

ponderados como um meio de obter estimadores BLUE. As estatísticas de testes da estimação MQP são perfeitamente válidas quando o termo de erro é normalmente distribuído ou assintoticamente válido sob não normalidade. Isso supõe, é claro, que temos o modelo de heteroscedasticidade apropriado.

Mais comumente, devemos estimar um modelo quanto à heteroscedasticidade antes de aplicarmos o MQP. O estimador MQG *factível* resultante não mais será não viesado, mas será consistente e assintoticamente eficiente. As estatísticas habituais da regressão MQP são assintoticamente válidas. Discutimos um método para assegurar que as variâncias estimadas sejam estritamente positivas para todas as observações, necessário para aplicar o método MQP.

Como discutimos no Capítulo 7, o modelo de probabilidade linear de uma variável dependente binária terá, necessariamente, um termo de erro heteroscedástico. Uma maneira simples de lidar com esse problema é calcular estatísticas robustas em relação à heteroscedasticidade. Alternativamente, se todos os valores ajustados (isto é, as probabilidades estimadas) estiverem estritamente entre zero e um, os mínimos quadrados ponderados poderão ser usados para a obtenção de estimadores assintoticamente eficientes.

## Termos-chave

Erro padrão robusto em relação à heteroscedasticidade	Estimador MQG factível (MQGF)	Heteroscedasticidade da forma desconhecida
Estatística $F$ robusta em relação à heteroscedasticidade	Estimadores de mínimos quadrados generalizados (MQG)	Teste de Breusch-Pagan da heteroscedasticidade (Teste BP)
Estatística $LM$ robusta em relação à heteroscedasticidade	Estimadores de mínimos quadrados ponderados (MQP)	Teste de White para heteroscedasticidade
Estatística $t$ robusta em relação à heteroscedasticidade		

## Problemas

- Quais das seguintes alternativas são consequências da heteroscedasticidade?
  - Os estimadores MQO,  $\hat{\beta}_j$ , são inconsistentes.
  - A estatística  $F$  usual não mais tem uma distribuição  $F$ .
  - Os estimadores MQO não são mais BLUE.
- Considere um modelo linear para explicar o consumo mensal de cerveja:

$$beer = \beta_0 + \beta_1 inc + \beta_2 price + \beta_3 educ + \beta_4 female + u$$

$$E(u|inc, price, educ, female) = 0$$

$$\text{Var}(u|inc, price, educ, female) = \sigma^2 inc^2.$$

Escreva a equação transformada que tenha um termo de erro homoscedástico.

- Verdadeiro ou Falso: O método MQP é preferido ao MQO quando uma variável importante for omitida do modelo.
- Utilizando os dados contidos no arquivo GPA3, a seguinte equação foi estimada para os alunos de uma universidade:

$$\widehat{trmgpa} = -2,12 + 0,900 \text{ crsgpa} + 0,193 \text{ cumgpa} + 0,0014 \text{ tothrs} \\
\begin{array}{cccc}
(0,55) & (0,175) & (0,064) & (0,0012) \\
[0,55] & [0,166] & [0,074] & [0,0012] \\
+ 0,0018 \text{ sat} - 0,0039 \text{ hsperc} + 0,351 \text{ female} - 0,157 \text{ season} \\
(0,0002) & (0,0018) & (0,085) & (0,098) \\
[0,0002] & [0,0019] & [0,079] & [0,080]
\end{array} \\
n = 269, R^2 = 0,465.$$

Aqui, *trmgpa* é a nota obtida pelo aluno no exame final do curso, no semestre corrente; *crsgpa* é uma média ponderada das notas nas diversas disciplinas cursadas no semestre; *cumgpa* é a nota do exame de final de semestre, no semestre anterior; *tothrs* é o total de créditos em horas, acumuladas até o semestre anterior; *sat* é a nota do aluno no exame de ingresso na Universidade; *hsperc* é o percentil do aluno no curso médio na escola em que o aluno se formou antes de ingressar na Universidade; *female* é uma *dummy* de gênero; e *season* é uma *dummy* igual a um se o esporte praticado pelo aluno for praticado durante o outono. Os erros padrão usuais e os robustos em relação à heteroscedasticidade estão registrados entre parênteses e colchetes, respectivamente.

