# Econometria para Avaliação de Políticas Públicas

## Aula 3: LATE

Prof. Cristine Campos de Xavier Pinto

Itaú Social

13/01/2016

- Auto-seleção nos não-observáveis.
- Como estimar o *ATE* quando

$$\Pr\left[T = 1 | Y(1), Y(0), X\right] \neq \Pr\left[T = 1 | X\right] = p(X)$$

- Na literatura econométrica, tem-se um problema equivalente:

$$\begin{aligned} Y_i &= \alpha + \beta \cdot T_i + \epsilon_i \\ T_i &= 1\left\{\gamma + \delta \cdot Z_i + \eta_i \geq 0\right\} \\ Cov\left(\eta_i, \epsilon_i\right) &\neq 0, \quad Cov\left(Z_i, \epsilon_i\right) = 0 \end{aligned}$$

- A principal diferença está em que o modelo anterior impõe efeitos constantes $\beta_i = \beta \; \forall i$ e "single-index model" no primeiro estágio.
- Mostraremos aqui como interpretar e identificar $\beta$ como sendo *ATE* para uma subpopulação particular. Precisamente, $\beta$ será o *ATE* para aqueles indivíduos para quem um instrumento $Z$ afeta apenas numa direção a probabilidade de tratamento. Esse parâmetro é chamado de *LATE* (local average treatment effects).

# Objetivos

- Como interpretar os estimadores de variáveis instrumentais permitindo que o efeito da variável endógena seja heterogêneo?
- **Idéia**: Os estimadores de variáveis instrumentais geralmente estimam o efeito médio do tratamento, com a média dependendo da escolha do instrumento.
- Por que estimar um efeito heterôgeneo é importante?

# Validade de uma estratégia empírica (research design)

- **Validade Interna**: se dada estratégia conseguiu obter (identificar e estimar) os efeitos causais de interesse para certa população.
  - Exemplos: Boa variável instrumental ou um experimento aleatório
- **Validade Externa**: se o valor encontrado para o efeito é válido em contextos diferentes daquele estudo
  - Exemplo: experimentos aleatórios tem geralmente pequena validade externa.

  - Um instrumental econometrico com efeito heterôgeneo nos ajuda a ter uma idéia da validade interna e externa das estimativas baseadas em variáveis instrumentais.

# A questão da identificação

- Com efeitos de tratamento heterogêneos, endogeneidade cria graves problemas para a identificação de média populacionais.
- Para estimar o efeito populacional médio causal, precisamos de hipóteses fortes sob o efeito do instrumento na variável endógena:
    - Hipóteses do efeito do tratamento é constante.
- Sem esta hipóteses, só identificamos efeitos médios para subpopulações que são **induzidas** pelo instrumento a ter uma mudança no valor das variáveis endógenas.

# Terminologia

- **Compliers**: subpopulações que são induzidas pelo instrumento a ter uma mudança de comportamento na variável endógena.
- **Efeito local médio do tratamento** (LATE): efeito médio do tratamento que é pontualmente identificado para esta subpopulação de compliers.
  - Idéia: Como se uma subpopulação fosse alocada ao tratamento, mas somente uma parte dos indivíduos de fato participam do programa. Para sabermos o efeito médio do tratamento para outras populações, temos que extrapolar.

# Terminologia

- **Compliers**: subpopulações que são induzidas pelo instrumento a ter uma mudança de comportamento na variável endógena.
- **Efeito local médio do tratamento** (LATE): efeito médio do tratamento que é pontualmente identificado para esta subpopulação de compliers.
  - Idéia: Como se uma subpopulação fosse alocada ao tratamento, mas somente uma parte dos indivíduos de fato participam do programa. Para sabermos o efeito médio do tratamento para outras populações, temos que extrapolar.
  - Com o LATE, podemos separar quais são as hipóteses necessárias para fazer a extrapolação para outras subpopulações. Aprender sobre a validade externa vs validade interna das estimativas.

