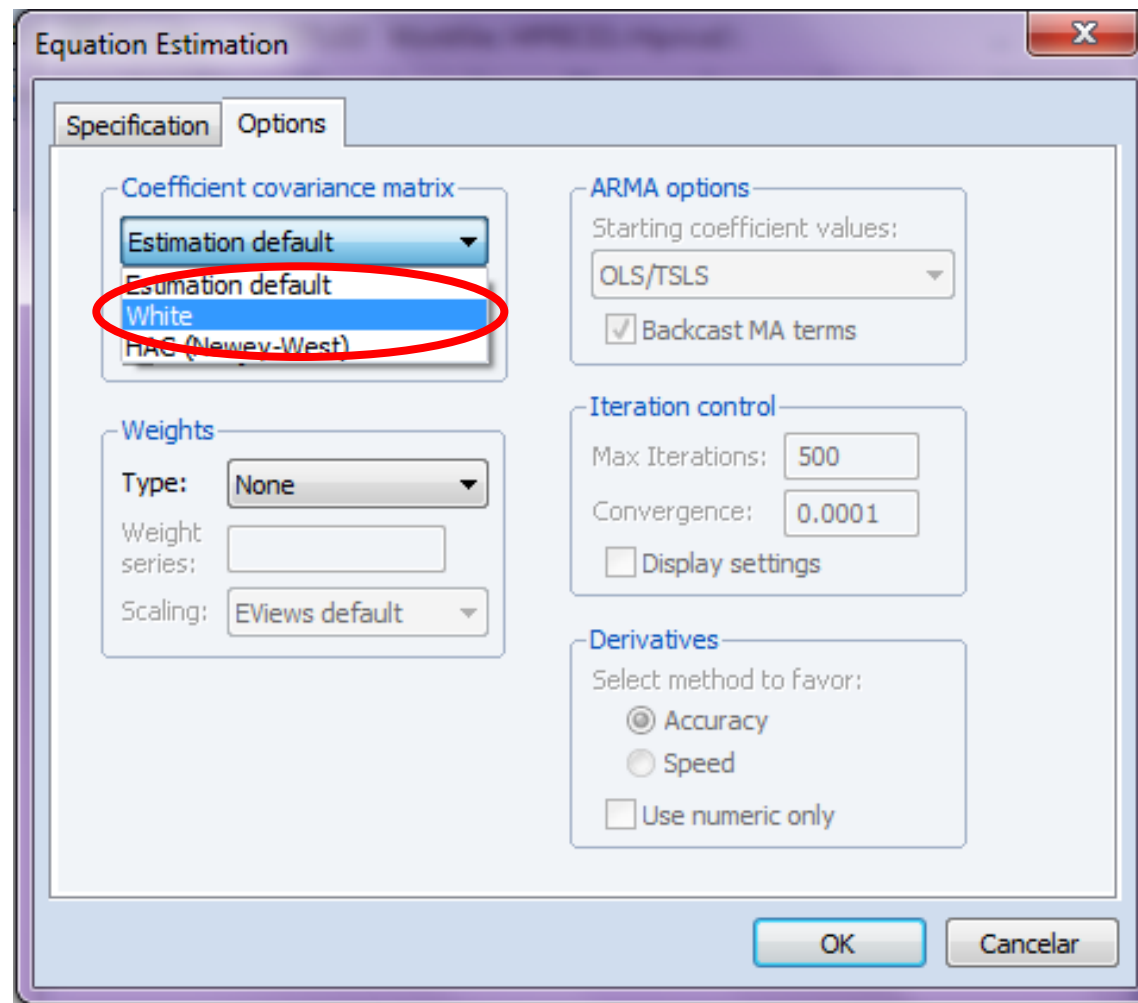
PRICE=C(1)+C(2)*LOTSIZE+C(3)*SQRFT+C(4)*BDRMS

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-21.77031	29.47504	-0.738601	0.4622
C(2)	0.002068	0.000642	3.220096	0.0018
C(3)	0.122778	0.013237	9.275093	0.0000
C(4)	13.85252	9.010145	1.537436	0.1279

R-squared	0.672362	Mean dependent var	293.5460
Adjusted R-squared	0.660661	S.D. dependent var	102.7134
S.E. of regression	59.83348	Akaike info criterion	11.06540
Sum squared resid	300723.8	Schwarz criterion	11.17800
Log likelihood	-482.8775	Hannan-Quinn criter.	11.11076
F-statistic	57.46023	Durbin-Watson stat	2.109796
Prob(F-statistic)	0.000000		

Clique no ícone estimate, na janela com o último modelo ajustado e, na seqüência clique em options.

Estimação via procedimento robusto:



Em Coefficient Covariance Matrix, selecionar a opção White.

Estimação via procedimento robusto:

Dependent Variable: PRICE

Method: Least Squares

Date: 10/29/11 Time: 11:43

Sample: 1 88

Included observations: 88

White heteroskedasticity-consistent standard errors & covariance

PRICE=C(1)+C(2)*LOTSIZE+C(3)*SQRFT+C(4)*BDRMS

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-21.77031	37.13821	-0.586197	0.5593
C(2)	0.002068	0.001251	1.652283	0.1022
C(3)	0.122778	0.017725	6.926707	0.0000
C(4)	13.85252	8.478625	1.633817	0.1060
R-squared	0.672362	Mean dependent var		293.5460
Adjusted R-squared	0.660661	S.D. dependent var		102.7134
S.E. of regression	59.83348	Akaike info criterion		11.06540
Sum squared resid	300723.8	Schwarz criterion		11.17800
Log likelihood	-482.8775	Hannan-Quinn criter.		11.11076
F-statistic	57.46023	Durbin-Watson stat		2.109796
Prob(F-statistic)	0.000000			

INFERÊNCIA APÓS A ESTIMAÇÃO VIA MÍNIMOS QUADRADOS (ESTIMADORES ROBUSTOS)

OBSERVAÇÃO

Usando os métodos de estimação robustos, é possível obter as estatísticas de testes, como a LM, robustas em relação à heterocedasticidade. Para mais detalhes, vide, por exemplo, Greene (2008), Heij *et al.* (2004) e Wooldridge (2006, 3 ed., Seção 8.2).

