Heterocedasticidade
Análise de Regressão com Dados de Séries Temporais

Profa. R. Ballini

Bibliografia Básica:

Wooldridge, J. (2016) *Introductory Econometric - A Modern Approach*, Caps. 8, 10, 11, 12.

Greene, W. (2012). *Econometric Analysis*, Cap. 20.

Patterson, K. (2000). *An Introduction to Applied Econometrics - a time series approach*, Cap. 2-5.

Heterocedasticidade

- Verificada heterocedasticidade considera-se as estatísticas robustas
- Alternativamente, utiliza-se a técnica de mínimos quadrados generalizados (MQG);
- Método mais eficiente que MQO no caso de presença de heterocedasticidade;
- Suponha $Var(u|\mathbf{x}) = \sigma^2 h(\mathbf{X})$;
- $h(\mathbf{x}) > 0$ é uma função de $\mathbf{X}$;

Heterocedasticidade

- Podemos reescrever o modelo de regressão como:

$$\frac{Y_i}{\sqrt{h_i}} = \frac{\beta_0^*}{\sqrt{h_i}} + \frac{\beta_1^* X_{1i}}{\sqrt{h_i}} + \ldots + \frac{\beta_k^* X_{ki}}{\sqrt{h_i}} + \frac{u}{\sqrt{h_i}}$$

em que $h_i = h(X)$. Tem-se:

$$E\left[\left(\frac{u_i}{\sqrt{h_i}}\right)^2\right] = \frac{1}{h_i}E(u_i^2) = \sigma^2$$

- $\beta_j^* \rightarrow$ **estimadores de MQG**:
- **Estimadores de mínimos quadrados ponderados**;
- O erro ao quadrado é ponderado por $1/h_i$;
- **Problema:** determinação da função de ponderação h_i.

Estimador MQG factível $\rightarrow$ baseado na seguinte relação:

$$Var(u|X) = \sigma^2 \exp(\delta_0 + \delta_1 X_1 + \ldots + \delta_k X_k) \tag{1}$$

sendo $h(\mathbf{x}) = \exp(\delta_0 + \delta_1 X_1 + \ldots + \delta_k X_k) > 0$

Para a estimação dos δ_j reescrevemos (1) como:

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 X_1 + \ldots + \delta_k X_k)\nu$$

Tomando o logaritmo, temos:

$$ln(u^2) = \alpha_0 + \delta_1 X_1 + \ldots + \delta_k X_k + e \tag{2}$$

em que $e = ln(\nu)$ tem média zero e é independente de $\mathbf{X}$.

Heterocedasticidade

- Regredimos $log(\hat{u}^2)$ sobre $X_1, X_2, \ldots, X_k$;

- Com os valores ajustados ($\hat{g}_i$), calculamos as estimativas de h_i:

$$\hat{h}_i = exp(\hat{g}_i) = exp(\hat{\alpha}_0 + \hat{\delta}_1 x_1 + \hat{\delta}_2 x_2 + \ldots + \hat{\delta}_k x_k)$$

- aplicamos o método de MQG com pesos $1/\hat{h}_i$;

- A interpretação dos parâmetros é a mesma que utilizando MQO.

MQGF para corrigir heterocedasticidade

1. Regrida Y sobre $X_1, \ldots, X_k$ e obtenha $\hat{u}$;
2. Crie $log(\hat{u}^2)$;
3. Regrida $log(\hat{u}^2)$ sobre $X_1, \ldots, X_k$ e obtenha $\hat{g}$ (valores estimados);
4. Calcule $\hat{h} = exp(\hat{g})$;
5. Estime $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + u$ por MQP usando pesos $1/\hat{h}$.

- MQGF é consistente e assimptoticamente mais eficiente que MQO.

Exemplo

A partir dos dados hprice.xlsx, estime o modelo de regressão linear múltipla:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

Use MQG para fazer a correção da heterocedasticidade.

Heterocedasticidade

- Análise gráfica dos resíduos (resíduos $\times$ Y ou resíduos $\times$ X);
- Aplicação de testes de homocedasticidade: Breush-Pagan ou White;
- Em caso positivo: utilizar estatísticas robustas ou MQGF;
- Se mesmo após aplicar MQGF restar dúvidas, calcular estatísticas robustas.
- Limitação no uso MQGF: especificação correta da forma funcional $h(\cdot)$.

