AULA 9 - MQO em regressão múltipla: Propriedades Estatísticas (Valor Esperado)

Econometria I

Área quantitativa - IE/UFRJ

Valor esperado dos estimadores MQO

- Nesta aula derivamos o valor esperado dos estimadores MQO.
- Em particular, vamos formular e discutir 4 hipóteses, que são extensões diretas das hipóteses do modelo de regressão simples.
- Também derivaremos o viés em MQO, quando uma variável importante for omitida da regressão.

Hipóteses do MQO múltiplo

H1 **[Linear nos parâmetros]**
O modelo da população pode ser escrito como:

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

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]**
O erro u tem um valor esperado igual a zero, dados quaisquer valores das variáveis independentes.

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

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.

H1 - Linearidade

O modelo da população pode ser escrito como:

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

em que $\alpha, \beta_1, ..., \beta_k$ são os parâmetros desconhecidos (constantes) e u é um erro aleatório não observável.

- Modelo linear **simples**: $y = \alpha + \beta_1 x_1 + ... + \beta_k x_k + u$
- Modelo **quadrático**: $y = \alpha + \beta_1 x + \beta_2 x^2$
  Ex: ganhos $= \alpha + \beta_1$ educação $+ \beta_2$ idade $+ \beta_3$ idade$^2 + u$.
- Modelo **loglinear**: $\ln y = \alpha + \beta_1 \ln x_1 + \cdots + \beta_k \ln x_k + u$
  Forma de elasticidade constante, a elasticidade de y em relação às mudanças em x é $\partial \ln y / \partial \ln x_k = \beta_k$

Linearidade significa linear nos parâmetros, não linear nas variáveis.

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

- Às vezes, precisamos escrever a equação de uma observação particular i: para uma observação extraída aleatoriamente da população. temos

$$y_i = \alpha + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i$$

  i refere-se à observação, enquando o segundo subscrito em x é o número da variável.
- Vimos na aula anterior que MQO escolhe as estimativas de uma amostra particular, de modo que os resíduos sejam, em média, iguais a zero e a correlação amostral em cada variável independente e os resíduos seja zero.
- Para que MQO seja não viesado, é preciso que a versão populacional dessa condição seja verdadeira.

H3 - Média condicional zero

O erro u tem um valor esperado igual a zero, dados quaisquer valores das variáveis independentes.

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

- Essa hipótese pode ser violada quando a relação funcional entre as variáveis explicativas está mal-especificada. Por exemplo: esquecemos uma varíavel, usamos o nível da variável mas o log é que aparece no modelo populacional.
- Podemos ter erros de medida em uma variável explicativa.
- Quando a H3 se mantém, dizemos que temos variáveis explicativas exógenas. Se x_j for correlacionado com u por alguma razão própria do modelo, então se diz que x_j é uma variável explicativa endógena.

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.

- Aqui a colinearidade não perfeita é somente em relação às variáveis explicativas.
- Agora devemos examinar as relações entre todas as variáveis explicativas (ou independentes).
- Se uma variável independente é combinação linear exata de outras variáveis independentes, dizemos que existe colinearidade perfeita e não podemos estimar por MQO. Exemplo: x $=$ [1 , renda não formal (x_1), renda formal (x_2), renda total (x_3)]
- Note que é possível que as variáveis independentes sejam correlacionadas, entretanto, não pode ser perfeitamente.

Valor esperado dos estimadores MQO

- Sob as hipóteses H1 a H4 temos que

$$E(\hat{\beta}_j) = \beta_j,\ j = 0, 1, ..., k$$

para qualquer valor do parâmetro populacional β_j.