LEITURA COMPLEMENTAR I

ANÁLISE GRÁFICA

A primeira forma de detectar a existência de heterocedasticidade é através da análise gráfica dos resíduos. Assim, sugere-se a construção dos seguintes gráficos de dispersão:

- i. resíduos ao quadrado **versus** cada uma das variáveis explicativas;
- ii. resíduos ao quadrado **versus** os valores ajustados da variável resposta.

EXEMPLO

Nesta questão utilizaremos o arquivo de dados *HPRICE1.wf1*, para estimar o modelo a seguir:

$$price = \beta_1 + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + \varepsilon$$

price: preço da casa (em milhares de dólares);

lotsize: tamanho do terreno (em pés²);

sqrft: área construída (em pés²);

bdrms: número de quartos.

Encontre as estimativas de mínimos quadrados dos parâmetros deste modelo e faça uma análise gráfica dos resíduos. Comente os resultados encontrados.

i) Estimação dos dos parâmetros do modelo, por MQO.

Dependent Variable: PRICE

Method: Least Squares

Date: 10/12/10 Time: 10:54

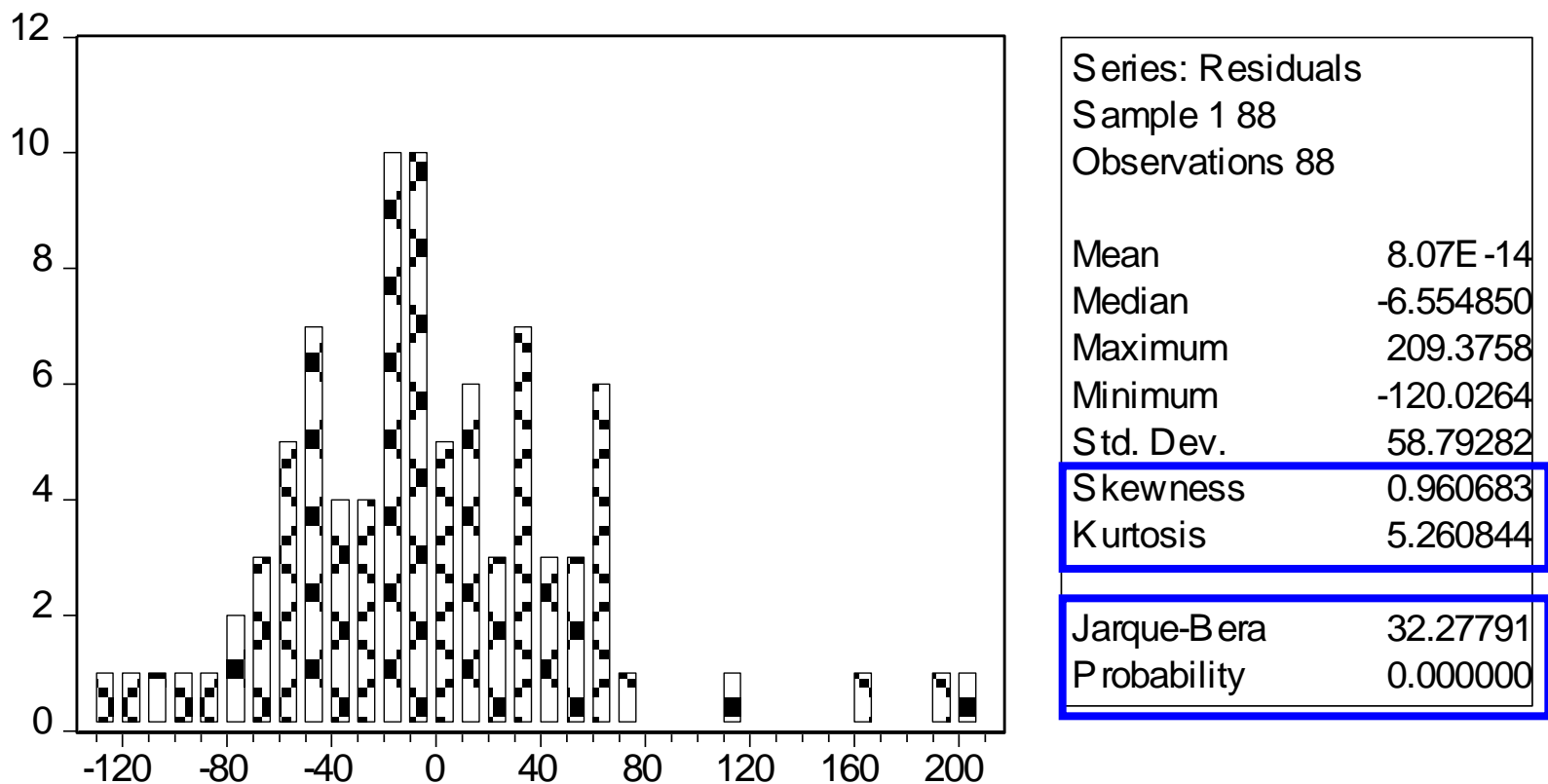
Sample: 1 88

Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-21.77031	29.47504	-0.738601	0.4622
LOTSIZE	0.002068	0.000642	3.220096	0.0018
SQRFT	0.122778	0.013237	9.275093	0.0000
BDRMS	13.85252	9.010145	1.537436	0.1279
R-squared	0.672362	Mean dependent var		293.5460
Adjusted R-squared	0.660661	S.D. dependent var		102.7134
S.E. of regression	59.83348	Akaike info criterion		11.06540
Sum squared resid	300723.8	Schwarz criterion		11.17800
Log likelihood	-482.8775	Hannan-Quinn criter.		11.11076
F-statistic	57.46023	Durbin-Watson stat		2.109796
Prob(F-statistic)	0.000000			

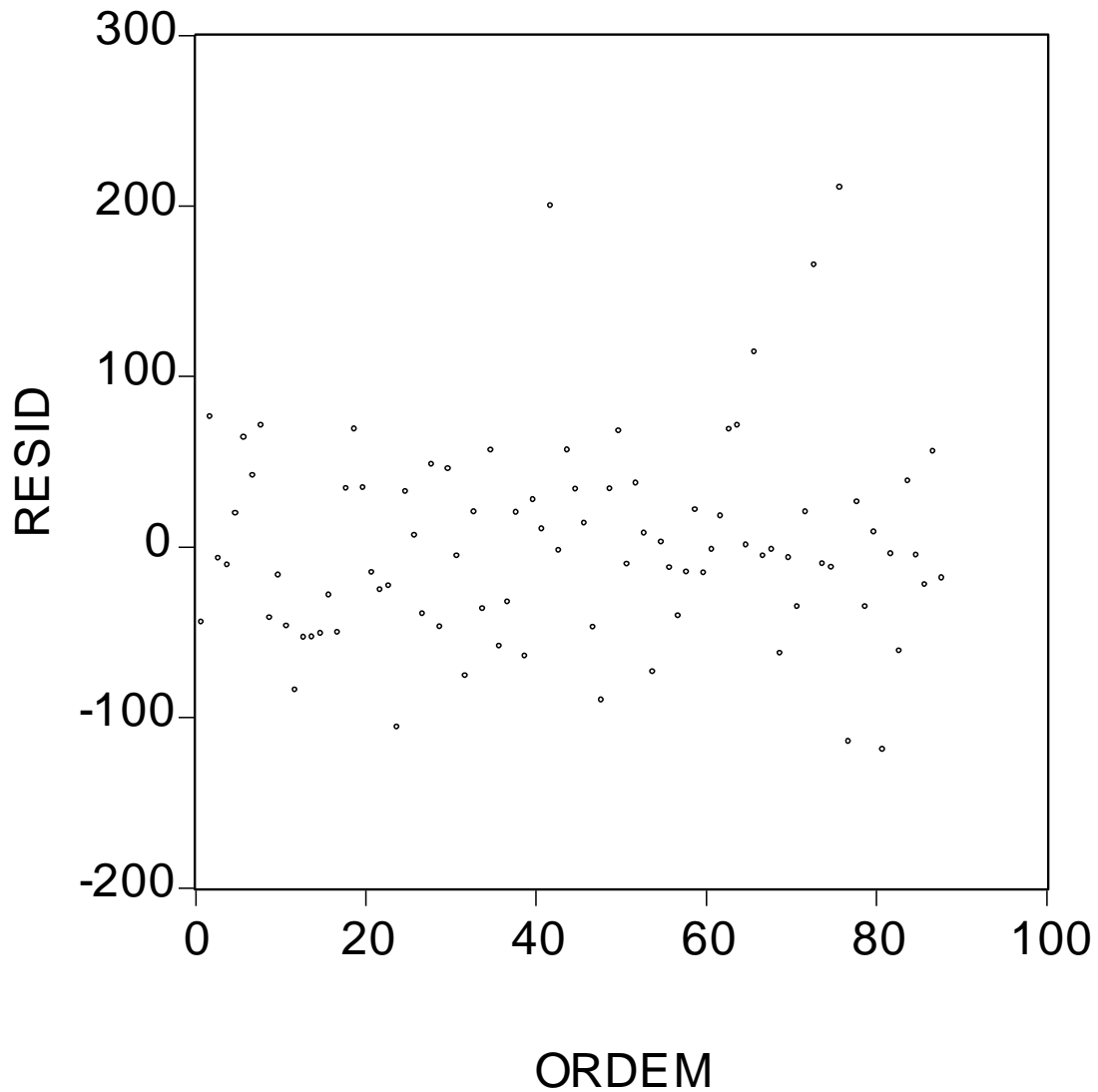
ii) Análise completa dos resíduos e comentários.

H_0 : Os erros são normalmente distribuídos.

