

# Variáveis instrumentais

Wooldridge, capítulo 15

- Começar com um modelo de regressão simples:

$$y = \beta_0 + \beta_1 x + u$$

Onde acreditamos que  $x$  e  $u$  são correlacionados,  $cov(x, u) \neq 0$

Para obter estimadores consistentes de  $\beta_0$  e  $\beta_1$  quando  $x$  e  $u$  são correlacionados, é necessária alguma informação adicional

A informação virá por meio de uma nova variável que deve satisfazer certas propriedades

# Variável instrumental ou instrumento

A fim de que uma variável ,  $z$ , sirva como um instrumento válido para  $x$ , o seguinte deve ser verdade:

- 1) O instrumento deve ser exógeno, ou seja,  $\text{Cov}(z,u) = 0$
- 2) O instrumento deve ser correlacionado com a variável endógena  $x$ , ou seja,  $\text{Cov}(z,x) \neq 0$

# Discussão

Requisito 1) envolve a covariância entre  $z$  e o erro não observado  $e$ , assim, não é possível ser testado

Devemos usar o senso comum e a teoria econômica para decidir se faz sentido assumir  $\text{Cov}(z, u) = 0$

Requisito 2): Podemos testar se  $\text{Cov}(z, x) \neq 0$

Simplemente testamos  $H_0: \pi_1 = 0$  em  $x = \pi_0 + \pi_1 z + v$

Algumas vezes nos referimos a esta regressão como regressão de primeiro estágio (*first-stage regression*)

## Regressão $\log(\text{salário})$ em educ

Uma variável instrumental  $z$  de educ deve ser:  
(i) não correlacionado com a aptidão (e com quaisquer outros fatores não observados que afetem o salário); (ii) correlacionada com educação

- VI para educação: escolaridade da mãe. A escolaridade de mãe é positivamente correlacionada com a educação dos filhos. O problema é que pode também estar correlacionada com a habilidade dos filhos
- VI para educação: número de irmãos. Ter mais irmãos está associado com a níveis mais baixos de educação. O problema é que o número de irmãos pode estar relacionado com aptidão uma vez que mais irmãos significa, em média, menos atenção dos pais o que pode resultar em menor habilidade

## Regressão de notas no exame final de faltas

- VI para faltas: distância até o campus. Alunos que moram longo devem faltar mais às aulas
- Problema: alguns fatores em  $u$  podem estar correlacionados com distância. Por exemplo, alunos de famílias de mais baixa renda provavelmente moram mais longe. Se a renda afetar o desempenho dos alunos isto pode fazer com que distância seja correlacionada com  $u$

Onde os economistas vão buscar bons instrumentos?

- 1) Teoria combinada com coleta inteligente de dados
- 2) Variação exógena em políticas ou implementação de programa ao longo do tempo ou do espaço (experimento natural)
- 3) Natureza

# Nos seminários veremos três artigos

- Acemoglu, D., Johnson, S., Robinson, J.A. The colonial origins of comparative development: an empirical investigation
- Dell, M. Path dependence in development: evidence from the Mexican Revolution
- Angrist, J.D., Krueger, A.B. Estimating the payoff to schooling using the Vietnam-era draft lottery Angrist e Krueger

Juvenile incarceration, human capital, and future crime: evidence from randomly assigned judges Aizer e Doyle Jr

$Y = f(\text{Indicador que o jovem foi preso})$

$Y =$  educação (completar ensino médio) e reincidência quando adulto

Correlação positiva entre encarceramento juvenil e fatores como severidade do crime, histórico criminal e características dos jovens. Esses fatores também estão relacionados com as variáveis de resultado

Medida da propensão de um juiz aleatoriamente designado decidir que o jovem deve ser posto sob custódia como instrumento

Estimates of the economic returns to schooling from a new sample of twins Ashenfelter e Krueger

Income, schooling and ability: evidence from a new sample of identical twins Ashenfelter e Rouse

Log salário = f(educação)

Lembrar da aula sobre erros de medida

- A regressão com as variáveis em diferenças acentuava o viés de atenuação
- Ashenfelter, Krueger e Rouse anteciparam o viés de atenuação e pediram para cada gêmeo reportar não só a sua própria escolaridade, mas também a escolaridade do seu gêmeo
- Dados de Twinsburg contêm duas medidas de escolaridade para cada gêmeo, a auto-reportada e a reportada pelo gêmeo

- Usaram a declaração do gêmeo A da escolaridade do gêmeo B como instrumento para a escolaridade auto-reportada pelo gêmeo B e vice-versa
- Gêmeo A e gêmeo B cometem erros quando reportam a escolaridade um do outro, assim como quando reportam a sua própria escolaridade. Desde que os erros feitos pelo gêmeo A ao reportar a escolaridade do gêmeo B sejam não correlacionados com os erros da escolaridade auto-reportada pelo gêmeo B e vice-versa, a escolaridade reportada pelo gêmeo A da escolaridade do gêmeo B pode ser usada como instrumento para a escolaridade auto-reportada do gêmeo B e vice-versa

What are some examples of really clever instrumental variables approaches in econometrics?

