Variáveis Instumentais

Raphael Corbi

Universidade de São Paulo

Agosto 2017

Variáveis Instrumentais

- grande questão da econometria aplicada é causalidade
- variéveis instrumentais é a grande ferramenta para:
  - (i) viés de omissão de variáveis
  - (ii) problema de erros de medida
  - (iii) sistema de equações simultâneas
- qual a intuição?

IV e Causalidade

- foco inicial no caso de efeito tratamento homogêneo
- salário e educação (sempre eles!)

$$
\begin{aligned}
Y_{si} &= \alpha + \rho s + \eta_i \\
\eta_i &= A_i'\gamma + \vee \\
E[s_i \vee_i] &= 0
\end{aligned}
$$

- caso Ai seja não-observado, IV para estimar modelo completo

IV e Causalidade

- seja Z_i correlacionada com s_i e nada mais que determina Y_i:
  - isto é, $Cov(Z_i, \eta_i) = 0$
  - **(1) restrição da exclusão**
- dada a restrição, o **estimador de VI** é:

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(s_i, Z_i)} = \frac{Cov(Y_i, Z_i)/Var(Z_i)}{Cov(s_i, Z_i)/Var(Z_i)}$$

- como chamamos o numerador e denominador?
- **(2) denominador = zero?**
- **(1)+(2)** são necessários para interpretação causal de $\hat{\rho}$

Consistência do Estimador IV

- quando o modelo é just-identified, em linguagem matricial:

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$$
$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'[X\beta + \eta]$$
$$\hat{\beta}_{IV} = \beta + (Z'X)^{-1}Z'\eta$$
$$\hat{\beta}_{IV} = \beta + (N^{-1}Z'X)^{-1}N^{-1}Z'\eta$$

- logo temos o resultado de consistência do estimador IV se:
- plim $N^{-1}(Z'\eta) = 0$ ou plim $N^{-1}(Z'X) \neq 0$
- IV é não-viesado? diagrama de Venn (lousa)

Onde encontrar boas VI?

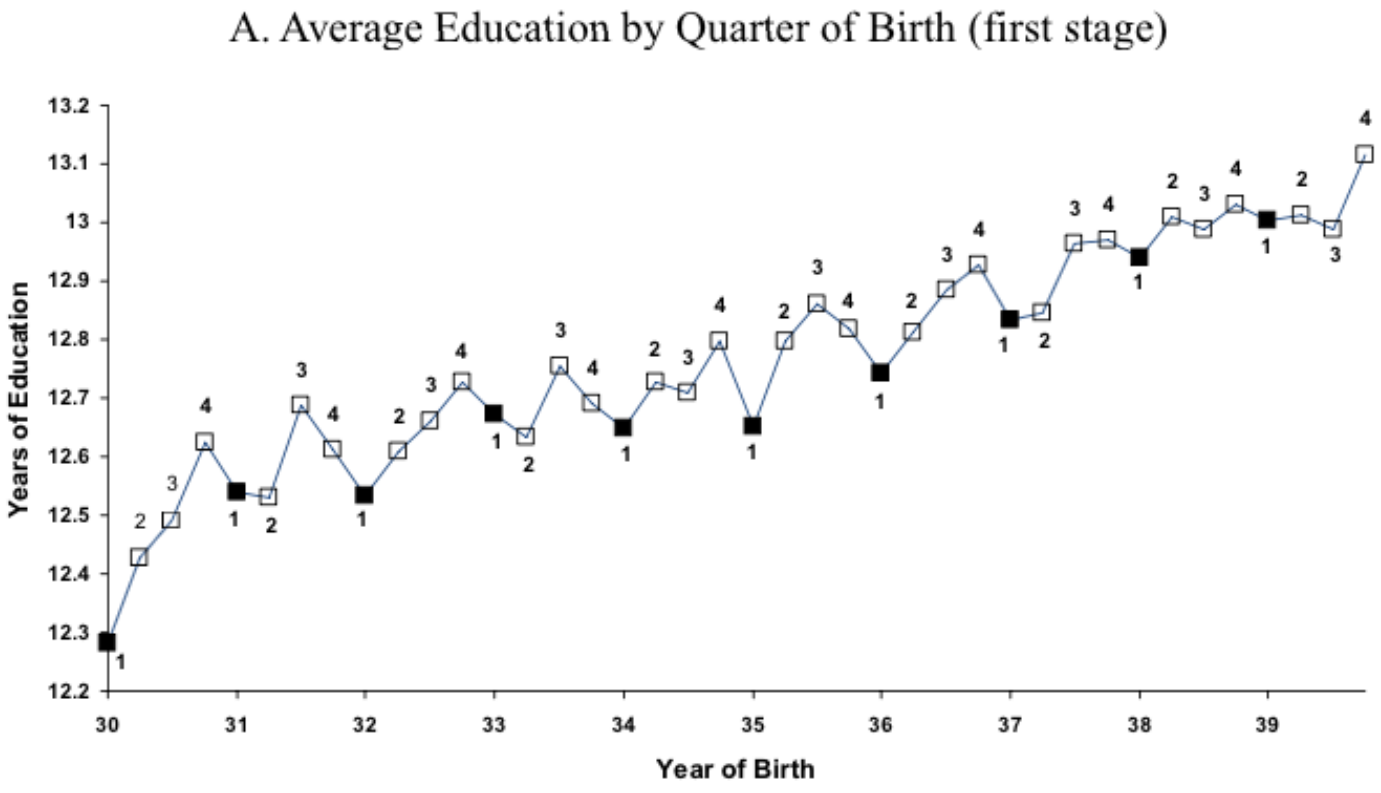
- restrições institucionais / legais (Angrist & Krueger, 1991)
- custo/benefício que afetam determinação do regressor
- natureza / terrorismo (próxima aula)