Modelo de Regressão com Séries de Tempo

- **Séries Temporais** $\rightarrow$ conjunto de dados ordenados no tempo;
- Pressuposto de que o passado pode afetar o futuro (**dependência**);
- Variáveis aleatórias indexadas no tempo $\rightarrow$ **processo estocástico**;
- Cada conjunto de dados é uma possível realização do processo estocástico;

Processo estocástico:

Família $\{y_t,\ t \in \mathcal{Z}_+\}$ tal que y_t é uma variável aleatória, e $\mathcal{Z}_+$ representa o conjunto de inteiros positivos, definida num espaço de probabilidade.

1) **Modelos Estáticos:**

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

em que y e z são duas séries temporais relacionadas contemporaneamente (efeito imediato em t).

2) **Modelos de Defasagens Distribuídas Finitas (DDF)**:

$$y_t = \beta_0 + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_q z_{t-q} + u_t$$

o qual é um modelo DDF de ordem q.

- $\delta_j \rightarrow$ propensão (tendência) de impacto;
- $\delta_0 + \delta_1 + \ldots + \delta_k \rightarrow$ propensão (impacto) de longo prazo;
- Correlação em z e suas defasagens $\rightarrow$ multicolinearidade.

Pressupostos de Amostra Finita do MQO sob as Hipóteses Clássicas

ST.1 Linear nos Parâmetros:
processo estocástico $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t), t = 1, 2, \ldots, T\}$ segue o modelo linear

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t$$

ST.2 Inexistência de Colinearidade Perfeita

ST.3 Média Condicional Zero:

$$E(u_t|\mathbf{X}) = 0, t = 1, \ldots, T$$

Pressupostos de Amostra Finita do MQO sob as Hipóteses Clássicas

Hipótese subjacente:

$$E(u_t|x_{1t}, x_{2t}, \ldots, x_{kt}) = E(u_t|\mathbf{x}_t) = 0$$

- x_{jt} são **contemporaneamente exógenos**: $Corr(x_{jt}, u_t) = 0, \ \forall \ j$;
- Hipótese ST.3 exige mais que exogeneidade contemporânea;
- Se ST.3 se mantém as variáveis explicativas são **estritamente exógenas**;
- Hipótese pouco realista.

Pressupostos de Amostra Finita do MQO sob as Hipóteses Clássicas

Teorema: Inexistência de Viés do MQO

Sob as hipóteses ST.1, ST.2 e ST.3 os estimadores de MQO são não viesados condicionados em $\mathbf{X}$:

$$E(\hat{\beta}_j) = \beta_j, j = 0, \ldots, k$$

Pressupostos de Amostra Finita do MQO sob as Hipóteses Clássicas

ST.4 Homocedasticidade:

$$Var(u_t|\mathbf{X}) = \sigma^2$$

ST.5 **Inexistência de Correlação Serial**:

$$corr(u_t, u_s|\mathbf{X}) = 0, \text{ para todo } t \neq s.$$

Variâncias Amostrais do MQO

Sob as hipóteses de séries temporais ST.1 a ST.5 de Gauss-Markov, a variância de $\hat{\beta}$, condicional a $\mathbf{X}$ é:

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

Teorema de Gauss-Markov

Sob as hipóteses de séries temporais ST.1 a ST.5, os estimadores de MQO são os melhores estimadores lineares não-viesados condicionais em $\mathbf{X}$.

ST.6 Normalidade: os erros u_t são independentes de $\mathbf{X}$ e são idênticos e independentemente distribuídos como $Normal(0, \sigma^2)$

Distribuições Amostrais Normais:

Sob as hipóteses ST.1 a ST.6, as hipóteses do modelo linear clássico para as séries temporais, os estimadores de MQO são normalmente distribuídos, condicionais em $\mathbf{X}$.

