Estatística II

1º Sem. 2015/2016

Lic. Economia e Finanças

Heterocedasticidade

Heterocedasticidade

- Recordando, **MLR.5: Homocedasticidade:**

$$Var(u|x_1, ..., x_k) = \sigma^2$$

ou, em notação matricial,

$$Var(u|X) = E(uu'|X) = \sigma^2 I$$

- Isto implica que:
  - $Var(u_i|x) = \sigma^2 \;\; \forall\; i = 1, \ldots, n$ (homoced.)
  - $Cov(u_i, u_j|x) = 0 \;\; \forall\; i, j = 1, \ldots, n$ (ausência de autocorr.)

Heterocedasticidade: exemplo

- Exemplo:

$$cerveja_i = \beta_0 + \beta_1 rend_i + u_i$$

com

$$Var(u_i|rend_i) = h(rend_i) = \sigma_i^2.$$

onde σ_i^2 não é constante.

- Por exemplo,

$$Var(u_i|rend_i) = h(rend_i) = \sigma^2 rend_i.$$

Heterocedasticidade: exemplo

- **Heterocedasticidade:** Se MLR.5 for violada, então

$$Var(u_i|x) = h(x_i) = \sigma_i^2,$$

com σ_i^2 não constante, ou, em notação matricial,

$$Var(u|X) = E(uu'|X) = \Sigma = \sigma^2 D,$$

onde

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix}$$

(continua a ser diagonal).

Motivação

Há duas motivações importantes:

1. Variáveis não negativas $y = x\beta + u \geq 0$
   - Particularmente importante quando a média é próxima de zero
   - O valor mínimo que o erro pode assumir é $-x\beta$ (o intervalo de variação de u depende de x), o que induz heterocedasticidade

Motivação

2. Variação de parâmetros (heterogeneidade nos efeitos das variáveis): $\beta_{1i} = \beta_1 + \nu_i$

- O modelo torna-se

$$y_i = \beta_0 + \beta_1 x + \nu_i x + u_i.$$

- Então

$$E\left(\nu_i x + u_i\right) = 0,$$

$$V\left(\nu_i x + u_i\right) = \sigma_u^2 + \sigma_\nu^2 x^2 + 2\sigma_{u\nu} x = h(x).$$

Estes problemas são mais importantes em dados seccionais, mas podem surgir no caso de dados temporais.

Consequências

- A hipótese de **homocedasticidade** não foi usada para obter o valor esperado de $\hat{\beta}$: $\hat{\beta}$ OLS é **centrado** se as hipóteses MLR.1 a MLR.4 são satisfeitas.

- Esta hipótese também não é necessária para provar a **consistência** de $\hat{\beta}$.

Assim, com **heterocedasticidade** ainda se tem

$$E\left(\hat{\beta}\right) = \beta \quad \text{e} \quad P\lim\left(\hat{\beta}\right) = \beta.$$

Consequências

- No entanto, a **homocedasticidade** é essencial para a determinação da matriz de covariâncias de $\hat{\beta}$ e para provar que este é BLUE.

- Com **heterocedasticidade**

$$Var(\widehat{\beta}|X) = \left(X'X\right)^{-1} X'\Sigma X \left(X'X\right)^{-1} \neq \sigma^2(X'X)^{-1}$$

e $\hat{\beta}$ OLS **deixa de ser BLUE**.

Consequências

- O estimador

$$\widehat{Var(\hat{\beta}|X)}_{OLS} = \hat{\sigma}^2(X'X)^{-1}$$

é um **estimador inconsistente** de $Var(\hat{\beta}|X)$.

- Em presença de **heterocedasticidade**, **toda a inferência** que estudamos (testes t e F, e intervalos de confiança) **deixa de ser válida**.

Estimação consistente da matriz de covariâncias

Considere-se o modelo de regressão linear simples

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Já mostrámos que

$$\hat{\beta}_1 = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \beta_1 + \frac{\sum u_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

$$Var\left(\hat{\beta}_1 | x\right) = Var\left(\frac{\sum u_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}\middle| x\right) = \frac{\sum Var(u_i|x)(x_i - \bar{x})^2}{\left(\sum(x_i - \bar{x})^2\right)^2}$$

Estimação consistente da matriz de covariâncias

- Sob homocedasticidade

$$Var\left(\hat{\beta}_1|x\right) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2},$$

- enquanto que, na presença de **heterocedasticidade**,

$$Var\left(\hat{\beta}_1|x\right) = \frac{\sum \sigma_i^2 (x_i - \bar{x})^2}{\left(\sum(x_i - \bar{x})^2\right)^2}.$$

Estimador robusto de $Var(\hat{\beta}|X)$ de White

- Mas sob heterocedasticidade $V\left(\hat{\beta}\right)$ pode ser estimada consistentemente por

$$\widehat{Var\left(\hat{\beta}_1|x\right)}_{WHITE} = \frac{\sum \hat{u}_i^2 (x_i - \bar{x})^2}{\left(\sum (x_i - \bar{x})^2\right)^2}$$

qualquer que seja a forma da heterocedasticidade (incluindo homocedasticidade).

- Os $\hat{u}_i$ são, como de costume, os resíduos da regressão OLS.

