

Lista 7 - Econometria I - 2017

Professor: Denisard Alves
Monitor: Adriano Teixeira

Data de entrega: 29 de junho (horário: início da prova)

Exercício 1: Considere a seguinte passagem de John Kay (professor da *London School of Economics*) sobre **heteroscedasticidade**.

“Every serious subject has its jargon. Economists need to know about heteroscedasticity. I take this example because it is virtually impossible to pronounce, and impossible to use the word in front of a class without everyone bursting out into laughter. Indeed, most spell-check programmes reject it, and offer improbable or embarrassing alternatives. Yet heteroscedasticity is an important concept”.

a) Defina formalmente o conceito. Discuta algumas das razões para a presença de heteroscedasticidade em casos empíricos.

“Heteroscedasticity has never been a reason to throw out an otherwise good model” – Gregory Mankiw

“But it should not be ignored either” – Damodar Gujarati

b) Heteroscedasticidade afeta viés/consistência? Os estimadores de Gauss-Markov continuam BLUE? Os testes de hipótese continuam válidos? Faça um resumo das consequências de heteroscedasticidade.

Podemos formalizar um pouco mais.

c) Seja $\hat{\beta}$ o estimador de MQO. Em formato matricial, sob heteroscedasticidade $E(uu') = \Sigma$. Mostre que $Var(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$ sob as hipóteses usuais. Qual deve ser Σ para que os erros sejam homoscedásticos?

d) Se Σ é conhecido, podemos obter estimadores de MQO eficientes e estatísticas adequadas usando a expressão de $Var(\hat{\beta})$ acima. No entanto, em geral Σ é desconhecido. White propôs usar os resíduos da regressão inicial e estimar Σ , o que permitiria obter os erros padrão robustos à heteroscedasticidade. Escreva $\hat{\Sigma}$.

Considere agora um caso numérico.

e) Seja $Y = X\beta + u$, com $E(u|X) = 0$. Sabe-se que

$$\Sigma = E(uu') = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 6 \end{pmatrix}, X = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ e } Y = \begin{pmatrix} 0 \\ 7 \\ 5 \\ 0 \\ 0 \end{pmatrix}.$$

Compute $Var(\hat{\beta})$.

Exercício 2: Você recebeu a base *SuspiciousHetero.dta*. Não há nenhuma informação prévia sobre a natureza dos dados, apenas uma suspeita de que há heteroscedasticidade nos erros provenientes da regressão de Y sobre X . Investigue esta suspeita, seguindo as pistas abaixo.

Pista I: Um método informal de detecção de heteroscedasticidade consiste em fazer um **diagnóstico gráfico** do comportamento dos resíduos. A ideia é fazer a regressão sob a hipótese de homoscedasticidade e observar se \hat{u}_i^2 exibe algum padrão sistemático.

a) Um modo de realizar este exame gráfico é verificar como o valor médio estimado de Y se relaciona com \hat{u}_i^2 . Faça um gráfico de \hat{u}_i^2 contra \hat{Y}_i . Alguma evidência?

b) Outro modo seria investigar como \hat{u}_i^2 se comporta à medida que uma variável explicativa varia. Faça um gráfico de \hat{u}_i^2 contra X . Alguma evidência?

Pista II: Em artigo seminal na *Econometrica* (vol. 34), R. E. Park formalizou o método gráfico ao sugerir que σ_i^2 pode ser uma função dos regressores. Sendo X_i a variável explicativa, a forma sugerida foi:

$$\sigma_i^2 = \sigma^2 X_i^\beta e^{v_i}$$

ou

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i.$$

Como σ_i^2 geralmente é desconhecido, Park sugere estimar a seguinte regressão:

$$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$

e observar a significância de β .

c) Execute o **teste de Park**. Alguma evidência?