- (i) As variáveis *crsgpa*, *cumgpa* e *tothrs* têm os efeitos estimados esperados? Quais dessas variáveis são estatisticamente significantes ao nível de 5%? Importa quais erros padrão são usados?
  - (ii) Por que a hipótese  $H_0: \beta_{crsgpa} = 1$  faz sentido? Teste essa hipótese contra a alternativa bicaudal no nível de 5%, usando ambos os erros padrão. Descreva suas conclusões.
  - (iii) Verifique se existe um efeito sazonal sobre a variável *trmgpa*, usando ambos os erros padrão. O nível de significância no qual a hipótese nula pode ser rejeitada depende do erro padrão usado?
5. A variável *smokes* é uma variável binária igual a um se a pessoa fuma, e zero caso contrário. Utilizando os dados contidos no arquivo SMOKE, estimamos um modelo de probabilidade linear de *smokes*:

$$\widehat{smokes} = 0,656 - 0,069 \log(\text{cigpric}) + 0,012 \log(\text{income}) - 0,029 \text{ educ} \\
\begin{array}{cccc}
(0,855) & (0,204) & (0,026) & (0,006) \\
[0,856] & [0,207] & [0,026] & [0,006] \\
+ 0,020 \text{ age} - 0,00026 \text{ age}^2 - 0,101 \text{ restaurn} - 0,026 \text{ white} \\
(0,006) & (0,00006) & (0,039) & (0,052) \\
[0,005] & [0,00006] & [0,038] & [0,050]
\end{array} \\
n = 807, R^2 = 0,062.$$

A variável *white* é igual a um se a pessoa for branca, e zero caso contrário; as outras variáveis independentes foram definidas no Exemplo 8.7. Tanto os erros padrão usuais como os robustos em relação à heteroscedasticidade estão informados.

- (i) Existe alguma diferença importante entre os dois conjuntos de erros padrão?
- (ii) Mantendo os outros fatores fixos, se a educação aumentar em quatro anos, o que acontece com a probabilidade de fumar estimada?
- (iii) Em que ponto mais um ano de idade reduz a probabilidade de fumar?

- (iv) Interprete o coeficiente da variável binária *restaurn* (uma variável *dummy* igual a um se a pessoa viver em um estado em que há restrições de fumar em restaurantes).
- (v) A pessoa número 206 no conjunto de dados tem as seguintes características: *cigpric* = 67,44, *income* = 6.500, *educ* = 16, *age* = 77, *restaurn* = 0, *white* = 0 e *smokes* = 0. Calcule a probabilidade de fumar dessa pessoa e comente o resultado.
6. Existem diferentes maneiras de se combinar recursos dos testes de Breusch-Pagan e White de heteroscedasticidade. Uma possibilidade não tratada no texto é executar a regressão

$$\hat{u}_i^2 \text{ sobre } x_{i1}, x_{i2}, \dots, x_{ik}, \hat{y}_i^2, i = 1, \dots, n,$$

em que  $\hat{u}_i$  refere-se aos resíduos dos MQO e  $\hat{y}_i$  aos valores ajustados dos MQO. Depois, testamos significância conjunta das  $x_{i1}, x_{i2}, \dots, x_{ik}$ , e  $\hat{y}_i^2$ . (É claro, sempre incluímos um intercepto nessa regressão.)

- (i) Quais são os *gl* associados com o teste *F* de heteroscedasticidade proposto?
- (ii) Explique por que o *R*-quadrado da regressão anterior sempre será, pelo menos, tão grande quanto os *R*-quadrados da regressão BP e os do caso especial do teste de White.
- (iii) O item (ii) conclui que o novo teste sempre produz um *p*-valor menor do que ou a BP ou do caso especial da estatística de White? Explique.
- (iv) Suponha que alguém também sugira incluir  $\hat{y}_i$  no teste recém-proposto. Qual sua opinião sobre essa ideia?
7. Considere um modelo em nível de empregado,

$$y_{i,e} = \beta_0 + \beta_1 x_{i,e,1} + \beta_2 x_{i,e,2} + \dots + \beta_k x_{i,e,k} + f_i + v_{i,e},$$

em que a variável não observada  $f_i$  é um “efeito da empresa” em cada empregado em determinada empresa  $i$ . O termo de erro  $v_{i,e}$  é específico do empregado  $e$  da empresa  $i$ . O erro de combinação é  $u_{i,e} = f_i + v_{i,e}$ , como na equação (8.28).