# O que fizeram até o momento..

- Vamos começar com o caso em que a variável endógena é binária (o efeito de ser veterano sob a renda).
- A relação causal que gostaríamos de estimar é:

$$Y_i = \beta_0 + \beta_1 \cdot D_i + \varepsilon_i$$

  - $D_i$: variável endógena
  - $Y_i$: variável explicativa
  - $Z_i$: vetor de instrumentos

- Assumimos que $Z_i$ é relacionado com $D_i$ e satisfaz a hipótese de identificação, $\mathbb{E}\left[\varepsilon_i \cdot Z_i\right] = 0$.

$$D_i = \pi_0 + \pi_1 Z_i + \nu_i$$

- Para entender o LATE, vamos usar uma notação um pouco diferente

**Resultados Potenciais**:

- $Y_i(0)$: resultado se a pessoa $i$ não serviu o exército, independente se a pessoa serviu ou não.
- $Y_i(1)$: resultado se a pessoa $i$ serviu o exército, independente se a pessoa serviu ou não.
- $D_i$: valor realizado para a variável endógena, 0 ou 1.
- Observamos $D_i$ e

$$Y_i = Y_i(D_i) = \begin{cases} Y_i(1) & \text{se } D_i = 1 \\ Y_i(0) & \text{se } D_i = 0 \end{cases}$$

- $Z_i$: variável instrumental binária (ex: se a pessoa teve um número baixo na loteria)
- $D_i(0)$ : o valor da variável endógena se a pessoa teve o número alto na loteria, independente dela ter tido ou não.
- $D_i(1)$ : o valor da variável endógena se a pessoa teve o número baixo na loteria, independente dela ter tido ou não.
- O valor realizado da variável endógena:

$$D_i = D_i(Z_i) = \begin{cases} D_i(1) & \text{se } Z_i = 1 \\ D_i(0) & \text{se } Z_i = 0 \end{cases}$$

Na nossa base de dados, observamos a tríplice:

- $Z_i$
- $D_i(Z_i)$
- $Y_i = Y_i(D_i(Z_i))$

# Hipóteses

Hipótese 1: INDEPÊNCIA

$$Z_i \perp \left(Y_i(0), Y_i(1), D_i(0), D_i(1)\right)$$

**Interpretação**:

- O instrumento é tão bom quanto se eles fossem aleatoriamente alocado, i.e, ele é independente do vetor de resultados potenciais e do vetor de tratamentos potenciais.
- O instrumento não tem um efeito **direto** no resultado de interesse.
- Esta hipótese não é consequência de $Z$ ser aleatoriamente alocado. Por que?

# Hipóteses

- Existem 4 resultados potenciais $Y_i\left(z,d\right)$ que é o resultado potencial que seria observado se o instrumento fosse $Z_i = z$ e o tratamento $D_i = d$.
- Podemos reescrever esta hipótese de independência como a combinação de duas outras hipóteses.

# Hipóteses

- Hipótese 2: Alocação Aleatória

$$Z_i \perp (Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), D_i(0), D_i(1))$$

- Hipótese 3: Restrição de Exclusão

$$Y_i(d,z) = Y_i(d,\acute{z}) \text{ para todo } z, \acute{z}, d$$

# Hipóteses

## Hipótese 2

- Esta hipótese vem da variável instrumental ser alocada de forma aleatória.
- Ela é suficiente para uma interpretação causal de $Z_i$ em $D_i$ :

$$
\begin{aligned}
& \mathbb{E}\left[ D_i \middle| Z_i = 1\right] - \mathbb{E}\left[ D_i \middle| Z_i = 0\right] \\
= \; & \mathbb{E}\left[ D_i(1) \middle| Z_i = 1\right] - \mathbb{E}\left[ D_i(0) \middle| Z_i = 0\right] \\
= \; & \mathbb{E}\left[ D_i(1) - D_i(0)\right]
\end{aligned}
$$

# Hipóteses

## Hipótese 3

- Não é consequência de $Z_i$ ser alocada de forma aleatória.
- Implica que o instrumento só opera atravé de um canal conhecido que chamamos de restrição de exclusão.
- De outra forma, controlando por $d$, $Y(d, z)$ não é uma função de $z$, somente de $d$.
- Como expresamos esta hipótese anteriormente no modelo linear com efeito constante?

# Hipóteses

- Devido a esta restrição de exclusão, podemos definir os resultados potenciais baseados somente no status do tratamento:

$$\begin{aligned} Y_{1i} &\equiv Y_i(1,1) = Y_i(1,0) \\ Y_{0i} &\equiv Y_i(0,1) = Y_i(0,0) \end{aligned}$$

- Neste caso, o resultado potencial pode ser escito como:

$$Y_i = Y_i(0, Z_i) + [Y_i(1, Z_i) - Y_i(0, Z_i)] \cdot D_i$$

- De outra forma,

$$Y_i = \alpha_0 + \rho_i D_i + \varepsilon_i$$

onde $\alpha_0 \equiv \mathbb{E}[Y_{0i} | D_i]$ e $\rho_i \equiv \mathbb{E}[Y_{1i} - Y_{0i} | D_i]$.