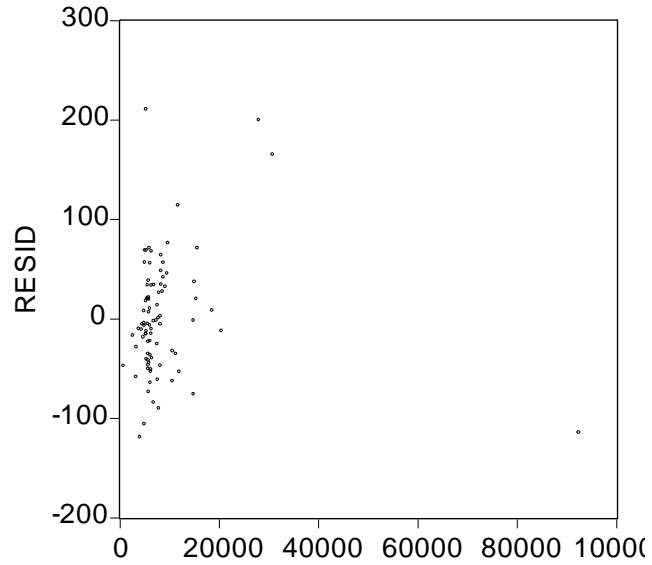


$$JB = \frac{n}{6} A\hat{s}^2 + \frac{n}{24} (C\hat{u}rt - 3)^2 \sim \chi_2^2$$

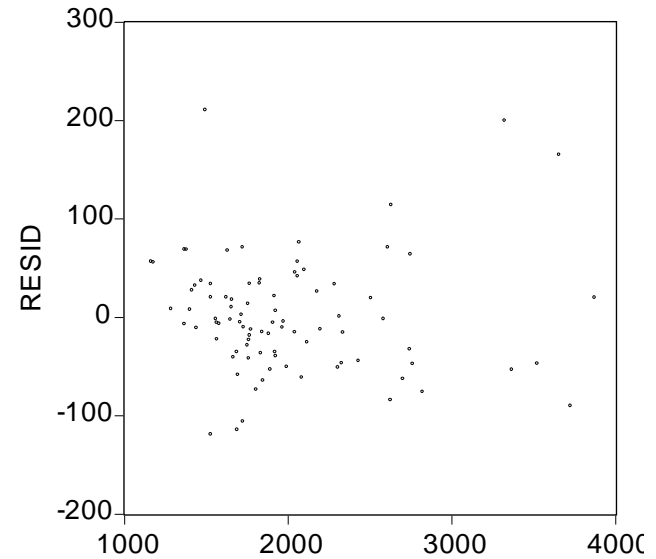
ii) Análise completa dos resíduos e comentários.



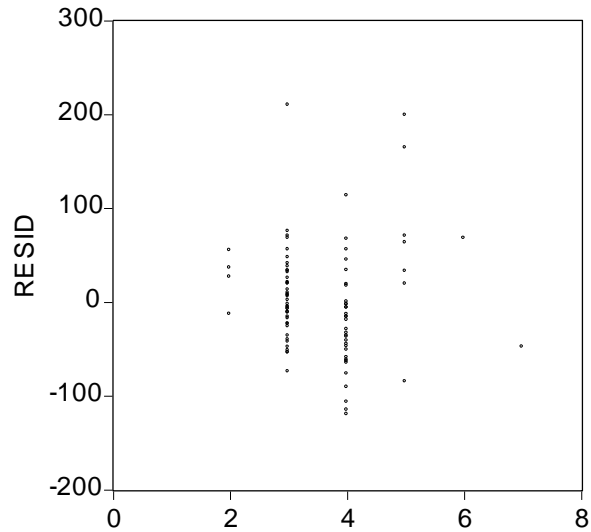
ii) Análise completa dos resíduos e comentários.



