Econometria Aplicada

Heterocedasticidade

João Ricardo Costa Filho

“The most important questions of life are, for the most part, really only problems in probability."

Laplace (1812)

“In God we trust. All others must bring data."

William Edwards Deming

A regressão linear

Motivação (tudo começa com uma pergunta)

O que explica o comércio bilateral entre os países?

Referências

- Como responder a questão que motivou a nossa análise?

Referências

- Como responder a questão que motivou a nossa análise?
  - Econometria: capítulo 8 de Wooldridge (2006).

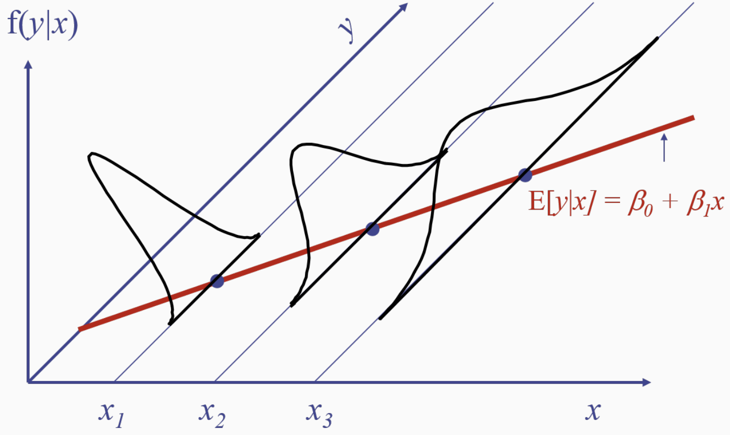
Consequências da heteroscedasticidade

Regressão múltipla

Consideremos o modelo geral:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i,$$

onde Y_i é a variável dependente, X_{ji} é o j-ésimo regressor cujo coeficiente populacional é β_j; ϵ_i é o termo de erro, com $E[\epsilon_i|\mathbf{X}] = 0$ e $Var[\epsilon_i|\mathbf{X}] = \sigma^2$.



f(y|x)
y
E[y|x] = β0 + β1x
x1
x2
x3
x

Sobre a variância dos erros

Sobre a variância dos erros

- A hipótese de homoscedasticidade assume que $\mathrm{Var}(\epsilon_i \mid X_1, X_2, \ldots, X_k) = \sigma^2$.

Sobre a variância dos erros

- A hipótese de homoscedasticidade assume que $\text{Var}(\epsilon_i \mid X_1, X_2, \ldots, X_k) = \sigma^2$.
- A heteroscedasticidade surge quando essa variância depende dos valores das variáveis explicativas.

Sobre a variância dos erros

- A hipótese de homoscedasticidade assume que $\mathsf{Var}(\epsilon_i \mid X_1, X_2, \ldots, X_k) = \sigma^2$.
- A heteroscedasticidade surge quando essa variância depende dos valores das variáveis explicativas.
- Não torna o estimador enviesado;

Sobre a variância dos erros

- A hipótese de homoscedasticidade assume que $\text{Var}(\epsilon_i \mid X_1, X_2, \ldots, X_k) = \sigma^2$.
- A heteroscedasticidade surge quando essa variância depende dos valores das variáveis explicativas.
- Não torna o estimador enviesado;
- Não altera a interpretação do R^2 (ajustado ou não).

Sobre a variância dos erros

- A hipótese de homoscedasticidade assume que $\text{Var}(\epsilon_i \mid X_1, X_2, \ldots, X_k) = \sigma^2$.
- A heteroscedasticidade surge quando essa variância depende dos valores das variáveis explicativas.
- Não torna o estimador enviesado;
- Não altera a interpretação do R^2 (ajustado ou não).
- $Var\left[\hat{\beta}_j\right]$ é enviesada sob a presença de heteroscedasticidade:

Sobre a variância dos erros

- A hipótese de homoscedasticidade assume que $\text{Var}(\epsilon_i \mid X_1, X_2, \ldots, X_k) = \sigma^2$.
- A heteroscedasticidade surge quando essa variância depende dos valores das variáveis explicativas.
- Não torna o estimador enviesado;
- Não altera a interpretação do R^2 (ajustado ou não).
- $Var\left[\hat{\beta}_j\right]$ é enviesada sob a presença de heteroscedasticidade:
  - Erros-padrão não são mais válidos porque a estatística t não segue mais uma distribuição t, estatísticas F não seguem mais uma distribuição F e estatísticas LM não seguem mais uma distribuição χ^2 [Wooldridge (2006); p. 244].