# Hipóteses

- Para entender a próxima hipótese, temos que pensar no comportamento induzido pelo instrumento para cada um dos indivíduos.
- Como cada indivíduo responde a diferentes valores do instrumento em termos de tratamento recebido.
- Tipos de "compliance":

| | | $D_i(0)$ | |
|---|---|---|---|
| | | 0 | 1 |
| $D_i(1)$ | 0 | never-taker | defier |
| | 1 | complier | always-taker |

# Hipóteses

- **Problema**: Não podemos estabelecer o tipo da unidade observacional baseado no que estamos observando dado que só observamos o par $(D_i, Z_i)$ e não $(D_i(0), D_i(1))$.
- A informação que obtemos sobre status de tratamento/instrumento é a seguinte:

| | | $Z_i$ 0 | $Z_i$ 1 |
|---|---|---|---|
| $D_i$ | 0 | complier/never-taker | never-taker/defier |
| | 1 | always-taker/defier | complier/always-taker |

# Hipóteses

- Hipótese 4: MONOTONICIDADE

$$D_i(1) \geq D_i(0) \text{ para todo } i$$

  **Interpretação**: Todas as pessoas afetadas pelo instrumento são afetadas da mesma maneira. Existem poucos indivíduos que farão o oposto do que foi alocado a eles.

- Sem esta hipótese, não temos garantia que os estimadores de variável instrumental estão estimando a média ponderada do efeito causal, $Y_{1i} - Y_{0i}$.

# Hipóteses

- Dado a hipótese de monotonicidade, a informação que podemos extrair do comportamento dos indivíduos aumenta:

| | | $Z_i$ = 0 | $Z_i$ = 1 |
|---|---|---|---|
| $D_i$ | 0 | complier/never-taker | never-taker |
| | 1 | always-taker | complier/always-taker |

# Hipóteses

- $\pi_c$: proporção de complies
- $\pi_n$: proporção de never-takers
- $\pi_a$: proporção de always-takers
- Podemos estimar da distribuição populacional do status do tratamento e do instrumento

$$\mathbb{E}\left[D_i | Z_i = 0\right] = \pi_a \text{ e } \mathbb{E}\left[D_i | Z_i = 1\right] = \pi_a + \pi_c$$

- Podemos obter a proporção na população dos diferentes tipos:

$$\begin{aligned} \pi_a &= \mathbb{E}\left[D_i | Z_i = 0\right], \ \pi_c = \mathbb{E}\left[D_i | Z_i = 1\right] - \mathbb{E}\left[D_i | Z_i = 0\right] \\ \pi_n &= 1 - \mathbb{E}\left[D_i | Z_i = 1\right] \end{aligned}$$

# Hipóteses

- Considere o resultado médio para cada um dos status do instrumento e do tratamento:

$$\begin{aligned}\mathbb{E}\left[\left.Y_i\right|D_i=0,Z_i=0\right] &= \frac{\pi_c}{\pi_c+\pi_n}\mathbb{E}\left[\left.Y_i(0)\right|\text{complier}\right] \\ &+ \frac{\pi_n}{\pi_c+\pi_n}\mathbb{E}\left[\left.Y_i(0)\right|\text{never-taker}\right]\end{aligned}$$

$$\mathbb{E}\left[\left.Y_i\right|D_i=0,Z_i=1\right] = \mathbb{E}\left[\left.Y_i(0)\right|\text{never-taker}\right]$$

$$\mathbb{E}\left[\left.Y_i\right|D_i=1,Z_i=0\right] = \mathbb{E}\left[\left.Y_i(1)\right|\text{always-taker}\right]$$

$$\begin{aligned}\mathbb{E}\left[\left.Y_i\right|D_i=1,Z_i=1\right] &= \frac{\pi_c}{\pi_c+\pi_n}\mathbb{E}\left[\left.Y_i(1)\right|\text{complier}\right] \\ &+ \frac{\pi_a}{\pi_c+\pi_a}\mathbb{E}\left[\left.Y_i(1)\right|\text{always-taker}\right]\end{aligned}$$