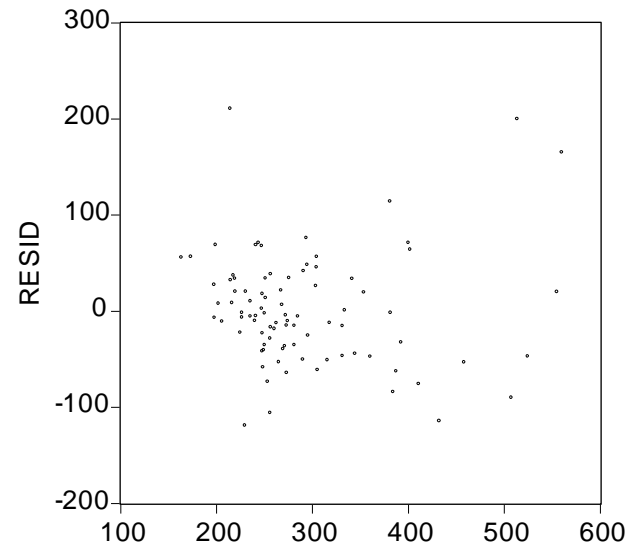
LOTSIZE



SQRFT

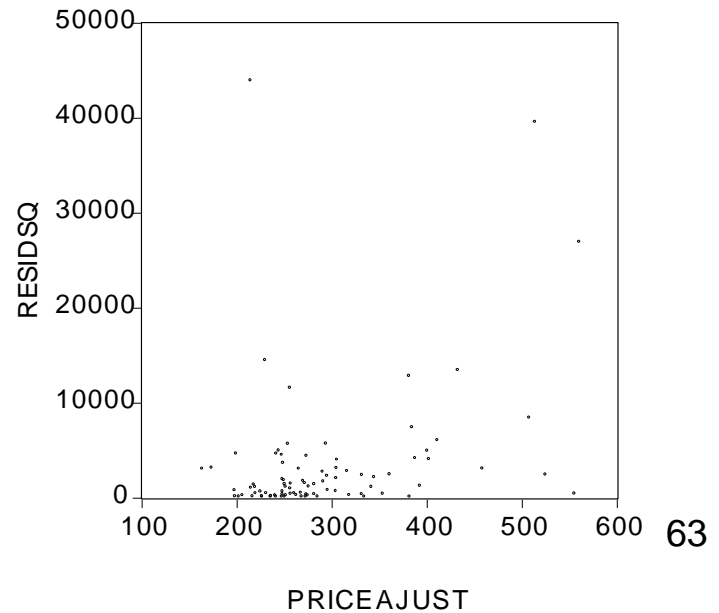
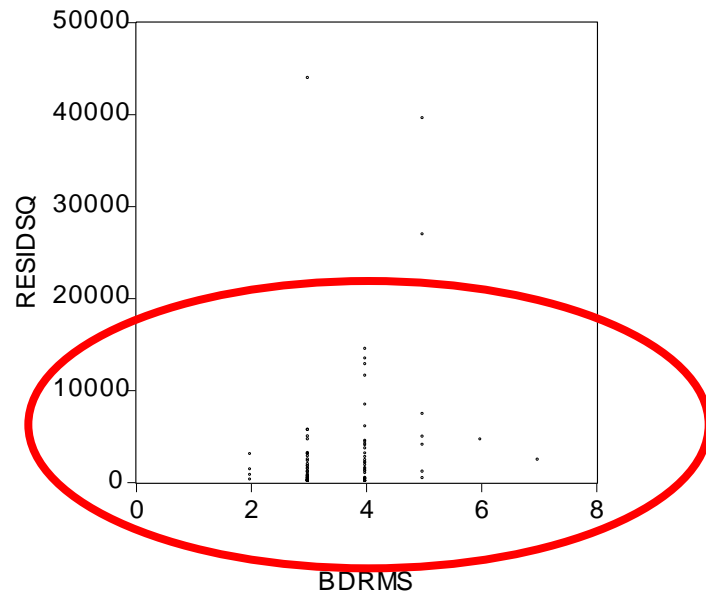
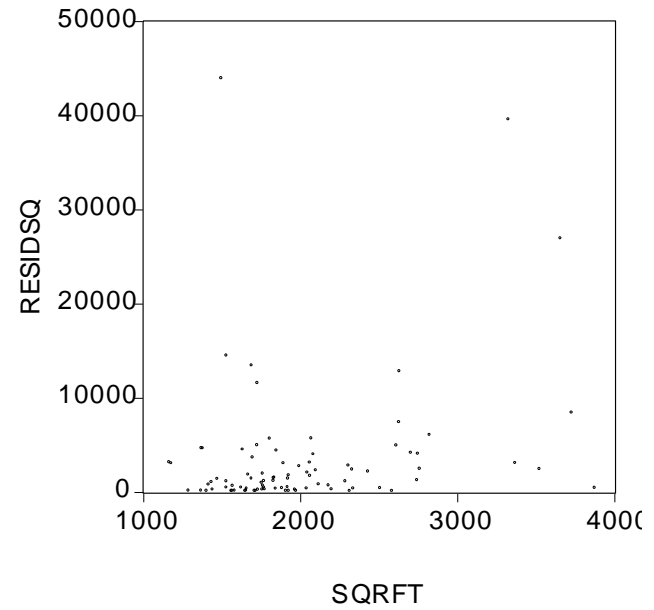
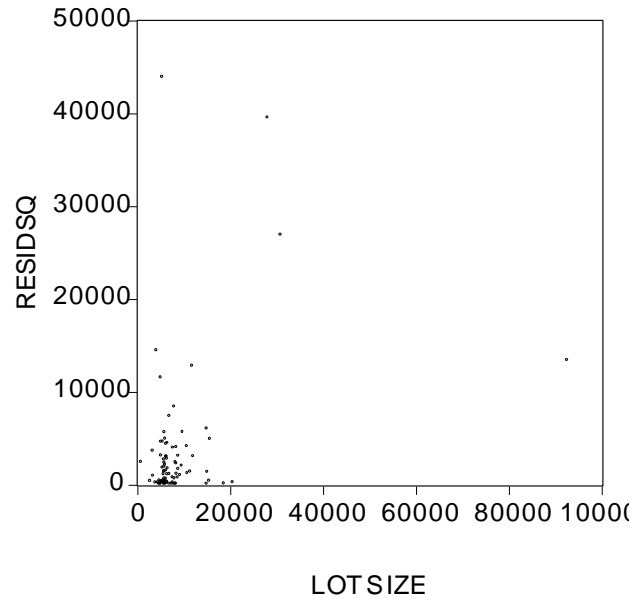


BDRMS



PRICEAJUST

ii) Análise completa dos resíduos e comentários.



LEITURA COMPLEMENTAR II

ESTIMADOR DE MÍNIMOS QUADRADOS GENERALIZADOS (GLS)

ESTIMAÇÃO VIA MÍNIMOS QUADRADOS GENERALIZADOS

Como já observamos, as técnicas inferenciais são componentes importantes em muitas análises de dados. Entretanto, na presença de heterocedasticidade, como discutido nos *slides* iniciais, toda a análise baseada em testes de hipóteses se torna inválida.

ESTIMAÇÃO VIA MÍNIMOS QUADRADOS GENERALIZADOS

Nos próximos slides estaremos interessados em ajustar os erros-padrão e as estatísticas t e F , obtidos por mínimos quadrados, na presença de heterocedasticidade cuja forma é conhecida.

MÍNIMOS QUADRADOS GENERALIZADOS

Pode-se demonstrar que no caso em que a **matriz Ω é conhecida**, os parâmetros do modelo de regressão podem ser estimados a partir de um método chamado de **Mínimos Quadrados Generalizados (GLS)**.