Acemoglu and Robinson have really turned the use of instrumental variables into an art:

Using the **Colombian colonial royal road network** to estimate the effect state capacity on development: [Page on mit.edu](#)

Using **access to the Atlantic** to understand trade and development: [AER \(95,3\) p. 546 - The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth](#)

Using the tragedy of the **Holocaust** to estimate the effect of middle class on development: [Social Structure and Development: A Legacy of the Holocaust in Russia](#)

Stephen Levitt, of the *Freakonomics* fame (and his co-authors), is also a master of clever IV:

Using **prison overcrowding** to estimate the effect of prison populations on crime: [The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation](#)

Using **Electoral Cycles in Police Hiring** to Estimate the Effect of Police on Crime (though there is a controversy about the results: [Page on berkeley.edu](#)).

Using **campaign spending** to understand when incumbent politicians make an effort: [The Impact of Federal Spending on House Election Outcomes](#)

Other papers that use clever IVs include:

Using **human wealth** to test the permanent income hypothesis: [Page on jstor.org](#) (Hayashi)

Using **dams** to understand the relationship between agricultural productivity and vulnerability to weather: [Dams](#) (Duflo and Pande)

Using **characteristics of indirect friends** to estimate peer effects in networks: [Identification of peer effects through social networks](#) (Bramouille et al.)

Using **rivers** to see whether school choice affect school quality: [Does Competition Among Public Schools Benefit Students and Taxpayers?](#) (Hoxby; this paper was also very controversial: [Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby \(2000\)](#))

Como a disponibilidade de uma variável instrumental pode ser utilizada para estimar consistentemente os parâmetros da regressão?

Para  $y = \beta_0 + \beta_1 x + u$  e dadas nossas hipóteses

$\text{Cov}(z, y) = \beta_1 \text{Cov}(z, x) + \text{Cov}(z, u)$ , então

$$\beta_1 = \text{Cov}(z, y) / \text{Cov}(z, x)$$

Desta forma, dada uma amostra aleatória da população, o estimador de VI para  $\beta_1$  é

$$\hat{\beta}_1 = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})(x_i - \bar{x})}$$

- Estimador VI de  $\beta_0$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- Estimador VI é consistente
- Estimador VI tem uma distribuição aproximadamente normal em amostras grandes

# Inferência

Impor a hipótese de homoscedasticidade que agora é  $E(u^2 | z) = \sigma^2 = Var(u)$

Dadas as hipóteses 1, 2 e de homoscedasticidade é possível mostrar que a variância assintótica de  $\hat{\beta}_1$  é:

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{n\sigma_x^2\rho_{x,z}^2}$$

Todas as quantidades na fórmula da variância podem ser consistentemente estimadas dada uma amostra aleatória

- Para estimar  $\sigma_x^2$ : calculamos a variância amostral de x
- Para estimar  $\rho_{x,z}^2$  fazemos a regressão de  $x_i$  em  $z_i$  para obter o  $R_{x,z}^2$
- Para estimar  $\sigma^2$  usamos os resíduos de VI

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad i = 1, 2, \dots, n$$

Onde  $\hat{\beta}_0$  e  $\hat{\beta}_1$  são as estimativas de VI

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-2}$$

- Erro padrão assintótico de  $\hat{\beta}_1$  é a raiz quadrada da variância assintótica estimada dada por

$$\widehat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{SQT_x R_{xz}^2}$$

Em que  $SQT_x$  é a soma dos quadrados total de x

- O erro padrão é usado para construir estatísticas  $t$  que envolvem  $\beta_1$  ou intervalos de confiança para  $\beta_1$
- Estimativa  $\hat{\beta}_0$  obviamente tem também um erro padrão (não apresentado)

# VI x MQO

- Erro padrão no caso VI difere do MQO somente pelo fato do  $R^2$  da regressão de  $x$  em  $z$  aparecer no denominador da variância VI
- Uma vez  $R^2 < 1$ , os erros padrão IV são maiores
- Contudo, VI é consistente, enquanto MQO é inconsistente, quando  $\text{Cov}(x, u) \neq 0$
- Quanto mais forte a correlação entre  $z$  e  $x$ , menores os desvios padrão de VI

Preço a pagar quando se usa VI quando  $x$  e  $u$  são não correlacionados: a variância assintótica do estimador de VI é maior e algumas vezes muito maior que a variância assintótica do estimador MQO

# Olhem que iniciativa legal!

imds instituto mobilidade e desenvolvimento social

<https://imdsbrasil.org/em-pauta/materias/50/imds-lanca-portal-agregador-de-politicas-publicas-e-programas-sociais-pelo-mundo>