# Hipóteses

- Destas relações, obtemos o resultado médio para os complier: $\mathbb{E}\left[\left. Y_i(1)\right| \text{complier}\right]$ e $\mathbb{E}\left[\left. Y_i(0)\right| \text{complier}\right]$.
- E obtemos o efeito médio do tratamento para os compliers:

$$
\begin{aligned}
\beta_{IV} &= \mathbb{E}\left[\left. Y_i(1) - Y_i(0)\right| \text{complier}\right] \\
&= \mathbb{E}\left[\left. Y_i(1) - Y_i(0)\right| D_i(1) > D_i(0)\right] \\
&= \mathbb{E}\left[\left. Y_i(1)\right| \text{complier}\right] - \mathbb{E}\left[\left. Y_i(0)\right| \text{complier}\right] \\
&= \frac{\mathbb{E}\left[\left. Y_i\right| Z_i = 1\right] - \mathbb{E}\left[\left. Y_i\right| Z_i = 0\right]}{\mathbb{E}\left[\left. D_i\right| Z_i = 1\right] - \mathbb{E}\left[\left. D_i\right| Z_i = 0\right]}
\end{aligned}
$$

- A nossa base de dados é informativa somente para o efeito médio dos compliers. Por que a base de dados não é informativa sobre o efeito médio para os never-takers e always-takers?

# Caso Especial

- Bloom (1984)
- Suponha que $D_i(0) = 0$. Quando isso acontece?
- Depois somente dois tipos: compliers e always-takers.
- A hipótese de monotonicidade é automaticamente satisfeita.

$$\mathbb{E}\left[ Y_i(1) - Y_i(0) \middle| \text{complier} \right] = \mathbb{E}\left[ Y_i(1) - Y_i(0) \middle| D_i = 1 \right]$$

- Qual é este estimador?

$$\beta_{IV} = p\lim\left(\widehat{\beta}_{2SLS}\right)$$

- onde $\widehat{\beta}_{2SLS}$ é o estimador de dois estágios. No caso de instrumentos binários, $\widehat{\beta}_{2SLS}$ é chamado de estimador de Wald:

$$\widehat{\beta}_{2SLS} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{T}_1 - \overline{T}_0} = \left(\widehat{\mathbf{T}}^{\intercal} \cdot \widehat{\mathbf{T}}\right)^{-1} \cdot \widehat{\mathbf{T}}^{\intercal} \cdot \mathbf{Y}$$

- Hipótese de Independência Condicional:

$$Z_i \perp \left(Y_i\left(0\right), Y_i\left(1\right), D_i\left(0\right), D_i\left(1\right)\right) \mid X_i$$

- Simplesmente para diminuir a variabilidade da variável dependente

- **Abordagem 1**: MQO em dois estágios

$$D_i = \pi_0 + \pi_1 Z_i + \pi_2 X_i + \omega_i$$

$$Y_i = \beta_0 + \beta_1 \cdot D_i + \beta_2 X_i + \varepsilon_i$$

- **Abordagem 2**: Normalidade (Heckman)
Vamos pensar no modelo em duas equações:

$$D_i(z) = \mathbf{I}\{\pi_0 + \pi_1 \cdot z + \pi_2 X_i + \nu_i \geq 0\}$$

$$Y_i(w) = \beta_0 + \beta_1 \cdot w + \beta_2 X_i + \varepsilon_i$$

onde $(\nu_i, \varepsilon_i)$ são normalmente distribuídos.
Como definimos os quatro grupos nesta abordagem?

- **Abordagem 3**: Modelar a distribuição como um todo

$$f_{Y(d)|X,Z}\left(y\left(d\right)|\,x,z\right)=f\left(y|\,x,\theta_{dz}\right)$$

para $(d,z)=(0,n)\,,\ (0,c)\,,\ (1,c),\ (1,d)$. Um modelo natural para a distribuição dos tipos é um modelo trinomial logit.

- Com estas duas distribuições, montamos a função de log-verossimilhança.

**Exemplo**: Efeito do serviço militar na renda

- Angrist (1989)
- Regressão Simples:

$$
\begin{array}{rcccc}
\log(\widehat{\text{earnings}})_i & = & 5.4364 & - & 0.0205 \cdot \widehat{\text{veteran}}_i \\
 & & (0079) & & (0.0167)
\end{array}
$$

Table 4: Treatment Status by Assignment

| | | $Z_i$ = 0 | $Z_i$ = 1 |
|---|---|---|---|
| $W_i$ | 0 | 5,948 | 1,915 |
| $W_i$ | 1 | 1,372 | 865 |