MÍNIMOS QUADRADOS GENERALIZADOS

O objetivo do método é estimar os parâmetros do modelo, assim como fazíamos com OLS. Entretanto, agora precisamos levar em consideração a nova configuração da matriz de variâncias e covariâncias dos erros. Ou seja, aqui, será necessário encontrar uma forma adequada de transformar o modelo, de tal sorte a obtermos um termo de erro homocedástico.

MÍNIMOS QUADRADOS GENERALIZADOS

Com esse objetivo em mente, tomaremos uma matriz Ψ , não singular, que chamaremos de matriz de transformação, e faremos a seguinte operação no modelo original

$$\underset{\sim}{\Psi} \underset{\sim}{\mathbf{y}} = \underset{\sim}{\Psi} \underset{\sim}{\mathbf{X}} \underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\Psi} \underset{\sim}{\boldsymbol{\varepsilon}}, \quad (1)$$

MÍNIMOS QUADRADOS GENERALIZADOS

Dessa transformação, temos os seguintes resultados:

$$E(\underset{\sim}{\Psi} \underset{\sim}{\boldsymbol{\varepsilon}}) = \underset{\sim}{\Psi} E(\underset{\sim}{\boldsymbol{\varepsilon}}) = \underset{\sim}{\mathbf{0}},$$

e

$$Var(\underset{\sim}{\Psi} \underset{\sim}{\boldsymbol{\varepsilon}}) = \underset{\sim}{\Psi} E(\underset{\sim}{\boldsymbol{\varepsilon}} \underset{\sim}{\boldsymbol{\varepsilon}}') \underset{\sim}{\Psi}' = \sigma^2 \underset{\sim}{\Psi} \underset{\sim}{\boldsymbol{\Omega}} \underset{\sim}{\Psi}',$$

MÍNIMOS QUADRADOS GENERALIZADOS

Se a matriz Ψ for construída de maneira tal que

$$\underset{\sim}{\Psi} \underset{\sim}{\Omega} \underset{\sim}{\Psi}' = \underset{\sim}{\mathbf{I}}$$

Então, será possível aplicar OLS em (1), pois

$$\underset{\sim}{Var}(\underset{\sim}{\Psi} \underset{\sim}{\boldsymbol{\varepsilon}}) = \underset{\sim}{\Psi} \underset{\sim}{E}(\underset{\sim}{\boldsymbol{\varepsilon}} \underset{\sim}{\boldsymbol{\varepsilon}}') \underset{\sim}{\Psi}' = \sigma^2 \underset{\sim}{\Psi} \underset{\sim}{\Omega} \underset{\sim}{\Psi}' = \sigma^2 \underset{\sim}{\mathbf{I}}.$$

MÍNIMOS QUADRADOS GENERALIZADOS

Aplicando OLS em (1), vem que:

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

Mas,

$$\Psi \Omega \Psi' = \mathbf{I}$$

pode ser reescrito como,

$$\Omega^{-1} = \Psi' \Psi$$

MÍNIMOS QUADRADOS GENERALIZADOS

Obtendo, assim

$$\hat{\beta}_{\sim}^{(GLS)} = \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\mathbf{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\mathbf{\Omega}}^{-1} \underset{\sim}{\mathbf{y}} \quad (2)$$

Ainda, não é difícil mostrar que, sob certas suposições,

$$E \left[\hat{\beta}_{\sim}^{(GLS)} \right] = \underset{\sim}{\beta}$$

MÍNIMOS QUADRADOS GENERALIZADOS

Substituindo (1) em

$$\hat{\beta}_{\sim}^{(GLS)} = \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\Psi}' \underset{\sim}{\Psi} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\Psi}' \underset{\sim}{\Psi} \underset{\sim}{\mathbf{y}}$$

vem que

$$\begin{aligned} \hat{\beta}_{\sim}^{(GLS)} &= \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\Psi}' \underset{\sim}{\Psi} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\Psi}' \left(\underset{\sim}{\Psi} \underset{\sim}{\mathbf{X}} \underset{\sim}{\beta} + \underset{\sim}{\Psi} \underset{\sim}{\boldsymbol{\varepsilon}} \right) = \\ &= \underset{\sim}{\beta} + \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\Psi}' \underset{\sim}{\Psi} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\Psi}' \underset{\sim}{\Psi} \underset{\sim}{\boldsymbol{\varepsilon}} = \underset{\sim}{\beta} + \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\boldsymbol{\varepsilon}} \end{aligned}$$

MÍNIMOS QUADRADOS GENERALIZADOS

Logo,

$$\begin{aligned} \text{Var} \left[\hat{\beta}^{(GLS)} \right] &= \text{Var} \left[\left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\boldsymbol{\varepsilon}} \right] = \\ &= \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \text{Var} \left[\underset{\sim}{\boldsymbol{\varepsilon}} \right] \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1} = \\ &= \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1} \underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \sigma^2 \underset{\sim}{\boldsymbol{\Omega}} \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1} = \\ &= \sigma^2 \left(\underset{\sim}{\mathbf{X}}' \underset{\sim}{\boldsymbol{\Omega}}^{-1} \underset{\sim}{\mathbf{X}} \right)^{-1}. \end{aligned}$$

MÍNIMOS QUADRADOS GENERALIZADOS

Observações

- (i) É possível demonstrar que o vetor de estimadores, via GLS, é o melhor estimador linear não tendencioso (BLUE) para o vetor de parâmetros do modelo de interesse.
- (ii) Quando os erros são homocedásticos, não é difícil ver que a matriz Ω se reduz à matriz identidade, o que nos leva aos resultados de OLS.