- (i) Suponha que  $\text{Var}(f_i) = \sigma_f^2$ ,  $\text{Var}(v_{i,e}) = \sigma_v^2$ , e  $f_i$  e  $v_{i,e}$  são não correlacionadas. Demonstre que  $\text{Var}(u_{i,e}) = \sigma_f^2 + \sigma_v^2$ ; chame-a  $\sigma^2$ .
- (ii) Agora suponha que para  $e \neq g$ ,  $v_{i,e}$  e  $v_{i,g}$  são não correlacionadas. Demonstre que  $\text{Cov}(u_{i,e}, u_{i,g}) = \sigma_f^2$ .
- (iii) Que  $\bar{u}_i = m_i^{-1} \sum_{e=1}^{m_i} u_{i,e}$  seja a média dos erros de combinação em determinada empresa. Demonstre que  $\text{Var}(\bar{u}_i) = \sigma_f^2 + \sigma_v^2/m_i$ .
- (iv) Detalhe a relevância do item (iii) para a estimação MQP usando dados nivelados pela média no nível da empresa, em que o peso usado para a observação  $i$  é o tamanho usual da empresa.
8. As equações a seguir foram estimadas usando os dados do arquivo ECONMATH. A primeira equação é para homens e a segunda trata de mulheres. A terceira e a quarta combinam homens e mulheres.

$$\widehat{\text{score}} = 20,52 + 13,60 \text{ colgpa} + 0,670 \text{ act}$$

(3,72)    (0,94)    (0,150)

$$n = 406, R^2 = 0,4025, \text{SQR} = 38,781,38.$$

$$\widehat{\text{score}} = 13,79 + 11,89 \text{ colgpa} + 1,03 \text{ act}$$

(4,11)    (1,09)    (0,18)

$$n = 408, R^2 = 0,3666, \text{SQR} = 48,029,82.$$

$$\widehat{score} = 15,60 + 3,17 \text{ male} + 12,82 \text{ colgpa} + 0,838 \text{ act}$$

$$(2,80) \quad (0,73) \quad (0,72) \quad (0,116)$$

$$n = 814, R^2 = 0,3946, SSR = 87.128,96.$$

$$\widehat{score} = 13,79 + 6,73 \text{ male} + 11,89 \text{ colgpa} + 1,03 \text{ act} + 1,72 \text{ male} \cdot \text{colgpa} - 0,364 \text{ male} \cdot \text{act}$$

$$(3,91) \quad (5,55) \quad (1,04) \quad (0,17) \quad (1,44) \quad (0,232)$$

$$n = 814, R^2 = 0,3968, SQR = 86.811,20.$$

- (i) Calcule a estatística de Chow usual para testar a hipótese nula de que as equações de regressão são as mesmas para homens e mulheres. Encontre o  $p$ -valor do teste.
- (ii) Calcule a estatística de Chow usual para testar a hipótese nula de que os coeficientes de inclinação são os mesmos para homens e mulheres e registre o  $p$ -valor.
- (iii) Você tem informações suficientes para calcular as versões robustas em relação à heteroscedasticidade dos testes dos itens (i) e (ii)? Explique.

## Exercícios em computador

**C1** Considere o modelo a seguir para explicar o comportamento do sono:

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + \beta_2 \text{educ} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \beta_5 \text{yngkid} + \beta_6 \text{male} + u.$$