- Em outras palavaras, os estimadores de MQO são estimadores não-viesados dos parâmetros da população.
- Esperamos que para qualquer amostra aleatória possível tenhamos obtido uma amostra que nos dê uma estimativa próxima do valor da população, mas, infelizmente, isso não pode ser garantido.
- Agora iremos analisar os problemas relacionados com a inclusão de variáveis irrelevantes e omissão de variáveis no modelo de regressão

Inclusão de variáveis irrelevantes no modelo de regressão

- A inclusão de variáveis irrelevantes ou a superespeficação do modelo, significa que uma ou mais variáveis independentes está incluida no modelo, embora ela não tenha efeito parcial sobre y na população (seu coeficiente populacional é zero).
- Suponha que especificamos o modelo como:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

  e esse modelo satisfaz H1 a H4.
- Entretanto, x_3 não tem efeito sobre y após x_1 e x_2 terem sido controlados, o que significa que $\beta_3 = 0$. A variável x_3 pode ou não ser correlacionada com x_1 e x_2.
- Temos que a esperança condicional $E(y|x_1, x_2, x_3) = E(y|x_1, x2) = \alpha + \beta_1 x_1 + \beta_2 x_2$
- Como não sabemos que $\beta_3 = 0$, somos inclinados a estimar a equação com x_3

$$y = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

Inclusão de variáveis irrelevantes no modelo de regressão

- Qual é o efeito de incluir x_3, quando o modelo populacional é zero?
- Lembre-se do resultado da inexistência de viés $E(\hat{\beta}_j) = \beta_j$ para qualquer valor de β_j, incluindo $\beta_j = 0$.
- Podemos concluir que $E(\hat{\beta}_0) = \beta_0$, $E(\hat{\beta}_1) = \beta_1$, $E(\hat{\beta}_2) = \beta_2$, e $E(\hat{\beta}_3) = 0$.
- Mesmo que $\hat{\beta}_3$ não seja exatamente zero, seu valor médio obtido de muitas amostras aleatórias será zero.
- Assim sendo, incluir uma ou mais variáveis irrelevantes não afeta a inexistência de viés dos estimadores de MQO.
- Isso quer dizer que não tem efeito algum? Veremos na aula que vem que isso gera efeitos indesejáveis sobre as variâncias dos estimadores de MQO.

Variável omitida

- Suponha que agora omitimos uma variável que, realmente, pertence ao modelo populacional, em geral chamado de excluir uma variável relevante ou subespeficar o modelo. Neste caso temos um modelo com má-especificação.
- Veremos não só que existe viés no estimador, mas também o tamanho deste viés.
- Suponha o modelo populacional:

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

  e assumimos que esse modelo satisfaz de H1 a H4.
- Suponha que nosso interesse esteja em β_1. Por exemplo, y é o salário hora (ou em log), x_1 a educação e x_2 é uma medida de aptidão inada.

Variável omitida

- Devido a nosso falta de informação ou de dados, estimamos o modelo excluindo x_2, da seguinte forma:

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

- Vamos derivar o valor esperado de $\tilde{\beta}_1$ condicionado aos valores amostrais de x_1 e x_2
- Vimos que no modelo de regressão simples a estimativa de $\tilde{\beta}_1$ é:

$$\tilde{\beta}_1 = \frac{\sum(x_{i1} - \overline{x}_1)y_i}{\sum(x_{i1} - \overline{x}_1)^2}$$

- Considere o modelo populacional $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u$ e substitua no $\tilde{\beta}_1$
$\tilde{\beta}_1 = \frac{\sum(x_{i1} - \overline{x}_1)(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + u)}{\sum(x_{i1} - \overline{x}_1)^2} =$
$\beta_1 + \beta_2 \frac{\sum(x_{i1} - \overline{x}_1)x_{i2}}{\sum(x_{i1} - \overline{x}_1)^2} + \frac{\sum(x_{i1} - \overline{x}_1)u_i}{\sum(x_{i1} - \overline{x}_1)^2}$
- Desde que $E(\sum(x_{i1} - \overline{x}_1)u_i) = 0$ obtemos
$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum(x_{i1} - \overline{x}_1)x_{i2}}{\sum(x_{i1} - \overline{x}_1)^2}$

Variável omitida

- A razão que multiplica β_2 tem a seguinte interpretação: ela é exatamente o coeficiente de inclinação da regressão de x_2 sobre x_1, utilizando nossa amostra de variáveis independentes, que pode ser escrita como:

$$\tilde{x}_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1$$

  então podemos escrever:

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1$$

- Assim o viés em $\tilde{\beta}_1$ é $E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$.
- Se $\tilde{\delta}_1 = 0$, então $\tilde{\beta}_1$ é não viesado, mesmo que $\beta_2 \neq 0$.