Angrist & Krueger (1991) - resolvendo o VVO...

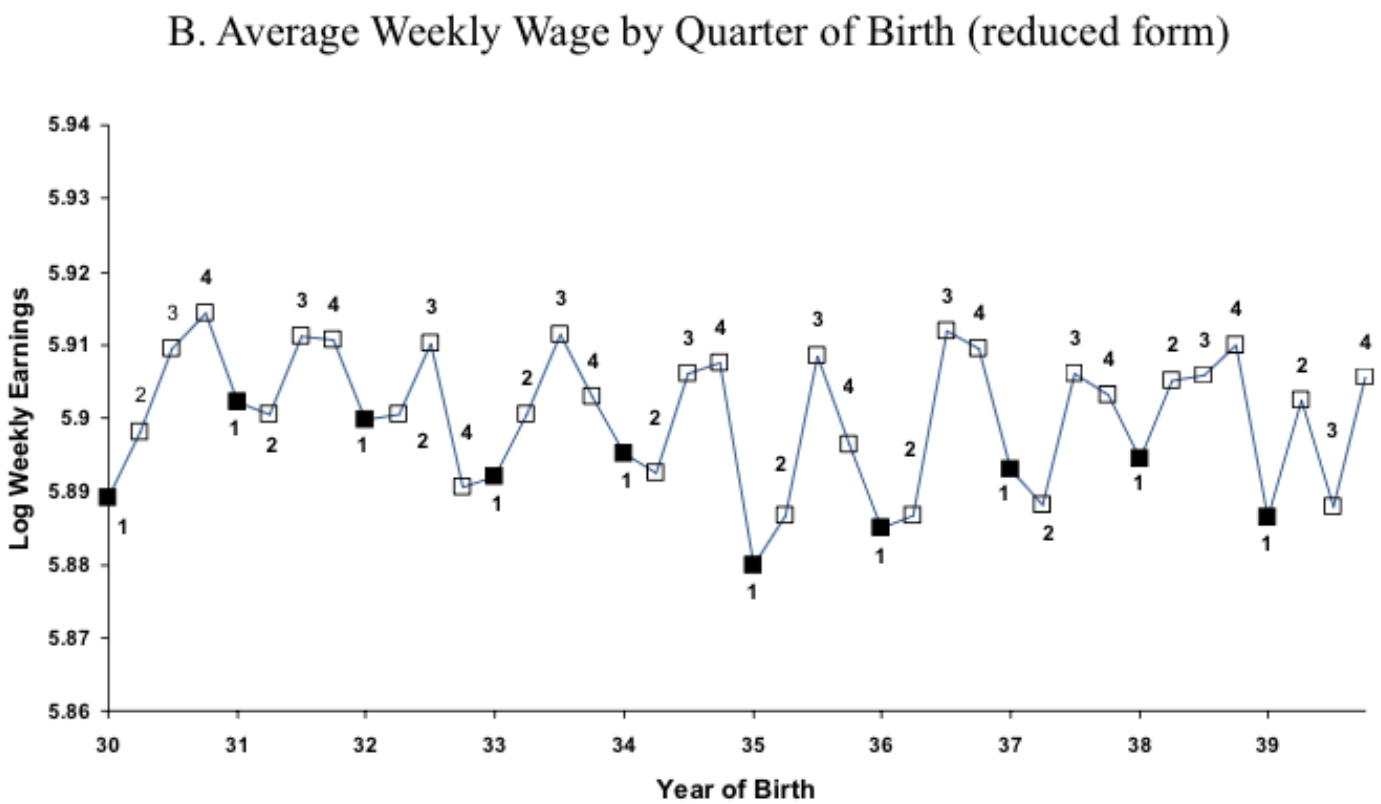
Estratégia de Identificação:

- (i) início da vida escolar no ano que completa 6 anos
  - quanto mais cedo no ano o nascimento, menos escolaridade
  - idade ao entrar na escola é função da data de nascimento
  - nascidos no último trimestre entram na escola com $\leq$ 6 anos
  - nascidos no primeiro trimestre entram na escola com 6.5 anos
- (ii) data legal para abandono e 16 anos

Primeiro Estágio



Forma Reduzida



Angrist & Krueger (1991) - resolvendo o VVO...

- **Primeiro Estágio:**
  - regressão da **variável de interesse** nos controles e instrumento
- **Forma Reduzida:**
  - regressão da **variável dependente** nos controles e instrumento
- Qual a hipótese fundamental deste modelo?
  - habilidade e data de nascimento

Formalizando...

- Primeiro Estágio e Forma Reduzida:

$$s_i = X_i'\pi_{10} + \pi_{11} z_i + \epsilon_{1i}$$
$$Y_i = X_i'\pi_{20} + \pi_{21} z_i + \epsilon_{2i}$$
$$\rho = \frac{\pi_{21}}{\pi_{11}} = \frac{Cov(Y_i, \tilde{z}_i)}{Cov(s_i, \tilde{z}_i)}$$

- ρ é o estimador de VI ajustado pelos controles

2SLS... um tipo de estimador VI

- Equação Estrutural (relação de interesse):

$$Y_i = \alpha' X_i + \rho s_i + \eta_i$$

- Substituindo o Primeiro Estágio na Equação Estrutural:

$$\begin{aligned} Y_i &= \alpha' X_i + \rho[X_i' \pi_{10} + \pi_{11} z_i + \epsilon_{1i}] + \eta_i \\ Y_i &= X_i'[\alpha + \rho\pi_{10}] + \rho\pi_{11} z_i + [\rho\epsilon_{1i} + \eta_i] \\ Y_i &= X_i' \pi_{20} + \pi_{21} z_i + \epsilon_{2i} \end{aligned}$$

2SLS... um tipo de estimador VI

- Rearranjado a Equação Estrutural:

$$Y_i = \alpha' X_i + \rho[\underbrace{X_i'\pi_{10} + \pi_{11} z_i}_{\hat{s}_i}] + \epsilon_{2i}$$

- 2SLS pode ser feito em dois estágios:
  - (i) estimar primeiro estágio usando OLS
  - (ii) estimar segundo estágio usando $\hat{s}_i$ ao invés de s_i
- porém SE incorretos - na prática se estima em conjunto
- $\epsilon_{2i} = \eta + \rho(\hat{s}_i - s_i)$

- qual é a intuição do 2SLS? no diagrama...