Estimador robusto de $Var(\hat{\beta}|X)$ de White

- Para o modelo de regressão múltipla $y_i = x_i\beta + u_i$

$$Var(\hat{\beta}|x) = \left(x'x\right)^{-1} x'\Sigma x \left(x'x\right)^{-1}$$

- Tem-se que o estimador assimptoticamente válido da matriz de covariâncias de $\hat{\beta}$ pode ser obtido como

$$\widehat{Var(\hat{\beta}|x)}_{WHITE} = \left(x'x\right)^{-1} x'Wx \left(x'x\right)^{-1}$$

em que

$$W = \begin{bmatrix} \hat{u}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{u}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{u}_n^2 \end{bmatrix}.$$

Estimador robusto de $Var(\hat{\beta}|X)$ de White

- Com base nesta **Matriz de Covariâncias Robusta** (à heterocedasticidade) podem calcular-se testes t, F e intervalos de confiança.
- A dedução deste resultado deve-se a Eicker, Huber e White.
- Se a matriz de covariâncias robusta é sempre consistente, porque é que não se usa só esta? Porque sob homocedasticidade a formula tradicional é melhor comportada em amostras finitas.

Estimador robusto de $Var(\hat{\beta}|X)$ de White

- Note-se que o teste F tem agora de ser calculado recorrendo ao cálculo matricial.

- As fórmulas habituais de cálculo da F baseadas na soma dos quadrados dos resíduos (SSR) ou nos R^2 **não são válidas**.

Testes de heterocedasticidade

- É importante testar a presença de heterocedasticidade para saber:
  1. que estimador da matriz de covariâncias usar
  2. se o estimador é BLUE
- Há muito testes de heterocedasticidade. Vamos considerar os seguintes:
  - Teste de **Breusch-Pagan** (BP)
  - Teste de **White**
  - Teste **Especial de White** ou **White simplificado**

Heteroskedasticity tests

- Em todos os testes a hipótese nula é sempre de **homoscedasticidade** do erro:

$$H_0 : E\left(u^2 | x_1, x_2, \ldots\right) = \sigma^2$$

- Se u fosse observado, podia fazer a regressão de u^2 nas variáveis relevantes e testar a nulidade conjunta de todos os parâmetros, excepto a constante.

- Como u não é observado, faz-se a mesma coisa mas com $\hat{u}^2$.

- Isto é **assimptoticamente válido**!

Teste de Breusch-Pagan

Procedimento do teste:

$$H_0 : Var(u|x_1, x_2, \ldots, x_k) = \sigma^2$$
$$H_1 : Var(u|x_1, x_2, \ldots, x_k) = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k.$$

1. Estimar a regressão inicial com OLS

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \xrightarrow{OLS} \hat{u}$$

2. Regressão auxiliar

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + erro$$

Teste de Breusch-Pagan

3. Na regressão auxiliar, testar a nulidade conjunta ds coeficientes de declive:

$$H_0' : \delta_1 = \delta_2 = \cdots = \delta_k = 0$$

$$H_1 : \exists\ \delta_j \neq 0,\ j = 1, \ldots, k$$

Para issso, calcular o R_u^2 da regressão auxiliar, e obter uma estatística de teste appropriada.

Teste de Breusch-Pagan

Podemos utilizar a estatística F do teste de significância global da regressão auxiliar, rejeitando H_0 para valores grandes:

$$F = \left[\frac{n-k-1}{k}\right] \frac{R_u^2}{(1-R_u^2)} \sim F_{(k,n-k-1)}$$

$$W = \left\{F : F > F_{\alpha;k,n-k-1}\right\}$$

onde k é o número derestrições, e n o número de observações.

Teste de Breusch-Pagan

Como alternativa, podemos utilizar o teste LM (Lagrange Multiplier), sempre rejeitando H_0 para valores grandes.

- Estatística de teste:

$$LM = n\,R_u^2 \;\overset{a}{\sim}\; \chi^2(k)$$

onde, como anteriormente, k é o número derestrições, e n o número de observações.

- Região crítica:

$$W = \left\{LM : LM > \chi^2_\alpha(k)\right\}$$

Teste de Breusch-Pagan

- Se **rejeitamos H_0 há evidência de heteroscedasticidade**.
- O teste pode ser adaptado para:
  - um subconjunto dos regressores x_j
  - transformações adicionas dos x_j
  - regressores addicionais além dos x_j

  O número de graus de liberdade tem de ser alterado conforme a escolha acima.
- A questão fulcral é quais variáveis utilizar.

Teste de White

- White notou que as únicas variáveis relevantes são os elementos distintos da matriz $x'x$, pelo que sugeriu um teste tipo **BP** usando como variáveis de teste:
  - os regressores iniciais $x_1, x_2, \ldots, x_k$
  - os seus quadrados
  - os seus produtos cruzados.
- Neste caso, a regressão auxiliar torna-se:

$$\begin{aligned}\hat{u}^2 =& \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + \delta_{k+1} x_1^2 + \cdots \\ & + \delta_{2k} x_k^2 + \delta_{2k+1} x_1 x_2 + \cdots + \delta_p x_{k-1} x_k + erro\end{aligned}$$

Teste de White

- A hipótese nula é:

$$H_0' : \delta_1 = \delta_2 = \cdots = \delta_p = 0$$

a testar com a estatística $F \sim F\,(p, n-p-1)$ ou $LM{\sim}\chi^2(p)$ como anteriormente.

- Para modelos com alguns regressores, o teste de White é baseado numa regressão auxiliar com imensos regressores, e porta-se relativamente mal em amostras finitas. Além disso, tende a ter pouca potência.

Um caso especial do Teste de White

- Uma alternativa é fazer um teste de BP usando como variáveis explicativas $\hat{y}$ e $\hat{y}^2$:

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k \\ \hat{y}^2 &= \left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k\right)^2 \end{aligned}$$

- Este teste usa as mesmas variáveis que o teste de White mas impõe restrições aos parâmetros da regressão.

Um caso especial do Teste de White

- A regressão auxiliar é dada por:

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

- A hipótese nula

$$H_0 : \delta_1 = \delta_2 = 0$$

é testada usando $F \sim F(2, n-3)$ ou $LM \sim \chi^2(2)$, e rejeitando H_0 para valores grandes.