Econometria I

Teste de Hipóteses Lineares

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024

Sumário I

Teste de Hipótese

- Dado o modelo

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{1}$$

podemos estar interessados em testar várias hipóteses sobre os parâmetros $\beta_1, \beta_2, ..., \beta_k$. Por exemplo:

$$H_0 : \beta_2 = 0$$
$$H_0 : \beta_2 = -1$$
$$H_0 : \beta_2 + \beta_3 = 1$$
$$H_0 : \beta_2 = \beta_4$$
$$H_0 : \beta_2 - \beta_4 = 0$$
$$H_0 : \beta_2 = \beta_3 = \beta_4 = \cdots = \beta_k = 0$$
$$H_0 : \beta_2 = \beta_3 = 0$$

- Qualquer uma dessas hipóteses pode ser rescrita no formato matricial:

$$\boldsymbol{R\beta} = \boldsymbol{c} \tag{2}$$

em que $\boldsymbol{R}$ é uma matriz de dimensão $q \times k$, $\boldsymbol{\beta}$ é um vetor de coeficiente lineares de dimensão $k \times 1$ e $\boldsymbol{c}$ é um vetor de dimensão $q \times 1$.

- A matriz $\boldsymbol{R}$ "codifica" as hipóteses a serem testadas. Cada linha corresponde a uma restrição linear sobre o vetor $\boldsymbol{\beta}$. Logo, q é o número de restrições a serem testadas.

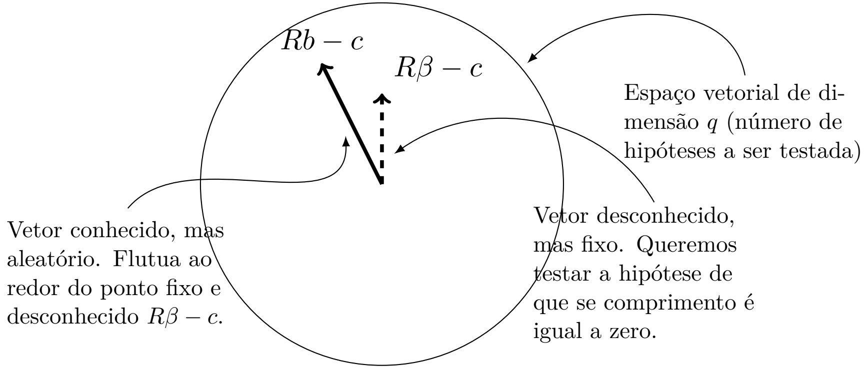
- Assim, temos que um conjunto qualquer de hipóteses lineares é substituído por uma única hipótese matricial:

$$\boldsymbol{H}_0\colon \boldsymbol{R\beta} - \boldsymbol{c} = \boldsymbol{0} \tag{3}$$

- **Questão**: o vetor $(\boldsymbol{R\beta} - \boldsymbol{c})$ tem um comprimento maior do que zero?
- Duas situações
  1. Se esse comprimento é nulo, significa aceitar o conjunto das hipóteses codificadas em $\boldsymbol{R}$ e $\boldsymbol{c}$.
  2. Se o comprimento é maior do que zero, corresponde a rejeição de um ou mais das hipóteses conjuntas originais.
- Como $\boldsymbol{\beta}$ é não conhecido, temos que o vetor $(\boldsymbol{R\beta} - \boldsymbol{c})$ também será não conhecido. É por isso que vamos testar a hipótese nula através do estimador de mínimos quadrados. Assim, dado o estimador $\widehat{\boldsymbol{\beta}}$, podemos calcular o vetor $(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})$.

Espaço Geométrico

Figura 1: Geometria do teste de hipótese



- Quanto mais longe o vetor $(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})$ estiver de $\mathbf{0}$, menos provável é que o vetor $(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})$ seja igual a zero. Com isso, tenderemos a rejeitar a hipótese nula.
- Como em qualquer teste de hipótese, a questão principal é se o desvio de $(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})$ em relação a $\mathbf{0}$ pode ser atribuído a erro de amostragem, ou se é de fato significativo.
- Para testar H_0, vamos investigar a distribuição do quadrado do comprimento de $(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})$, sob H_0.
- Esse vetor nada mais é do que uma transformação linear do vetor aleatório $\widehat{\boldsymbol{\beta}}$, cuja distribuição é dada por:

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \tag{4}$$

- Temos que:

$$\mathbb{E}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right) = \mathbb{E}\left(\boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{c}\right) = 0 \tag{5}$$

  sob a hipótese nula.
- E

$$\begin{aligned}
\operatorname{var}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right) = \operatorname{var}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}\right) &= \mathbb{E}\left[\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{R}\boldsymbol{\beta}\right)\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{R}\boldsymbol{\beta}\right)'\right] \\
&= \mathbb{E}\left[\boldsymbol{R}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \boldsymbol{R}'\right] \\
&= \boldsymbol{R}\operatorname{var}\left(\widehat{\beta}\right)\boldsymbol{R}' \\
&= \sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}' \qquad (6)
\end{aligned}$$

- Assim,

$$\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right) \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right) \tag{7}$$

