# Econometria I

## Variáveis instrumentais

Luis A. F. Alvarez

9 de abril de 2025

# Instrumento

- Considere um modelo linear causal da forma:

$$Y = X\beta + U\,,$$

  onde $X$ é uma causa observada escalar, $U$ são causas não observadas, e $\beta$ é o efeito causal de $X$.
- Considere uma situação em que não é razoável supor que $\text{cov}(X, U) = 0$.
  - Nesse caso, a inclinação de $X$ no melhor preditor linear de $Y$ em $X$ e um intercepto não identificará $\beta$, de modo que o estimador de MQO de $Y$ em $X$ e um intercepto não estimará consistentemente o efeito causal de interesse.
- Suponha que observemos uma variável $Z$ que satisfaz:
  - (Relevância) $\text{cov}(Z, X) \neq 0$.
  - (Exogeneidade ou exclusão) $\text{cov}(X, U) = 0$
- À variável $Z$ damos o nome de instrumento:
  - Trata-se de variável que exibe associação, na população, com $X$ (relevância), e cuja **única associação com $Y$ se dá através de $X$** (exogeneidade ou exclusão).

# Estimando de Wald e Identificação sob variável instrumental

- Defina **o estimando de Wald** como o **parâmetro**:

$$\gamma_{\mathsf{Wald}} := \frac{\mathrm{cov}(Z, Y)}{\mathrm{cov}(Z, X)}\,,$$

- Sob as hipóteses de relevância e exogeneidade do instrumento, estimando identifica $\beta$, i.e.:

$$\gamma_{\mathsf{Wald}} = \beta\,.$$

# Estimação e Inferência

- Dada uma amostra aleatória $(Y_i, X_i, Z_i) \sim (Y, X, Z)$, resultado anterior sugere que estimemos o efeito causal como:

$$\hat{\gamma}_{\text{Wald}} = \frac{\widehat{\text{cov}(Z,Y)}}{\widehat{\text{cov}(Z,X)}} = \frac{\hat{b}_{Y,Z}}{\hat{b}_{Z,X}},$$

onde $\hat{b}_{S,Z}$ é o estimador de MQO para o coeficiente de $S$ numa regressão de $S$ num intercepto e $Z$.
- Sob condições de regularidade da aula anterior, vimos que, com $n \to \infty$

$$\hat{b}_{Y,Z} \xrightarrow{p} \frac{\text{cov}(Y,Z)}{\mathbb{V}(Z)}$$

$$\hat{b}_{X,Z} \xrightarrow{p} \frac{\text{cov}(X,Z)}{\mathbb{V}(Z)}$$

de onde segue, por aplicação do teorema do mapa contínuo que, sob as condições de relevância e exogeneidade:

$$\hat{\gamma}_{\text{Wald}} \xrightarrow{p} \gamma_{\text{Wald}} = \beta\,.$$

# Testando relevância

- Note que, como estimador de $\hat{b}_{X,Z}$ é consistente para $\frac{\text{cov}(X,Z)}{\mathbb{V}(Z)}$, podemos utilizar esse estimador para realizar um teste da nula de que que o instrumento **não** é relevante.
- Por outro lado, em nosso ambiente com uma variável instrumental e um modelo causal linear, a hipótese de exclusão é **intestável**.
    - Devemos suplementar a análise empírica com uma argumentação de porquê o único mecanismo através do qual variações em $Z$ produzem variações em $Y$ é através de $X$.

## Caso geral

- Vamos agora considerar um modelo causal geral da forma:

$$Y = X'\beta + U$$

onde $X$ é um vetor de $k$ causas observadas.

- Vamos supor a existência de um vetor $Z$ de $l$ instrumentos que satisfaz as hipóteses:

**Hipótese (H1-Relevância)**

$\mathbb{E}[ZX']$ *tem posto* $k$*.*

**Hipótese (H2-Exogeneidade)**

$\mathbb{E}[ZU] = 0$

- Note que a hipótese de exogeneidade é equivalente a $\text{cov}(Z, \epsilon) = 0$ quando $X$ inclui um intercepto, pois nesse caso podemos supor que $U$ tem média zero "absorvendo sua média" ao intercepto.

# Variáveis Endógenas e Exógenas

- A condição de relevância implica que necessitamos de $l \geq k$ variáveis que exibam suficiente variação com as causas $X$, e que não exibam associação com as causas $U$.
  - Se uma entrada $X_j$ é exógena por hipótese, i.e. $\mathbb{E}[X_j U] = 0$, podemos incluí-la entre os instrumentos $Z$.
  - Por outro lado, para cada variável $X_s$ endógena, i.e. tal que potencialmente $\mathbb{E}[X_j U] \neq 0$, precisamos de pelo menos um instrumento que exiba associação com as causas observadas, mas não com as causas não observadas.

# Identificação dos efeitos no modelo linear geral

- Sob as hipóteses H1 e H2, para qualquer matriz $A$ $k \times l$ de posto cheio, temos que:

$$\gamma_A := (\mathbb{E}[AZX'])^{-1}\mathbb{E}[AZY] = \beta$$

- Note que, o resultado anterior sugere que, para uma amostra aleatória $(Y_i, X_i, Z_i) \sim (Y, X, Z)$ e uma dada matriz $A$, podemos estimar $\beta$ por:

$$\hat{\gamma}_A = \left(\sum_{i=1}^{n} \hat{A} Z_i X_i'\right)^{-1} \sum_{i=1}^{n} \hat{A} Z_i Y_i = (\hat{A}\boldsymbol{Z}'\boldsymbol{X})^{-1}\hat{A}\boldsymbol{Z}'\boldsymbol{y}\,,$$

onde $\hat{A}$ é um estimador da matriz $A$ e:

$$\boldsymbol{Z} = \begin{bmatrix} Z_1' \\ Z_2' \\ \vdots \\ Z_n' \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{bmatrix}$$