- (i) Escreva um modelo que permita que a variância de  $u$  difira entre homens e mulheres. A variância não deve depender de outros fatores.
  - (ii) Use os dados do arquivo SLEEP75 para estimar os parâmetros do modelo para heteroscedasticidade. (Você precisa estimar a equação *sleep* por MQO primeiro para obter os resíduos MQO.) A variância estimada de  $u$  é maior para homens ou para mulheres?
  - (iii) A variância de  $u$  é estatisticamente diferente para homens e mulheres?
- C2** (i) Use os dados do arquivo HPRICE1 para obter erros padrão robustos em relação à heteroscedasticidade para a equação (8.17). Discuta as diferenças importantes em relação aos erros padrão usuais.
- (ii) Repita o item (i) para a equação (8.18).
- (iii) O que esse exemplo sugere a respeito da heteroscedasticidade e da transformação usada para a variável dependente?
- C3** Aplique o teste de White completo para heteroscedasticidade [ver equação (8.19)] na equação (8.18). Usando o qui-quadrado da estatística, obtenha o  $p$ -valor. O que você conclui?
- C4** Use o arquivo VOTE1 para este exercício.
- (i) Estime o modelo com *voteA* como a variável dependente e *prtystrA*, *democA*,  $\log(\text{expendA})$  e  $\log(\text{expendB})$  como variáveis independentes. Obtenha os resíduos de MQO,  $\hat{u}_i$ , e regreda os valores sobre todas as variáveis independentes. Explique por que você obtém  $R^2 = 0$ .
  - (ii) Agora, calcule o teste de Breusch-Pagan para heteroscedasticidade. Use a versão da estatística  $F$  e registre o  $p$ -valor.
  - (iii) Calcule o caso especial do teste de White para heteroscedasticidade, novamente usando a forma da estatística  $F$ . Quão forte é a evidência de heteroscedasticidade agora?
- C5** Use os dados do arquivo PNTSPRD para este exercício.

- (i) A variável *sprdcvr* é uma variável binária igual a um se o *point spread* de Las Vegas em um jogo de basquete universitário for coberto. O valor esperado de *sprdcvr*, digamos  $\mu$ , é a probabilidade de que o *spread* seja coberto em um jogo aleatoriamente selecionado. Teste  $H_0: \mu = 0,5$  contra  $H_1: \mu \neq 0,5$  ao nível de 10% de significância e discuta seus achados. (Dica: É mais fácil usar um teste *t* regressando *sprdcvr* sobre apenas um intercepto.)
- (ii) Quantas partidas na amostra de 553 foram jogadas em uma quadra neutra?
- (iii) Estime o modelo linear de probabilidade

$$sprdcvr = \beta_0 + \beta_1 favhome + \beta_2 neutral + \beta_3 fav25 + \beta_4 und25 + u$$

e registre os resultados na forma usual. (Registre os erros padrão de MQO usuais e os erros padrão robustos em relação à heteroscedasticidade.) Qual variável é mais significativa, dos pontos de vista prático e estatístico?

- (iv) Explique por que, sob a hipótese nula  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ , não há heteroscedasticidade no modelo.
- (v) Use a estatística *F* usual para testar a hipótese do item (iv). O que você conclui?
- (vi) Dada a análise anterior, você diria que é possível prever sistematicamente se o *spread* de Las Vegas será coberto usando as informações disponíveis antes da partida?

**C6** No Exemplo 7.12, estimamos um modelo linear de probabilidade para saber se um homem jovem foi preso durante 1986:

$$arr86 = \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 tottime + \beta_4 ptime86 + \beta_5 qemp86 + u$$

- (i) Usando os dados do arquivo CRIME1, estime o modelo por MQO e verifique se todos os valores ajustados estão exatamente entre zero e um. Quais são os menores e os maiores valores ajustados?
- (ii) Estime a equação por mínimos quadrados ponderados, conforme discutido na Seção 8.5.
- (iii) Use os estimadores de MQP para determinar se *tottime* e *avgsen* são conjuntamente significantes a um nível de 5%.

**C7** Use os dados do arquivo LOANAPP para este exercício.

- (i) Estime a equação do item (iii) do Exercício em computador C8, do Capítulo 7, calculando os erros padrão robustos em relação à heteroscedasticidade. Compare o intervalo de confiança de 95% em  $\beta_{white}$  com o intervalo de confiança não robusto.
- (ii) Obtenha os valores ajustados da regressão do item (i). Algum deles é menor do que zero? Algum deles é maior do que um? O que isso quer dizer a respeito de aplicar mínimos quadrados ponderados?

