AULA:

Regressão por Variáveis Instrumentais (VI)

Prof. Victor Hugo Lachos Davila

Motivação para o uso de VI

Vimos anteriormente que:

Sob a hipótese **cov(u,x) = 0**,

(I) MQO é consistente

Sob a hipótese **E(u|x) = 0**,

(II) MQO é não-viesado

Se essas hipóteses forem violadas, MQO será viesado e inconsistente, sendo necessário buscar um novo método de estimação.

O método de regressão por "**variáveis instrumentais**" (**VI**) é uma solução possível que fornece estimadores consistentes dos parâmetros de interesse

Principais causas do viés do estimador de MQO

- As razões mais comuns para a existência de correlação entre o distúrbio (u) e alguma variável explicativa (x) são:

(1) Omissão de variáveis relevantes
(2) Erros de mensuração
(3) Simultaneidade

Erros de mensuração

Considere o modelo de regressão simples:

$$\boxed{y = \beta_0 + \beta_1 x^* + u}$$

onde cov(x*,u) = E(x*u) = 0.

Nesse modelo, a estimação por MQO deveria gerar estimadores consistentes dos parâmetros.

Supõe-se, porém, que a **variável x* seja observada com erro**. Isto é, o que observamos na prática é:

$$\boxed{x = x^* + e}$$

onde E(e) = 0
cov(x*,e) = E(x*e) = 0
cov(e,u) = E(eu) = 0

Exemplo:

- Para explicar o CR de um aluno da UNICAMP, podemos estar interessados em usar como variáveis explicativas (dentre outras): renda familiar, número de horas dedicadas ao estudo, tempo necessário para o trajeto casa-UNICAMP etc.

- Todas essas variáveis estão sujeitas a **erros de mensuração**, pois os alunos podem errar (deliberadamente ou não) ao responder à pesquisa.

- Se os erros forem **puramente aleatórios**, isto é, não estiverem correlacionados com outras variáveis relevantes, as hipóteses do modelo acima serão satisfeitas.

- Reescrevendo o modelo em função da variável observada x:

$$y = \beta_0 + \beta_1 x^* + u = \beta_0 + \beta_1 (x - e) + u = \beta_0 + \beta_1 x + (u - \beta_1 e)$$
$$= \beta_0 + \beta_1 x + \varepsilon$$

- Agora, a estimação por MQO **não** <u>gera estimadores consistentes</u> dos parâmetros, pois:

$$\mathrm{cov}(x, \varepsilon) = E(x\varepsilon) = E[(x^* + e)(u - \beta_1 e)]$$
$$= E[x^* u + eu - \beta_1 x^* e - \beta_1 e^2] = E(x^* u) + E(eu) - \beta_1 E(x^* e) - \beta_1 E(e^2)$$
$$= -\beta_1 \sigma_e^2$$
$$\neq 0$$

Erros de mensuração

- Lembre que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- E note que

$$\mathrm{var}(x) = \mathrm{var}(x^*) + \mathrm{var}(e) = \sigma^2_{x*} + \sigma^2_e$$

$$\lim(\hat{\beta}_1) = \beta_1 + \frac{\mathrm{cov}(x, \varepsilon)}{\mathrm{var}(x)}$$

$$= \beta_1 - \frac{\beta_1 \sigma^2_e}{\sigma^2_{x*} + \sigma^2_e} = \beta_1\left(1 - \frac{\sigma^2_e}{\sigma^2_{x*} + \sigma^2_e}\right)$$

- Considere a equação:

$$\boxed{y = \beta_0 + \beta_1 x + \boldsymbol{\gamma}' \mathbf{v} + u}$$

onde y a taxa de criminalidade num determinado estado, x é o número de policias e **v** é um vetor que inclui outras variáveis relevantes para explicar y, tal que cov(**v**,u) = 0.

- Não seria razoável esperar que o "**modelo estrutural**" que relaciona as variáveis acima contivesse uma segunda equação,

$$\boxed{x = \alpha_0 + \alpha_1 y + \boldsymbol{\delta}' \mathbf{w} + e},$$

ou seja, que x também dependesse de y?

De fato, mostraremos a seguir que, na primeira equação, em geral a condição

cov(x,u) = 0 é violada e, portanto, o estimador de MQO <u>é inconsistente.</u>

- O fato de que x e u devem ser correlacionados na equação 1 pode ser verificado facilmente, já que

  (1) quando u varia, y varia na mesma direção, pela equação 1;

  (2) quando y varia, x também varia, pela equação 2;

  (3) logo, há correlação entre u e x: **quando u varia, x também varia**!

