AULA 10 - MQO em regressão múltipla: Propriedades Estatísticas (Variância)

Econometria I

Área quantitativa - IE/UFRJ

Variância dos estimadores MQO

- Vamos incluir mais uma hipótese:

H1 **[Linear nos parâmetros]**

$$y = \alpha + \beta_1 x_1 + ... + \beta_k x_k + u$$

H2 **[Amostragem aleatória]**
Temos uma amostra aleatória de n observações, $\{(x_{i1}, x_{i2}, ..., x_{ik}; y_i) : i = 1, 2, ..., n\}$ do modelo populacional

H3 **[Média condicional zero]**

$$E(u|x_{i1}, x_{i2}, ..., x_{ik}) = 0$$

H4 **[Colinearidade não perfeita]**
Na amostra (e na população), nenhuma das variáveis independentes é constante, e não há relações lineares exatas entre as variáveis indepedentes.

H5 **[Homecedasticidade]**

$$Var(u|x_{i1}, x_{i2}, ..., x_{ik}) = \sigma^2$$

Hipótese H5

- Adicionar H5 não só simplifica as fórmulas, mas também faz com que o MQO tenha a propriedade de eficiência.
- A H5 significa que a variância do termo erro u, condicional as variáveis explicativas, é a mesma para todas as combinações de resultados das variáveis explicativas.
- Se a variância varia com qualquer uma das três variáveis explicativas, então a heterocedasticidade está presente.
- As hipóteses H1 a H5 são necessárias para o teorema de Gauss-Markov em regressão múltipla (para dados de corte transversal).

Variância dos estimadores MQO

- Usaremos a notação **x** para representar o conjunto de variáveis explicativas $(x_1, ..., x_k)$.
- De H1 a H3 podemos escrever

$$E(y|\mathbf{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

  o que significa que o valor esperado de y, dado **x**, é linear nos parâmetros e depende de $(x_1, ..., x_k)$.
- De H5 escrevemos $Var(y|\mathbf{x}) = \sigma^2$, o que significa que a variância de y, dado **x**, não depende dos valores das variáveis indepedentes.

Variância dos estimadores MQO

- Sob H1 a H5, condicionadas aos valores amostrais das variáveis independentes

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{[SQT_j(1 - R_j^2)]}$$

para $j = 1, 2, ..., k$, em que $SQT_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$ é a variação amostral total em x_j e R_j^2 é o R-quadrado da regressão x_j sobre todas as outras variáveis independentes (incluindo o intercepto).
- A variância do erro σ^2: quanto maior for σ^2 maior será a variância do estimador. Uma forma de reduzir a variância do erro é incluir variáveis explicativas.
- SQT_j: quanto maior, menor é a $Var(\hat{\beta}_j)$. Podemos fazer isso aumentando o tamanho da amostra.
- R_j^2 não é o R^2 da regressão. Um elevado grau de relação linear entre as variáveis explicativas pode levar a variâncias grandes. $R_j^2 = 1$ está excluído por H4 (multicolinearidade).

Variância em modelos mal especificados

- Incluir ou não uma variável particular em um modelo pode ser feita ao analisar o dilema entre viés e variância.
- Na aula passada vimos o viés produzido pela omissão, vamos agora incluir nesta análise a variância.
- Suponha o modelo populacional, qua satisfaz H1 a H5

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u$$

- Considere dois estimadores de β_1.
  Da regressão múltipla:

$$\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

  Da regressão simples:

$$\tilde{y} = \tilde{\alpha} + \tilde{\beta}_1 x_1$$

Variância em modelos mal especificados

- Se $\beta_2 \neq 0$ vimos na aula anterior que induz ao viés em $\tilde{\beta}_1$, assim $\hat{\beta}_1$ é preferível a $\tilde{\beta}_1$, entretanto, se calcularmos a variância
  Da regressão múltipla: $Var(\hat{\beta}_1) = \frac{\sigma^2}{[SQT_1(1-R_1^2)]}$
  Da regressão simples: $Var(\tilde{\beta}_1) = \frac{\sigma^2}{SQT_1}$
  Neste caso temos que $Var(\tilde{\beta}_1)$ é sempre menor que $Var(\hat{\beta}_1)$
- Se $\beta_2 = 0$, temos que $\hat{\beta}_1$ e $\tilde{\beta}_1$ não são viesado, mas $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$. Neste caso fica claro que a inclusão de uma variável irrelevante causa apenas maior variância.
- No caso que $\beta_2 \neq 0$ temos que $\hat{\beta}_1$ é não viesado e $\tilde{\beta}_1$ é viesado com $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$

Estimação de σ^2

- Vamos escolher um estimador não viesado de σ^2, que nos permitirá obter estimadores não viesados de $Var(\beta_j)$.
- Como vimos na regressão simples estimar σ^2 por $\hat{u}_i$ leva a um estimador viesado.
- O estimador não viesado de σ^2 no caso geral de regressão múltipla é uma extensão do caso simples

$$\hat{\sigma^2} = \left(\sum u_i^2\right) / (n - k - 1)$$

  no caso da regressão simples $k = 1$ por isso descontamos $(n - 2)$.
- O termo $n - k - 1$ representa os graus de liberdade (gl) do problema geral de MQO com n observações, k variáveis independentes e um intercepto.

$$\begin{aligned} gl &= n - (k + 1) \\ &= (\text{número de observações}) - (\text{número de parâmetros estimados}) \end{aligned}$$

Teorema de Gauss-Markov

- Sob H1 a H5 $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$, são os melhores estimadores lineares não viesados e eficientes de $\beta_1, \beta_2, ..., \beta_k$
- É valido apenas para dados de corte transversal