2SLS... um tipo de estimador VI

- 2SLS é um estimador de VI
  - no caso de um instrumento e uma variável endógena, eles são iguais
- no caso de instrumentos múltiplos, podemos combiná-los no primeiro estágio para produzir estimativas VI mais precisas:

$$s_i = X_i'\pi_{10} + \pi_{11}z_{1i} + \pi_{12}z_{2i} + \pi_{13}z_{3i} + \epsilon_{1i}$$

- a interpretação é a mesma de antes
- ex: dummies para trimestre de nascimento (A & K, 1991)

Angrist & Krueger - estimativas 2SLS

Table 4.1.1: 2SLS estimates of the economic returns to schooling

	OLS		2SLS					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Years of education	0.075 (0.0004)	0.072 (0.0004)	0.103 (0.024)	0.112 (0.021)	0.106 (0.026)	0.108 (0.019)	0.089 (0.016)	0.061 (0.031)
Covariates:								
Age (in quarters)								✓
Age (in quarters) squared								✓
9 year of birth dummies		✓			✓	✓	✓	✓
50 state of birth dummies		✓			✓	✓	✓	✓
Instruments:			dummy for QOB=1	dummy for QOB=1 or QOB=2	dummy for QOB=1	full set of QOB dummies	full set of QOB dummies int. with year of birth dummies	full set of QOB dummies int. with year of birth dummies

O Estimador de Wald

- mais simples dos estivadores de IV
  - uma dummy instrumento, um regressor endógeno, sem controles

$$Y_i = \alpha + \rho s_i + \eta_i$$

- seja $p = prob(z_i = 1)$, podemos escrever

$$Cov(Y_i, z_i) = \{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]\}p(1-p)$$
$$Cov(s_i, z_i) = \{E[s_i|z_i = 1] - E[s_i|z_i = 0]\}p(1-p)$$
$$\rho = \frac{\{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]\}}{\{E[s_i|z_i = 1] - E[s_i|z_i = 0]\}}$$

O Estimador de Wald

- forma simples e intuitiva de eliminarmos viés de seleção
- hipótese central: z_i afeta Y_i somente via primeiro estágio
- naturalmente, temos que ajustar pela diff médias do FS

- aplicação de Wald nos dados do Angrist & Krueger

Angrist & Krueger - estimativas Wald

Table 4.1.2: Wald estimates of the returns to schooling using quarter of birth instruments

	(1) Born in the 1st or 2nd quarter of year	(2) Born in the 3rd or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (weekly wage)	5.8916	5.9051	-0.01349 (0.00337)
Years of education	12.6881	12.8394	-0.1514 (0.0162)
Wald estimate of return to education			0.0891 (0.0210)
OLS estimate of return to education			0.0703 (0.0005)

Outro Exemplo...

- Angrist & Imbens (1995)
- 1970 - serviço militar por loteria nos USA
- sorteio não determina 100% do alistamento, mas afeta
  - voluntários, mesmo sem serem sorteados
  - sorteados, mas com problemas de saúde

- RF: efeito negativo no salário
- FS: efeito positivo no tratamento

- quick check para ver se instrumento e exógeno
- bo solução para experimento aleatório desvirtuado

Angrist & Imbens (1995)- estimativas Wald

Table 4.1.3: Wald estimates of the effects of military service on the earnings of white men born in 1950

	Earnings		Veteran Status		Wald Estimate of Veteran Effect
Earnings year	Mean	Eligibility Effect	Mean	Eligibility Effect	
	(1)	(2)	(3)	(4)	(5)
1981	16,461	-435.8	0.267	0.159	-2,741
		(210.5)		(0.040)	(1,324)
1971	3,338	-325.9			-2050
		(46.6)			(293)
1969	2,299	-2.0			
		(34.5)			

Teste de Especificação de Hausmann

- Caso haja endogeneidade (ex: variável omitida):
  - OLS é viesado, devemos usar VI
- Caso não haja endogeneidade:
  - OLS e 2SLS são consistentes, mas 2SLS é ineficiente
- Durbin-Wu-Hausmann testa se há endogeneidade nos dados
- idéia: testas se B_{OLS} e B_{2SLS} são significantemente diferentes
- $H = (b_{OLS} - b_{IV})'(V(b_{OLS}) - V(b_{IV}))^{-1}(b_{OLS} - b_{IV}) \sim \chi^2_k - 1$

Sobreidentificação - Teste J de Sargan

- quando temos mais instrumentos do que regressores endógenos
- múltiplos instrumentos podem aumentar a eficiência da estimativa
- instrumento não deve ser correlacionado com o erro
  - mas como testar isso? impossível?