**C8** Use o conjunto de dados do arquivo GPA1 para este exercício.

- (i) Use MQO para estimar um modelo que relacione *colGPA* a *hsGPA*, *ACT*, *skipped* e *PC*. Obtenha os resíduos de MQO.
- (ii) Calcule o caso especial do teste de White para heteroscedasticidade. Na regressão de  $\hat{u}_i^2$  sobre  $\overline{colGPA}_i$ ,  $\overline{colGPA}_i^2$ , obtenha os valores ajustados, digamos  $\hat{h}_i$ .
- (iii) Verifique se os valores ajustados do item (ii) são todos estritamente positivos. Em seguida, obtenha os estimadores de mínimos quadrados ponderados usando pesos  $1/\hat{h}_i$ . Compare os estimadores de mínimos quadrados ponderados para o efeito de faltar em aulas e o efeito de possuir PCs com os estimadores de MQO correspondentes. E em relação à sua significância estatística?

- (iv) Na estimação de MQP do item (iii), obtenha os erros padrão robustos em relação à heteroscedasticidade. Em outras palavras, permita o fato de que a função da variância estimada no item (ii) possa ser mal-especificada. (Ver Questão 8.4.) Os erros padrão mudam muito no item (iii)?
- C9** No Exemplo 8.7, calculamos os MQO e um conjunto de estimadores de MQP em uma equação de demanda por cigarros.
- Obtenha os estimadores de MQO na equação (8.35).
  - Obtenha o  $\hat{h}_i$  usado na estimação dos MQP da equação (8.36) e reproduza essa equação. A partir dela, obtenha os resíduos *não ponderados* e valores ajustados; chame-os de  $\hat{u}_i$  e  $\hat{y}_i$ , respectivamente. (Por exemplo, em Stata, os resíduos não ponderados e os valores ajustados são apresentados por predefinição.)
  - Defina  $\check{u}_i = \hat{u}_i/\sqrt{\hat{h}_i}$  e  $\check{y}_i = \hat{y}_i/\sqrt{\hat{h}_i}$  como as quantias ponderadas. Execute o caso especial do teste de White para heteroscedasticidade regredindo  $\check{u}_i^2$  sobre  $\check{y}_i$ ,  $\check{y}_i^2$ , certificando-se de incluir um intercepto, como sempre. Você encontrou heteroscedasticidade nos resíduos ponderados?
  - O que os achados do item (iii) indicam sobre a forma proposta da heteroscedasticidade usada para obter a equação (8.36)?
  - Obtenha os erros padrão válidos para os estimadores de MQP que permitem que a função de variância seja mal-especificada.
- C10** Use o conjunto de dados 401KSUBS neste exercício.
- Usando MQO, estime um modelo linear de probabilidade para  $e401k$ , usando como variável explicativa  $inc$ ,  $inc^2$ ,  $age$ ,  $age^2$  e  $male$ . Obtenha os erros padrão usuais dos MQO e as versões robustas em relação à heteroscedasticidade. Existem diferenças importantes?
  - No caso especial do teste de White para heteroscedasticidade, onde regredimos os resíduos quadrados de MQO sobre uma transformação quadrática nos valores ajustados de MQO,  $\hat{u}_i^2$  sobre  $\check{y}_i$ ,  $\check{y}_i^2$ ,  $i = 1, \dots, n$ , afirme que o limite de probabilidade do coeficiente de  $\hat{y}_i$  deve ser um, o limite da probabilidade do coeficiente de  $\hat{y}_i^2$  deve ser  $-1$  e o limite da probabilidade do intercepto deve ser zero. {Dica: Lembre-se que  $\text{Var}(y|x_1, \dots, x_k) = p(\mathbf{x})[1 - p(\mathbf{x})]$ , em que  $p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ .}
  - Do modelo estimado no item (i), obtenha o teste de White e veja se as estimativas de coeficiente correspondem aproximadamente aos valores teóricos descritos no item (ii).
  - Depois de verificar que os valores ajustados do item (i) estão todos entre zero e um, obtenha os estimadores de mínimos quadrados ponderados do modelo linear de probabilidade. Eles diferem de forma importante dos estimadores de MQO?
- C11** Use os dados do arquivo 401KSUBS para esta questão, restringindo a amostra a  $fsize = 1$ .
- Ao modelo estimado na Tabela 8.1, adicione o termo de interação,  $e401k \cdot inc$ . Estime a equação por MQO e obtenha os erros padrão usual e robusto. O que você conclui sobre a significância estatística do termo de interação?
  - Agora estime o modelo mais geral por MQP usando os mesmos pesos,  $1/inc$ , como na Tabela 8.1. Calcule o erro padrão usual e o robusto para o estimador de MQP. O termo de interação é estatisticamente significativo usando o erro padrão robusto?
  - Discuta o coeficiente de MQP sobre  $e401k$  no modelo mais geral. Ele atrai muito interesse por si só? Explique.