Pista III: Embora intuitivamente atraente, o teste de Park apresenta algumas limitações: em especial, não há garantias que $E(v_i|X_i) = 0$ e v_i também pode ser heteroscedástico. Duas décadas depois, os testes de Breusch-Pagan-Godfrey (*Econometrica*; vol. 47) e White (*Econometrica*; vol. 48) se consolidaram como os métodos mais utilizados de detecção de heteroscedasticidade.

d) A partir da regressão auxiliar $\hat{u}_i^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + \text{erro}$, o **teste de Breusch-Pagan-Godfrey** verifica a significância conjunta de $\delta_1 = \dots = \delta_k$. Proceda o teste. Alguma evidência?

e) A partir da regressão auxiliar $\hat{u}_i^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_1^2 + \delta_4 x_2^2 + \delta_5 x_1 x_2 + \text{erro}$, o **teste de White** verifica a significância conjunta de $\delta_1 = \dots = \delta_5$. Proceda o teste. Alguma evidência?

Pista IV: Em diversos casos, testes usuais apontam para a heteroscedasticidade quando na verdade há um **problema de forma funcional**. Estimamos o modelo nível-nível, mas o modelo mais adequado pode ser o quadrático, ou o log-linear, por exemplo. Em muitas aplicações empíricas, observa-se a ocorrência de menos heteroscedasticidade quando a variável dependente está em forma logarítmica.

f) Faça a regressão de $\ln Y$ sobre X . Proceda os testes de Breusch-Pagan-Godfrey e White. Alguma evidência?

g) Use o comando `graph twoway` e faça o gráfico da reta estimada da regressão de $\ln Y$ sobre X . Alguma evidência?

Exercício 3: Considere a seguinte função custo onde C representa o custo total e Q denota a quantidade produzida pela firma 1:

$$C_{1t} = \beta_1 + \beta_2 Q_{1t} + \beta_3 Q_{1t}^2 + \beta_4 Q_{1t}^3 + e_{1t}.$$

Estamos usando o subscrito t pois os dados, disponíveis em *micro.dta*, são de séries de tempo.

a) Sob a hipótese de homoscedasticidade, estime o modelo.

b) Conduza um teste de heteroscedasticidade. Interprete.

c) Estime a regressão robusta à heteroscedasticidade.

d) Suponha que $\text{var}(e_{1t}) = \sigma^2 Q_{1t}$. Estime a regressão por **mínimos quadrados generalizados**. O que ocorre se a função de heteroscedasticidade presumida estiver errada?

e) Suponha agora que a forma funcional de heteroscedasticidade é desconhecida. Proceda a estimação por **MQG factível**.

Exercício 4: Na época de discussão de leis antifumo, um argumento recorrente era que a proibição de fumar nos locais de trabalho desestimularia o vício diminuindo as oportunidades de fumar. Como econometrista, você tem a tarefa de estimar o efeito da proibição do fumo usando uma amostra de 10 mil indivíduos contendo informações individuais e dos locais de trabalho.

A base é *ProibidoFumar.dta*, usada no artigo “*Do Workplace Smoking Bans reduce Smoking?*”, publicado na *American Economic Review*. Uma descrição detalhada das variáveis está disponível em *Variables.pdf*.

a) Estime a probabilidade de fumar para (i) todos os trabalhadores, (ii) trabalhadores afetados pela proibição de fumar e (iii) trabalhadores não afetados.

b) Qual a diferença de probabilidade de fumar entre trabalhadores afetados pela proibição e trabalhadores não afetados? Estime um **modelo de probabilidade linear** (MPL) para testar se esta diferença é estatisticamente significativa. Explique por que a estimação robusta à heteroscedasticidade é mais conveniente para a estimação do MPL.

c) Estime agora o seguinte MPL:

$$\begin{aligned} \text{smoker} = & \beta_0 + \beta_1 \text{smkban} + \beta_2 \text{female} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \beta_5 \text{hsdrop} \\ & + \beta_6 \text{hsgrad} + \beta_7 \text{colsome} + \beta_8 \text{colgrad} + \beta_9 \text{black} + \beta_{10} \text{hispanic} + u \end{aligned}$$

Compare o efeito estimado da proibição de fumar com o efeito obtido no item anterior. Explique esta mudança.

d) Com base nesta última estimação, teste a hipótese de que o coeficiente de *smkban* é estatisticamente significativo a 5% de significância.

e) Ainda com base na regressão do item **c**, teste a hipótese de que a probabilidade de fumar não depende do nível de escolaridade. A probabilidade de fumar aumenta ou diminui com o aumento da escolaridade? Explique.