Exemplo: Efeitos da Inflação e dos Déficits sobre as Taxas de Juros

A partir dos dados do arquivo **INTDEF.xlsx**, os quais provêm do Relatório Econômico da Presidência dos EUA para 1997 e abrangem os anos de 1948 a 1996, estime o modelo:

$$i3_t = \beta_0 + \beta_1 inf_t + \beta_3 def_t + u_t$$

em que:

- $i3$: taxa de juros de títulos do Tesouro norte-americano de três meses;
- inf: taxa anual de inflação baseada no índice de preços ao consumidor (IPC);
- def: déficit orçamentário federal como uma percentagem do PIB.

Forma Funcional, Variáveis Dummy e Números-Índices

1. Forma Funcional: frequentemente usa-se logaritmo natural;

2. Variáveis Dummy: pode ser empregada para representar a ocorrência de um determinado evento, em cada período de tempo.

3. Número índices: dado que as magnitudes não são informativas, frequentemnete introduzimos na forma logarítmica, de modo que os coeficientes de regressão têm interpretações de mudanças percentuais.

Exemplo: Efeitos da Isenção de Impostos nas Taxas de Fertilidade

A partir dos dados do arquivo **FERTIL3.xlsx**, retirados do artigo de Whittington, Alm e Peters (1990), o seguinte modelo foi ajustado:

$$gfr_t = \beta_0 + \beta_1 pe_t + \beta_2 ww2_t + \beta_3 pill_t + u_t$$

em que:

> *grf*: é o número de crianças nascidas para cada 1.000 mulheres em idade fértil;
>
> *pe*: taxa de isenção de impostos;
>
> *ww2*: variável binária que recebe o valor um durante os anos de 1941 a 1945;
>
> *pill*: variável binária que recebe o valor um de 1963 em diante, quando a pílula anticoncepcional foi disponibilizada para controle de natalidade.

Exemplo: Efeitos da Isenção de Impostos nas Taxas de Fertilidade

Supondo que a taxa de fertilidade pode reagir a mudanças em *pe* com defasagens, estime o seguinte modelo:

$$gfr_t = \beta_0 + \beta_1 pe_t + \beta_2 pe_{t-1} + \beta_3 pe_{t-2} + \beta_3 ww2_t + \beta_4 pill_t + u_t$$

Caracterização de Séries Temporais

- Muitas séries temporais econômicas apresentam tendência temporal;

- Formulação para representar uma tendência temporal linear:

$$y_t = \beta_0 + \beta_1 t + u_t, t = 1, 2, \ldots$$

- β_1 mede a mudança em y_t devido o tempo.

Valor médio de y_t é:

$$E(y_t) = \beta_0 + \beta_1 t$$

e variância de y_t:

$$Var(y_t) = \sigma^2$$

Tendência exponencial em uma série temporal pode ser captada pela modelagem do log natural da série com uma tendência linear:

$$log(y_t) = \beta_0 + \beta_1 t + u_t, t = 1, 2, \ldots$$

Definindo $\Delta u_t = 0$:

$$\Delta log(y_t) = \beta_1$$

- β_1 representa a taxa de crescimento por período de tempo t.

Uso de Variáveis com Tendência na Regressão

- Devemos incluir o termo de tendência na regressão;
- Não incluí-lo pode resultar em **regressão espúria**

Exemplo: Investimento Imobiliário e Preços de Imóveis

Os dados do arquivo **HSEINV.xlsx** são observações anuais de investimento imobiliário e um índice de preços de imóveis nos EUA de 1947 a 1988. Estime o seguinte modelo:

$$log(invpc_t) = \beta_0 + \beta_1 log(price_t) + u_t$$

Após estime o modelo:

$$log(invpc_t) = \beta_0 + \beta_1 log(price_t) + \beta_2 t + u_t$$

em que $invpc_t$ é o investimento per capita em milhares de dólares e *price* um índice de preços de imóveis.

Procedimento para retirar a tendência de uma série temporal:

1. Estime $y_t = \alpha_0 + \alpha_1 t + e_t$;
2. Defina os resíduos $\hat{e}_t = w$;
3. Regrida w sobre as variáveis explicativas $\rightarrow$ estimadores sem tendência temporal;

- R^2 em séries com tendência, sempre deve ser calculado removida a tendência;
- Muitas séries temporais podem apresentar **sazonalidade**;
- Controle se dá por meio da inclusão de **variáveis dummies sazonais**.