Sobre a variância dos erros

- A hipótese de homoscedasticidade assume que $\mathrm{Var}(\epsilon_i \mid X_1, X_2, \ldots, X_k) = \sigma^2$.
- A heteroscedasticidade surge quando essa variância depende dos valores das variáveis explicativas.
- Não torna o estimador enviesado;
- Não altera a interpretação do R^2 (ajustado ou não).
- $Var\left[\hat{\beta}_j\right]$ é enviesada sob a presença de heteroscedasticidade:
  - Erros-padrão não são mais válidos porque a estatística t não segue mais uma distribuição t, estatísticas F não seguem mais uma distribuição F e estatísticas LM não seguem mais uma distribuição χ^2 [Wooldridge (2006); p. 244].
  - Os estimadores de MQO não são mais BLUE.

Variância amostral de $\hat{\beta}_j$

Consideremos, por esimplicidade, a seguinte relação:

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i.$$

Podemos escrever o estimador por MQO como:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 e_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

Temos que:

$$Var\left[\hat{\beta}_j\right] = \frac{\sum(x_i - \bar{x})^2 \sigma_i^2}{\left(\sum(x_i - \bar{x})^2\right)^2},$$

onde $Var\left[\hat{\beta}_j\right] = \sigma^2 / SQT$ se $\sigma_i^2 = \sigma^2, \forall i$.

Estimador válido de $Var\left[\hat{\beta}_j\right]$

Seja $\widehat{u}_i$ o resíduo da regressão de y sobre x. Temos que

$$\widehat{\mathrm{Var}}\left[\hat{\beta}_1\right] = \frac{\sum(x_i - \bar{x})^2 \hat{u}_i^2}{\left(\sum(x_i - \bar{x})^2\right)^2},$$

é estimativa robusta à heteroscedasticidade (White 1980).

Estimador válido de $Var\left[\hat{\beta}_j\right]$

Em uma regressão linear múltipla, temos que

$$\widehat{\text{Var}}\left[\hat{\beta}_1\right] = \frac{\sum \hat{r}_{ij}^2 \hat{u}_i^2}{SQR_j^2},$$

é estimativa robusta à heteroscedasticidade (White 1980), onde $\hat{r}_{ij}$ é resíduo da regressão de x_j nos demais regressores e $\hat{u}_i$ representa o resíduo da regressão MQO original.

Estatística t robusta

Com a estimativa da variância corrigida, a estatística t,

$$t = \frac{\text{estimativa} - \text{valor hipotético}}{\text{erro-padrão robusto}},$$

segue uma distribuição t e é **assintoticamente** válida mesmo com heteroscedasticidade.

Teste de Breusch-Pagan

Podemos definir a variância de y como uma função das variáveis explicativas (x_i):

$$\sigma_i^2 = E\left(e_i^2\right) = h\left(\alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_s x_{is}\right)$$

$$\mathcal{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_s = 0$$
$$\mathcal{H}_a : \text{ pelo menos um } \alpha_i \neq 0$$

Teste de White (1980)

- O teste de Breusch-Pagan detecta formas lineares de heterocedasticidade.

Teste de White (1980)

- O teste de Breusch-Pagan detecta formas lineares de heterocedasticidade.
- O teste de White (1980), ao inclui termos quadráticos e produtos cruzados dos regressores, perminte não-linearidades.

Por que testar?

- Por que não utilizamos erros-padrão robustos diretamente?

Por que testar?

- Por que não utilizamos erros-padrão robustos diretamente?
  - "Com amostras de tamanho pequeno, as estatísticas t robustas podem ter distribuições que não sejam muito próximas da distribuição t, e isso pode ofuscar a nossa inferência" [Wooldridge (2006); p. 248].

Mínimos Quadrados Ponderados

Mínimos Quadrados Ponderados (WLS)

Assuma que conheçamos a forma da heterocedasticidade:

$$\text{Var}[\epsilon_i|x_i] = \sigma^2 h(x_i).$$

Mínimos Quadrados Ponderados (WLS)

Assuma que conheçamos a forma da heterocedasticidade:

$$\text{Var}[\epsilon_i|x_i] = \sigma^2 h(x_i).$$

* $\mathbb{E}\left[\frac{\epsilon_i}{\sqrt{h(x_i)}}\right] = 0$. * $\mathbb{E}\left[\left(\frac{\epsilon_i}{\sqrt{h(x_i)}}\right)^2\right] = \frac{\mathbb{E}\left[\epsilon_i^2\right]}{h(x_i)} = \frac{\mathbb{E}\left[\sigma^2 h(x_i)\right]}{h(x_i)} = \sigma^2$.