- dado um instrumento válido, podemos testar a validade do outro
- podemos formalmente testar a hipótese - Sargan-Hansen
- (i) estimar modelo com z_1, (ii) predict $\hat{u}$, (iii) reg $\hat{u}$ z_2 z_3
- (iv) $S = nR^2 \sim \chi^2_k - 1$

IV com Resultados Potenciais Heterogêneos

- até agora focamos no caso de efeito tratamento homogêneo

$$Y_{1i} - Y_{0i} = \rho \quad \forall_i, \ \textit{caso tratamento discreto}$$
$$Y_{si} - Y_{s-1,i} = \rho \quad \forall_{i,s}, \ \textit{caso tratamento contínuo}$$

- visão estilizada - impõe homogeneidade e linearidade
- porque efeitos heterogêneos são importantes?
  - (i) validade interna
  - (ii) validade externa

LATE - Local Average Treatment Effects

- seja $Y_i(d, z)$ o resultado potencial do indivíduo i
  - $D_i \in \{0, 1\}$ status de veterano
  - $Z_i \in \{0, 1\}$ eligibilidade para convocação
  - $Y_i(1, Z_i)$-$Y_i(0, Z_i)$ qual a interpretação?
  - $Y_i(D_i, 1)$-$Y_i(D_i, 0)$ qual a interpretação?
- podemos pensar na VI como iniciando uma cadeia de reação causal
- $Z_i \Rightarrow$D$_i$ e $D_i \Rightarrow$Y$_i$

LATE - Local Average Treatment Effects

- status de tratamento observado pode ser escrito como:

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \pi_0 + \pi_{1i}Z_i + \epsilon_i$$

- onde π_{1i} é o efeito heterogêneo causal de Z_i em D_i
- Hipótese 1: Independência $(Y_i(d,z) \vee_{d,z}, D_{1i}, D_{0i}) \perp Z_i$
- Hipótese 2: Restrição de Exclusão ($Y_i(d,z)$ é função somente de d)

$$Y_i(d,0) = Y_i(d,1) \vee_d$$

LATE - Local Average Treatment Effects

- mas qual a diferença entre independência e restrição de exclusão?
  - Angrist e Krueger (1992), Angrist e Chen (2007)
  - o que garante a independência?
  - o que garante a exclusão?

- ex: sorteados podiam pedir dispensa caso estivessem na universidade
- restrição da exclusão = efeito do instrumento tem canal único?

- independência vale, mas R.E. não

LATE - Local Average Treatment Effects

- dada a restrição de exclusão, temos:

$$Y_{1i} \equiv Y_i(1,1) = Y_i(1,0)$$
$$Y_{0i} \equiv Y_i(0,1) = Y_i(0,0)$$

- como antes, reescrevemos Y_i obervado como (note ρ_i):

$$Y_i = Y_i(0, Z_i) + [Y_i(1, Z_i) - Y_i(0, Z_i)]D_i$$
$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$
$$Y_i = \alpha_0 + \rho_i D_i + \eta_i$$

LATE - Local Average Treatment Effects

- Hipótese 3: Existência do Primeiro Estágio
  - $E[D_{1i} - D_{0i}] \neq 0$
- Hipótese 4: Monotonicidade
  - ou $[D_{1i} \geq D_{0i}]$ ou $[D_{1i} \leq D_{0i}] \quad \vee_i$
  - instrumento não precisa afetar todos i, mas todos os afetados devem ser afetados na mesma direção
  - sem monotonicidade, não podemos interpretar IV como média ponderada dos efeitos causais individuais

Teorema do LATE

- A1. Independência: $(Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}) \perp Z_i$
- A2. Exclusão: $Y_i(d, 0) = Y_i(d, 1) \ \forall_d$
- A3. Primeiro Estágio: $E[D_{1i} - D_{0i}] \neq 0$
- A4. Monotonicidade: $[D_{1i} \geq D_{0i}] \ \ \forall_i$