- (iv) Reestime o modelo por MQP, mas use o termo de interação  $e401k \cdot (inc - 30)$ ; a renda média da amostra é cerca de 29,44. Agora interprete o coeficiente sobre  $e401k$ .

**C12** Use os dados do arquivo MEAP00 para responder esta questão.

- (i) Estime o modelo

$$math4 = \beta_0 + \beta_1 lunch + \beta_2 \log(enroll) + \beta_3 \log(exppp) + u$$

por MQO e obtenha os erros padrão usuais e os erros padrão completamente robustos. De que forma eles geralmente se comparam?

- (ii) Aplique o caso especial do teste de White para heteroscedasticidade. Qual é o valor do teste  $F$ ? O que você conclui?
- (iii) Obtenha  $\hat{g}_i$  como os valores ajustados da regressão  $\log(\hat{u}_i^2)$  sobre  $\widehat{math4}_i$ ,  $\widehat{math4}_i^2$ , em que  $\widehat{math4}_i$  são os valores ajustados de MQO e  $\hat{u}_i$  são os resíduos de MQO. Defina  $\hat{h}_i = \exp(\hat{g}_i)$ . Use  $\hat{h}_i$  para obter os estimadores de MQP. Existem grandes diferenças entre os coeficientes de MQO?
- (iv) Obtenha os erros padrão para MQP que permitem a especificação errada da função de variância. Eles diferem muito dos erros padrão usuais de MQP?
- (v) Para estimar o efeito de gastos sobre  $math4$ , os MQO ou MQP parecem ser mais precisos?

**C13** Use os dados do arquivo FERTIL2 para responder esta questão.

- (i) Estime o modelo

$$children = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 educ + \beta_4 electric + \beta_5 urban + u$$

e registre os erros padrão usual e robusto em relação à heteroscedasticidade. Os erros padrão robustos são sempre maiores do que os não robustos?

- (ii) Adicione três variáveis *dummy* de religião e teste se são conjuntamente significantes. Quais são os  $p$ -valores dos testes não robustos e robustos?
- (iii) A partir da regressão do item (ii), obtenha os valores ajustados  $\hat{y}$  e os resíduos  $\hat{u}$ . Regrida  $\hat{u}^2$  sobre  $\hat{y}$ ,  $\hat{y}^2$  e teste a significância conjunta dos dois regressores. Conclua que a heteroscedasticidade está presente na equação para *children*.
- (iv) Você diria que a heteroscedasticidade encontrada no item (iii) é importante na prática?

**C14** Use os dados do arquivo BEAUTY para esta questão.

- (i) Usando os dados coletados para homens e mulheres, estime a equação

$$lwage = \beta_0 + \beta_1 belavg + \beta_2 abvavg + \beta_3 female + \beta_4 educ + \beta_5 exper + \beta_5 exper^2 + u$$

e registre os resultados usando erros padrão robustos em relação à heteroscedasticidade abaixo dos coeficientes. Alguns dos coeficientes são surpreendentes nos sinais ou nas magnitudes? O coeficiente de *female* é grande na prática e estatisticamente significativo?

- (ii) Adicione interações de *female* com todas as outras variáveis explicativas na equação do item (i) (cinco interações ao todo). Calcule o teste  $F$  usual de significância conjunta das cinco interações e uma versão robusta em relação à heteroscedasticidade. O uso da versão robusta em relação à heteroscedasticidade muda o resultado de forma importante?
- (iii) No modelo completo com interações, determine se aquelas que envolvem as variáveis de aparência – *female* · *belavg* e *female* · *abvavg* – são conjuntamente significantes. Seus coeficientes são pequenos na prática?