Mínimos Quadrados Ponderados (WLS)

Portanto, podemos dividir a equação por $\sqrt{h(x_i)}$

$$\frac{Y_i}{\sqrt{h(x_i)}} = \frac{(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i)}{\sqrt{h(x_i)}} \implies$$

$$Y_i = \beta^*_{0,i} + \beta^*_1 X_{1i} + \beta^*_2 X_{2i} + \cdots + \beta^*_k X_{ki} + u_i.$$

Mínimos Quadrados Ponderados (WLS)

Portanto, podemos dividir a equação por $\sqrt{h(x_i)}$

$$\frac{Y_i}{\sqrt{h(x_i)}} = \frac{(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i)}{\sqrt{h(x_i)}} \implies$$

$$Y_i = \beta^*_{0,i} + \beta^*_1 X_{1i} + \beta^*_2 X_{2i} + \cdots + \beta^*_k X_{ki} + u_i.$$

Isso é um exemplo de **Mínimos Quadrados Generalizados (GLS)**.

Exemplo [Wooldridge (2006); p. 256]

Qual é a relação entre a poupança e a renda?

Exemplo [Wooldridge (2006); p. 256]

Qual é a relação entre a poupança e a renda?

$$poup_i = \beta_0 + \beta_1 renda_i + u_i,$$

com

$$\text{Var}[u_i | renda_i] = \sigma^2 renda_i,$$

de tal forma que a variância da poupança aumenta com a renda.

Aplicação

Qual é a relação entre a renda e a riquiza líquida?

Aplicação

Qual é a relação entre a renda e a riquiza líquida?

$$\text{nettfa}_i = \beta_0 + \beta_1 \cdot \text{inc}_i + \beta_2 \cdot (\text{age}_i - 25)^2 + \beta_3 \cdot \text{male}_i + \beta_4 \cdot \text{e401k}_i + u_i,$$

onde nettfa_i é a riqueza líquida do indivíduo, inc_i é a renda anual, $(\text{age}_i - 25)^2$ representa a idade centrada (em 25) ao quadrado, male_i é uma dummy de gênero, e401k_i é uma dummy de elegibilidade ao 401(k).

Mínimos Quadrados Generalizados

Mínimos Quadrados Generalizados

- O estimador de GLS é é um WLS onde minimiza-se a soma do erro ao quadrado ponderados por $1/h(x_i)$.

Mínimos Quadrados Generalizados

- O estimador de GLS é é um WLS onde minimiza-se a soma do erro ao quadrado ponderados por $1/h(x_i)$.
- O estimador de GLS é mais eficiente que MQO sob heterocedasticidade conhecida. (É BLUE).

Mínimos Quadrados Generalizados

- O estimador de GLS é é um WLS onde minimiza-se a soma do erro ao quadrado ponderados por $1/h(x_i)$.
- O estimador de GLS é mais eficiente que MQO sob heterocedasticidade conhecida. (É BLUE).
- Procedimento equivalente ao WLS, onde cada observação é ponderada por $1/\text{Var}[\epsilon_i|x_i]$.

Estimador GLS Factível (FGLS)

- Quando não conhecemos $h(x_i)$, o que fazemos?

Estimador GLS Factível (FGLS)

- Quando não conhecemos $h(x_i)$, o que fazemos? Estimamos:

$$\text{Var}[\epsilon_i|x_i] = \sigma^2 \exp(\delta_0 + \delta_1 x_{1,i} + \ldots + \delta_k x_{k,i}) \cdot \nu_i$$

onde σ^2 é a variância "base", ν_i é o erro multiplicativo, com $\mathbb{E}[\nu_i \mid \mathbf{x}_i] = 1$.

Estimador GLS Factível (FGLS)

Assim,

$$\ln\left(\epsilon_i^2\right) = \ln(\sigma^2) + \delta_0 + \delta_1 x_{1i} + \cdots + \delta_k x_{ki} + \ln(\nu_i).$$

Ao definirmos $\alpha_0 = \ln(\sigma^2) + \delta_0$, $e_i = \ln(\nu_i) \Rightarrow \mathbb{E}[e_i \mid \mathbf{x}_i] = 0$, podemos estimar:

$$\ln(u_i^2) = \alpha_0 + \delta_1 x_{1i} + \cdots + \delta_k x_{ki} + e_i$$

Considerações finais

Considerações finais

- WLS e FGLS visam **ganhos de eficiência**, não corrigem viés.

Considerações finais

- WLS e FGLS visam **ganhos de eficiência**, não corrigem viés.
- MQO continua sendo não-viesado e consistente.

Considerações finais

- WLS e FGLS visam **ganhos de eficiência**, não corrigem viés.
- MQO continua sendo não-viesado e consistente.
- Se resultados de MQO e WLS forem muito diferentes, talvez outra hipótese do modelo esteja violada.

Referências i

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica: Journal of the Econometric Society*, 817–38.

Wooldridge, Jeffrey M. 2006. *Introdução à Econometria: Uma Abordagem Moderna*. Cengage Learning.