Segue que:

$$\begin{aligned} \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]} &= E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}] \\ &= E[\rho_i|\pi_{1i} > 0] \end{aligned}$$

Teorema do LATE

- um instrumento quase-aleatório (A1) que afeta o resultado somente via um canal (A2), que tem um primeiro estágio (A3) e afeta esta canal somente numa direção (A4) pode ser usado para estimar o efeito causal no grupo afetado.
- ex: estimativas de VI do efeito do serviço militar usando loterias captam o efeito do serviço militar em homens que serviram porque foram sorteados e que não teriam servido caso contrário (exclui voluntários e excluídos por razões médicas).
- Prova!

Prova: Teorema do LATE

$$\begin{aligned}\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]} &= E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}] \\ &= E[\rho_i|\pi_{1i} > 0]\end{aligned}$$

Por exclusão (A2) e independência (A1), temos:

$$\begin{aligned}E[Y_i|Z_i = 1] &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_i|Z_i = 1] \\ &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}]\end{aligned}$$

O numerador do estimador de Wald

$$E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})] = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]P[(D_{1i} > D_{0i})]$$

Qual a utilidade do LATE?

- produz medidas do efeito tratamento com validade interna
  - veteranos da guerra do Vietnam recebem salários justos?
- mas e a validade externa? outra época ou lugar?
  - precisamos de uma teoria para isso (próxima aula)
  - menos experiência? Angrist e Krueger (1994) WWII
- porque precisamos da monotonicidade (A4) no LATE?
  - suponha um instrumento que aumente $prob(D = 1)$ para alguns (**compliers**) e diminua para outros (**defiers**)
  - efeito verdadeiro é positivo para todos, mas a estimativa é zero porque um lado cancela o outro

Qual a utilidade do LATE?

- a abordagem LATE divide a população em três grupos:
  - (i) **Compliers**, sendo $D_{1i} = 1$ e $D_{0i} = 0$
  - (i) **Always-takers**, sendo $D_{1i} = D_{0i} = 1$
  - (i) **Never-takers**, sendo $D_{1i} = D_{0i} = 0$
- LATE é o efeito sobre os compliers
- ex: pacientes são aleatoriamente escolhidos para tratamento experimental
- alguns escolhidos não participam, alguns não-escolhidos participam
- VI pode resolver este problema de triagem aleatório com partial compliance

Qual a utilidade do LATE?

- (LATE) não é a mesma coisa que (ATT)
  - quem são os tratados?

$$\underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{\text{efeito nos tratados}}$$

$$= \underbrace{E[Y_{1i} - Y_{0i}|D_{0i} = 1]}_{\text{efeito nos always-takers}} P[D_{0i} = 1|D_i = 1]$$

$$+ \underbrace{E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]}_{\text{efeito nos complacentes}} P[D_{1i} > D_{0i}, Z_i = 1|D_i = 1]$$

IV em Experimentos Aleatórios

- geralmente num experimento, a aleatorização é feita sobre uma oferta de tratamento, e não do tratamento em si
- indivíduos do grupo de controle não recebem
- indivíduos do grupo de tratamento "podem" receber tratamento
  - aqueles que de fato recebem se auto-selecionam

- experimento aleatório, mas ainda temos seleção?
- qual a direção do viés? quase sempre...

- qual a solução? qual o título do slide?

IV em Experimentos Aleatórios

- podemos usar o sorteio como instrumento para participação
- seja Z_i uma dummy que indica sorteado
- seja D_i uma dummy que indica tratamento
- JTPA training program -
  - takeup rate 60%, one-sided non-compliance

- neste caso, qual a diferença entre LATE e ATT?