- Resolvendo o sistema para y e x em função das variáveis exógenas (**v** e **w**) e dos distúrbios, obtemos a "**forma reduzida**":

$$y = \frac{1}{1-\alpha_1\beta_1}[\beta_0 + \beta_1\alpha_0 + \beta_1\boldsymbol{\delta}'\mathbf{w} + \boldsymbol{\gamma}'\mathbf{v} + \beta_1 e + u]$$

$$x = \frac{1}{1-\alpha_1\beta_1}[\alpha_0 + \alpha_1\beta_0 + \alpha_1\boldsymbol{\gamma}'\mathbf{v} + \boldsymbol{\delta}'\mathbf{w} + \alpha_1 u + e]$$

- Logo, é evidente que, em geral, há correlação entre x e u:

$$\mathrm{cov}(x,u) = E(xu) = \frac{\alpha_1\sigma_u^2}{1-\alpha_1\beta_1} \neq 0$$

• Portanto, o estimador de MQO aplicado à equação 1 e <u>inconsistente</u>!. Esse tipo de viés do estimador de MQO é denominado "**viés de equações simultâneas**" ou simplesmente "**viés de simultaneidade**".

• Outros exemplos:

- Horas trabalhadas X salário médio em determinado setor da indústria (oferta e demanda)
- Consumo de bebidas alcoólicas X desempenho do aluno
- Abertura comercial X crescimento econômico
- Democracia X crescimento econômico
- Corrupção X crescimento econômico

- Considere a equação:

$$\boxed{y = \beta_0 + \beta_1 x + u} \qquad (*)$$

onde: E(u) = 0

cov(x,u) $\neq$ 0

- Independentemente do motivo para a existência de correlação entre x e u, o método de **variáveis instrumentais** (**VI**) fornece um estimador consistente dos parâmetros de interesse.

- O método se baseia na utilização de uma variável adicional z, não incluída em (*), que satisfaça tais condições:

(1) Cov(z,u) = 0

(2) Cov(z,x) $\neq$ 0.

- Quando uma variável z satisfaz ambas as condições acima, dizemos que z é um **instrumento válido** para x.
- Vale notar que a condição (1) **não é testável**, pois refere-se à covariância entre z e um erro **não observável**
- Você precisa de uma boa "historinha" para justificar seu instrumento!
- A condição (2), porém, **pode ser testada** em uma regressão de x em z [teste de significância de qual coeficiente?]

- Vimos pela Lei dos Grandes números:

$$\boxed{p\lim(\hat{\beta}_1^{VI}) = \frac{\mathrm{cov}(z, y)}{\mathrm{cov}(z, x)} = \beta_1}$$

- Logo, vemos que um instrumento válido permite efetivamente obter um estimador consistente do parâmetro

- Infelizmente, não é sempre fácil encontrar instrumentos válidos para nossos modelos. Na verdade, é muito difícil!

- Uma das razões dessa dificuldade reside no fato de que as **duas condições** requeridas de um instrumento são muitas vezes **conflitantes**

- Exemplo: estimação de equação de salário em função da educação

  - **Variável omitida**: “habilidade” do indivíduo – viesa coeficiente da educação

  - **Possível instrumento**: educação da mãe (correlacionada com a educação do indivíduo)

  - **Mas**: educação da mãe também deve ser correlacionada com a habilidade do indivíduo presente no erro!

- Pode-se mostrar que

$$\boxed{\operatorname{var}(\hat{\beta}_1^{VI}) = \frac{\operatorname{var}(u)}{n \operatorname{var}(x)[corr(x, z)]^2}}$$

- Por essa razão, devemos procurar um instrumento que tenha a mais alta correlação possível com x

Mínimos Quadrados em 2 Estágios

- Suponha que temos **dois instrumentos válidos** para a "variável endógena" x.

- Ou seja, temos o modelo

$$\boxed{y = \beta_0 + \beta_1 x + u} \quad (*)$$

- onde: $E(u) = 0$
  $cov(x,u) \neq 0$
  $cov(z_1,u) = 0; cov(z_1,x) \neq 0$
  $cov(z_2,u) = 0; cov(z_2,x) \neq 0$

- Será melhor usar z_1 ou z_2 como instrumento?. A resposta é: **melhor usar os dois**!.

- É Claro que devemos escolher a combinação linear de z_1 e z_2 com a maior correlação possível com x.

- Além disso, como z_1 e z_2 tem correlação zero com u, <u>qualquer combinação linear dessas variáveis também terá correlação zero com u. Assim, temos</u> um **instrumento válido** "<u>relativamente eficiente</u>"