Table 5: Compliance Types: Estimated Proportions

| | | $W_i(0)$ | |
|---|---|---|---|
| | | 0 | 1 |
| $W_i(1)$ | 0 | never-taker (0.6888) | defier (0) |
| | 1 | complier (0.1237) | always-taker (0.1874) |

Table 7: COMPLIANCE TYPES: ESTIMATED AVERAGE OUTCOMES

| | | $W_i(0)$: 0 | $W_i(0)$: 1 |
|---|---|---|---|
| $W_i(1)$ | 0 | never-taker: $\widehat{\mathbb{E}[Y_i(0)]} = 5.4028$ | defier (NA) |
| | 1 | complier: $\widehat{\mathbb{E}[Y_i(0)]} = 5.6948$, $\widehat{\mathbb{E}[Y_i(1)]} = 5.4612$ | always-taker: $\widehat{\mathbb{E}[Y_i(1)]} = 5.4076$ |

- MQO em dois estágios:

$$\widehat{\log(\text{earnings})}_i = \underset{(0.0289)}{5.4836} - \underset{(0.1266)}{0.2336} \cdot \widehat{\text{veteran}}_i$$

# Múltiplos Instrumentos

- Consideres um par de instrumentos binários: $Z_{1i}$ e $Z_{2i}$.
- Suponha que a hipótese de monotonicidade é satisfeita para cada uma destas variáveis binárias.
- O LATE será uma média ponderada dos estimadores de variáveis instrumentais usando $Z_{1i}$ e $Z_{2i}$ separadamente.
- Usamos um estimador de MQO em dois estágios.
- Primeiro Estágio:

$$\widehat{D}_i = \widehat{\pi}_{11} Z_{1i} + \widehat{\pi}_{12} Z_{2i}$$

# Múltiplos Instrumentos

Segundo Estágio:

$$\begin{aligned} \rho_{2SLS} &= \frac{Cov\left(Y_i, \widehat{D}_i\right)}{Cov\left(D_i, \widehat{D}_i\right)} \\ &= \widehat{\pi}_{11} \cdot \frac{Cov\left(Y_i, Z_{1i}\right)}{Cov\left(D_i, \widehat{D}_i\right)} + \widehat{\pi}_{12} \cdot \frac{Cov\left(Y_i, Z_{2i}\right)}{Cov\left(D_i, \widehat{D}_i\right)} \end{aligned}$$

# Múltiplos Instrumentos

$$
\begin{aligned}
&= \widehat{\pi}_{11} \cdot \frac{Cov\left(D_i, Z_{1i}\right)}{Cov\left(D_i, \widehat{D}_i\right)} \cdot \underbrace{\frac{Cov\left(Y_i, Z_{1i}\right)}{Cov\left(D_i, Z_{1i}\right)}}_{\rho_1} \\
&\quad + \widehat{\pi}_{12} \cdot \frac{Cov\left(D_i, Z_{2i}\right)}{Cov\left(D_i, \widehat{D}_i\right)} \cdot \underbrace{\frac{Cov\left(Y_i, Z_{2i}\right)}{Cov\left(D_i, Z_{2i}\right)}}_{\rho_2} \\
&= \lambda \rho_1 + (1 - \lambda)\, \rho_2
\end{aligned}
$$

onde

$$
\lambda = \frac{\widehat{\pi}_{11} Cov\left(D_i, Z_{1i}\right)}{\widehat{\pi}_{11} Cov\left(D_i, Z_{1i}\right) + \widehat{\pi}_{12} Cov\left(D_i, Z_{2i}\right)}
$$

**Exemplo**: Retornos da Educação

- Angrist e Krueger (1991)
- Usam as regras de entrada na escola.
- Usam os trimestres de nascimento como vetor de instrumentos.

Table 8: Average Level of Education by Quarter of Birth

| quarter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| average level of education | 12.69 | 12.74 | 12.81 | 12.84 |
| standard error | 0.01 | 0.01 | 0.01 | 0.01 |
| number of observations | 81,671 | 80,138 | 86,856 | 80,844 |

- Formas Reduzidas:

$$\widehat{\text{educ}_i} = \underset{(0.006)}{12.797} - \underset{(0.013)}{0.109} \cdot \text{qob}_i$$

$$\log(\widehat{\text{earnings}})_i = \underset{(0.001)}{5.903} - \underset{(0.003)}{0.011} \cdot \text{qob}_i$$

$$\widehat{\beta}_{IV} = \frac{-0,1019}{-0,011} = 0,1020$$