JTPA training program

Table 4.4.1: Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status		Comparisons by Assignment Status		Instrumental Variable Estimates	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

Outro exemplo: Angrist 2006

- **Minneapolis Domestic Violent Effort**
- resposta a ligações de violência doméstica
- LATE útil em situações onde aleatorização é anti-ética
- policiais recebiam formulários de cores diferentes, associados a estratégias diferentes
  - aconselhamento, ordem de separação, prisão

- policiais podiam desviar da ação sugerida
- sugestão de prisão era de fato prisão (no "never-takers")
- LATE = ATnT
- prisão reduz repetição da agressão

Caracterizando a Subpopulação Complacente

- cada VI identifica um parâmetro específico específico a subpopulação de complacentes (salvo exceções)
- invalida o teste de superidentificação

- caso a subpopulação complacente seja similar a outra população de interesse, facilita a extrapolação dos resultados
  - Acemoglu e Angrist (2000) - quarter-of-birth vs compulsory attendance laws

- e se as subpopulações complacentes de cada instrumento sejam diferentes mas suas estimativas IV sejam parecidas?
- teste de superidentificação $\Rightarrow$ teste de homogeneidade

Generalizando LATE

- teorema do LATE se aplica a situação sem controles onde um único dummy Z_i
- podemos generalizar de 3 formas:
  - (i) instrumentos múltiplos
  - (ii) controles
  - (i) instrumentos contínuos
- em todos os casos, a estimativa VI é uma média do efeito causal de cada instrumento sobre as subpopulações complacentes
- 2SLS contínua sendo é ferramenta usada e a interpretação similar

Erros comuns do 2SLS - inferência

- Primeiro e Segundo Estágio:

$$s_i = X_i'\pi_{10} + \pi_{11}Z_i + \epsilon_{1i}$$
$$Y_i = \alpha' X_i + \rho\hat{s}_i + [\eta_i + \rho(s_i - \hat{s}_i)]$$

- qual o problema deste procedimento? inferência?

Erros comuns do 2SLS - controles

- mesma coisa, mas com dois tipos de controles: X_{0i} e X_{1i}
- Primeiro e Segundo Estágio:

$$s_i = X_{0i}'\pi_{10} + \pi_{11} Z_i + \epsilon_{1i}$$
$$Y_i = \alpha_0' X_{0i} + \alpha_1' X_{1i} + \rho \hat{s}_i + [\eta_i + \rho(s_i - \hat{s}_i)]$$

- qual o problema deste procedimento?
- resíduos do FS $(s_i - \hat{s}_i)$ pode ser correlacionado com X_{1i}

Peer Effects

- dois tipos de peer effect:
  - 1) média de uma variável afeta o resultado individual (outra variável)
  - ex: média de escolaridade de uma cidade aumento salário individual

$$Y_{ijt} = \delta_j + \lambda_t + \gamma \bar{S}_{jt} + \rho s_i + u_{jt} + \eta_{ijt}$$

- problema de VVO: $\bar{S}_{jt}$, u_{jt}
- Acemoglu e Angrist (2000) usam VI (historical compulsory attendance laws)
- além disso, $\bar{S}_{jt}$ é média de s_i - problema para interpretar

Peer Effects

- 2) média de uma variável afeta o resultado individual da mesma

$$s_{ij} = \mu + \pi_2 \bar{S}_j + \epsilon_{ij}$$

- o coeficiente é matematicamente 1
- podemos excluir s_{ij} de $\bar{S}_j$, mas ambos são afetados por choques do nível da escola
- qualquer choque de grupo cria peer effects espúrios
- Ammermueller and Pischke (2006) - usar livros na casa de $(i)j$

Viés do 2SLS

- OLS é não-viesado e consistence
- 2SLS também é consistente, mas é viesado
- quanto mais instrumentos e mais fracos, maior o viés
- o estimador 2SLS é viesado em direção a OLS

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\epsilon}}{\sigma_{\epsilon}^2} \frac{1}{1+F}$$

- LIML tem viés menor, mas menos eficiente.

Viés do 2SLS

- 1. Estimar o FS e ver se faz sentido. A magnitude e sinal são os esperados?
- 2. Reportar a estatística-F dos instrumentos. Quanto maior, melhor (F¿10)
- 3. Use o melhor instrumento e reporte uma 2SLS uninstrumental (viés menor, menos sujeito a crítica do instrumento fraco)
- 4. Veja a RF: coeficientes, t-stats e F-stats dos instrumentos. Essas estimativas são OLS e não são viesadas.
- **Se não há forma reduzida, não há paper**