MÍNIMOS QUADRADOS GENERALIZADOS

Observações (cont.)

(iii) Em muitos casos, na prática, σ^2 e Ω são desconhecidos. Assim sendo, uma das soluções é usar o método dos mínimos quadrados generalizados factíveis (FGLS). Aqui, inicialmente estimamos os parâmetros do modelo de interesse, por OLS, desconsiderando o problema da heterocedasticidade, geramos os resíduos e construímos a matriz de variâncias e covariâncias consistente estimada dos termos de erros.

MÍNIMOS QUADRADOS GENERALIZADOS

Observações (cont.)

(iv) É possível demonstrar que os estimadores de FGLS são consistentes para os estimadores de GLS). Aqui, inicialmente estimamos os parâmetros do modelo de interesse, por OLS, desconsiderando o problema da heterocedasticidade, geramos os resíduos e construímos a matriz de variâncias e covariâncias estimada dos termos de erros.

MÍNIMOS QUADRADOS GENERALIZADOS

Observações (cont.)

- (v) É importante notar que, para a estimação dos parâmetros do modelo de interesse, é necessário encontrar a matriz Ψ , tal que $\Psi'\Psi = \Omega^{-1}$. Feito isso, basta transformar o modelo original conforme (1).

MÍNIMOS QUADRADOS GENERALIZADOS

Observações (cont.)

(vi) O vetor de estimadores para β , via GLS, pode ser obtido de maneira análoga àquela obtida nos slides anteriores, minimizando a quantidade

$$S\left(\underset{\sim}{\beta}\right) = \underset{\sim}{\varepsilon}' \underset{\sim}{\Omega}^{-1} \underset{\sim}{\varepsilon} = \left(\underset{\sim}{y} - \underset{\sim}{X} \underset{\sim}{\beta}\right)' \underset{\sim}{\Omega}^{-1} \left(\underset{\sim}{y} - \underset{\sim}{X} \underset{\sim}{\beta}\right)$$

Decomposição de Cholesky

Suponha que $\Sigma = \sigma^2\Omega$ seja uma matriz de variâncias e covariâncias associada a um vetor de v.a. qualquer. Como Σ é uma matriz simétrica e positiva definida, existe uma matriz triangular inferior B , com elementos na diagonal principal iguais a 1, e uma matriz diagonal D , cujos elementos da diagonal são todos positivos, tal que

$$\Sigma = BDB'.$$

Decomposição de Cholesky

Usando os resultados anteriores, é fácil ver que

$$\Sigma = TT'$$

em que

$$T = BD^{1/2}.$$

De modo análogo, a matriz Σ^{-1} pode ser decomposta como $T'T$ usando

$$T = D^{-1/2}B'.$$

LEITURA COMPLEMENTAR III

ESTIMADOR DE MÍNIMOS QUADRADOS PONDERADOS (WLS)

MÍNIMOS QUADRADOS PONDERADOS

Considere o modelo

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon. \quad (1)$$

Seja

$$\mathbf{x} = (x_2, \dots, x_k)$$

um vetor que denota todas as variáveis explicativas da equação anterior e assumamos que

$$\text{Var}(\varepsilon|\mathbf{x}) = \sigma^2 h(\mathbf{x}). \quad (2)$$

MÍNIMOS QUADRADOS PONDERADOS

em que

$h(x)$ é alguma função das variáveis explicativas que determina a heterocedasticidade.

MÍNIMOS QUADRADOS PONDERADOS

Observação:

- 1) Como a variância retorna sempre um valor positivo, então, $h(\mathbf{x}) > 0$.
- 2) O desvio padrão de ε , condicionado a \mathbf{x} , é

$$DP(\varepsilon | \mathbf{x}) = \sigma \sqrt{h(\mathbf{x})}$$

Pergunta:

Como poderemos utilizar (2) para estimarmos $\hat{\beta}_j$?

MÍNIMOS QUADRADOS PONDERADOS

IDEIA!!

Seja o modelo dado em (1), que contém erro heterocedástico; vamos transformar tal modelo, num modelo com erro homocedástico.

MÍNIMOS QUADRADOS PONDERADOS

Como

$$\begin{aligned} \text{Var}(\varepsilon \mid x_2, \dots, x_k) &= E(\varepsilon^2 \mid x_2, \dots, x_k) = \\ &= \sigma^2 h(x_2, \dots, x_k) \end{aligned}$$

a variância de

$$\frac{\varepsilon}{\sqrt{h(x_2, \dots, x_k)}}$$

(condicionada a \mathbf{x}), é σ^2 :

MÍNIMOS CUADRADOS PONDERADOS

Como

$$\begin{aligned} \text{Var} \left[\left(\frac{\varepsilon}{\sqrt{h(x_2, \dots, x_k)}} \right) \middle| (x_2, \dots, x_k) \right] &= \frac{1}{\left(\sqrt{h(x_2, \dots, x_k)} \right)^2} \text{Var} [\varepsilon | (x_2, \dots, x_k)] = \\ &= \frac{1}{h(x_2, \dots, x_k)} E [\varepsilon^2 | (x_2, \dots, x_k)] = \frac{1}{h(x_2, \dots, x_k)} \cdot \sigma^2 h(x_2, \dots, x_k) = \sigma^2 \end{aligned}$$

MÍNIMOS CUADRADOS PONDERADOS

Logo, dividiendo (1) por

$$\sqrt{h(x_2, \dots, x_k)}, \quad i = 1, 2, \dots, n,$$

teremos

$$\begin{aligned} \frac{y}{\sqrt{h(x_2, \dots, x_k)}} &= \beta_1 \cdot \frac{1}{\sqrt{h(x_2, \dots, x_k)}} + \beta_2 \cdot \frac{x_2}{\sqrt{h(x_2, \dots, x_k)}} + \\ &+ \dots + \beta_k \cdot \frac{x_k}{\sqrt{h(x_2, \dots, x_k)}} + \frac{\varepsilon}{\sqrt{h(x_2, \dots, x_k)}} \end{aligned}$$

MÍNIMOS QUADRADOS PONDERADOS

ou,

$$y^* = \beta_1 x_1^* + \beta_2 x_2^* + \dots + \beta_k x_k^* + \varepsilon^*$$

Observação:

- 1) A interpretação dos parâmetros do modelo deve ser feita a partir de (1).
- 2) Este último modelo satisfaz às suposições usuais, se o modelo (1) também as satisfizer, exceto pela de homocedasticidade dos erros.

VOLTANDO AO EXEMPLO

Voltando ao arquivo *HPRICE1.wf1*. Estime os parâmetros do modelo

$$price = \beta_1 + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + \varepsilon$$

via WLS e realize um teste adequado para verificar se os erros se tornaram homocedásticos.

VOLTANDO AO EXEMPLO

Utilizando os dados do arquivo *HPRICE1.wf1*.

Estime os parâmetros do modelo

$$\log(\text{price}) = \beta_1 + \beta_2 \log(\text{lotsize}) + \beta_3 \log(\text{sqrft}) + \beta_4 \text{bdrms} + \varepsilon$$

e realize um teste adequado para verificar se os erros são homocedásticos. Ainda, faça um teste que analise a especificação da forma funcional.

Comente os resultados.

VOLTANDO AO EXEMPLO

Voltando ao arquivo *HPRICE1.wf1*. Para o modelo

$$price = \beta_1 + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + \varepsilon$$

proponha alterações, de tal sorte a capturar, de maneira mais adequada, a informação sobre o preço de determinadas residências cujos valores das variáveis (resposta e explicativas) podem ser considerados como discrepantes. Ainda, faça uma análise completa dos resíduos. Comente.

EXERCÍCIO

Considere o modelo

$$y_i = \alpha + \beta \cdot x_i + u_i$$

com $E(u_i) = 0$ e $\text{Var}(u_i | x_i) = \sigma^2 x_i^2$.

Dados os valores de X e Y para uma dada amostra de cinco observações

X	1	2	4	4	6
Y	5,0	7,5	10,0	11,5	15,0

- Obtenha os estimadores lineares não viciados de variância mínima de α e β .
- Teste a hipótese $H_0: \beta = 1$ contra $H_A: \beta > 1$.

MÍNIMOS QUADRADOS PONDERADOS FACTÍVEIS (FWLS)

Caso os erros do modelo de regressão sejam heterocedásticos, o caso mais comum é aquele em que a forma da heterocedasticidade é desconhecida. Sendo assim, precisamos encontrar uma estimativa para a quantidade $h(\mathbf{x})$, que seja gerada a partir de um estimador com boas propriedades.

MÍNIMOS QUADRADOS PONDERADOS FACTÍVEIS (FWLS)

Em Wooldridge (2003), temos uma sugestão bastante geral para a quantidade anterior, que é dada por

$$\text{Var}(\varepsilon_i/\mathbf{x}) = \sigma^2 \exp(\delta_1 + \delta_2 x_2 + \dots + \delta_k x_k).$$

Note que, sob tal proposta, $h(\mathbf{x}) > 0$, o que é uma vantagem da proposta.

MÍNIMOS QUADRADOS PONDERADOS FACTÍVEIS (FWLS)

De resultados anteriores, temos que

$$\varepsilon_i^2 = \sigma^2 \exp(\delta_1 + \delta_2 x_2 + \dots + \delta_k x_k) v$$

em que,

$$E(v/x) = 1.$$

Então, se $E(v) = 1$,

$$\ln(\varepsilon_i^2) = \alpha_1 + \delta_2 x_2 + \dots + \delta_k x_k + u \quad [1]$$

MÍNIMOS QUADRADOS PONDERADOS FACTÍVEIS (FWLS)

em que,

$$E(u) = 1;$$

u é independente de x .

Como sabemos, as estimativas para ε_i^2 podem ser obtidas diretamente dos quadrados dos resíduos do modelo de interesse, cujos parâmetros foram estimados por OLS.

MÍNIMOS QUADRADOS PONDERADOS FACTÍVEIS (FWLS)

Assim, ao estimarmos os parâmetros de [1], conseguiremos uma estimativa para a função $h(\cdot)$, que é obtida de

$$\hat{h} = \exp(\hat{g}), \quad [2]$$

em que

$$g = \ln(\varepsilon^2).$$

Finalmente, usamos [2] para encontrar as estimativas de WLS e os respectivos erros padrões do modelo de interesse.

VOLTANDO AO EXEMPLO

Voltando ao arquivo *HPRICE1.wf1*. Estime os parâmetros do modelo

$$price = \beta_1 + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + \varepsilon$$

via proposta discutida anteriormente e realize um teste adequado para verificar se os erros se tornaram homocedásticos.