- Se $\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right)$ é uma normal multivariada com média $\mathbf{0}$, o seu comprimento ao quadrado, dado por $\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right)\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right)'$ também será uma soma de quadrados de uma variáveis aleatórias normais.
- É uma variável aleatória não tabelada, mas com um forte "parentesco" com uma variável aleatória qui-quadrado.
- **Questão**: como torná-la uma qui-quadrado, com valores críticos conhecidos? Vamos usar um resultado para distribuições de formas quadráticas e mostrar que ela possui distribuição qui-quadrado.
- Se um vetor $\boldsymbol{x}$ de dimensão $q \times 1$ tem distribuição

$$\boldsymbol{x} \sim N(\mathbf{0}, \Sigma) \text{ então } x'\Sigma^{-1}\boldsymbol{x} \sim \chi^2(q) \tag{8}$$

- Assim, chegamos a uma variável aleatória tabelada, sobre a qual poderíamos realizar testes de hipóteses:

$$\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right)' \left[\sigma^2 \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right) \sim \chi^2(q) \tag{9}$$

- Essa expressão deve ser entendida como o quadrado do comprimento "padronizado" do vetor $(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})$, isto é, medido em desvios padrões.
- O problema prático que surge com a aplicação acima é a presença de σ^2, que é um parâmetro desconhecido.
- Contudo, sabemos que

$$\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2} \sim \chi^2(n-k) \tag{10}$$

e que esta estatística tem distribuição independente de $\widehat{\boldsymbol{\beta}}$.

Distribuição do Teste de Hipótese

- Também sabemos que a razão entre duas variáveis qui-quadrado independentes, divididas pelos respectivos graus de liberdade n_1 e n_2, gera uma variável com distribuição $F(n_1, n_2)$.
- Então podemos concluir que:

$$\frac{\frac{\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{c}\right)'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{c}\right)}{q}}{\frac{(\boldsymbol{e}'\boldsymbol{e})}{(n-k)}} \sim F(q, n-k) \quad (11)$$

- Usando a definição de s^2, obtemos:

$$\frac{\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{c}\right)'\left[s^2\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{c}\right)}{q} \backsim F(q, n-k) \quad (12)$$

- A estatística dada pela eq. (12) pode ser usada para testar hipóteses lineares sobre o vetor $\boldsymbol{\beta}$.
- A regra de decisão será: valores elevados da estatística apontam para a rejeição de H_0.
- **Observação**: a raiz quadrada de uma variável $F(1, n)$ é uma variável $t(n)$.
- Assim, no caso de uma única restrição ($q = 1$), a raiz quadrada da estatística F, eq. (12), equivale a uma estatística t.

Exemplo 1

- Vamos testar a hipótese nula $H_0 \colon \beta_2 = 0$.
- A hipótese escrita no formato geral se torna: $\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c} = \widehat{\beta}_2$.
- Assim, $\text{var}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}) = s^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' = \text{var}(\widehat{\beta}_2)$. Usando a eq. (12), temos:

$$\frac{\widehat{\beta}_2' \left[\text{var}(\widehat{\beta}_2)\right]^{-1} \widehat{\beta}_2}{1} = \frac{\widehat{\beta}_2^2}{\text{var}(\widehat{\beta}_2)} \sim F(1, n-k) \tag{13}$$

- Tirando a raiz quadrada, temos:

$$\frac{\widehat{\beta}_2}{d.p.(\widehat{\beta}_2)} \sim t(n-k) \tag{14}$$

Exemplo 2

- Vamos testar a hipótese nula $H_0\colon \beta_2 + \beta_3 = 1$. Com isso, $\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c} = \widehat{\beta}_2 + \widehat{\beta}_3 - 1$.
- Assim,

$$
\begin{aligned}
& s^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' = \\
& = s^2 \begin{bmatrix} 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} c_{11} & c_{21} & c_{31} & \cdots & c_{k1} \\ c_{12} & c_{22} & c_{32} & \cdots & c_{k2} \\ c_{13} & c_{23} & c_{33} & \cdots & c_{k3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{1k} & c_{2k} & c_{3k} & \cdots & c_{kk} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (15)
\end{aligned}
$$

- Desse modo,

$$\begin{aligned} s^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' &= s^2(c_{22}+c_{23}+c_{32}+c_{33}) \\ &= s^2(c_{22}+2c_{23}+c_{33}) \\ &= \operatorname{var}(\widehat{\beta}_2)+2\operatorname{cov}(\widehat{\beta}_2,\widehat{\beta}_3)+\operatorname{var}(\widehat{\beta}_3) \\ &= \operatorname{var}(\widehat{\beta}_2+\widehat{\beta}_3) \end{aligned} \tag{16}$$

- Logo,

$$\frac{\left(\widehat{\beta}_2+\widehat{\beta}_3-1\right)'\left[\operatorname{var}\left(\widehat{\beta}_2+\widehat{\beta}_3\right)\right]^{-1}\left(\widehat{\beta}_2+\widehat{\beta}_3-1\right)}{1}=\frac{\left(\widehat{\beta}_2+\widehat{\beta}_3-1\right)^2}{\operatorname{var}\left(\widehat{\beta}_2+\widehat{\beta}_3\right)}$$

que segue $\sim F(1,n-k)$.

- Tirando a raiz quadrada, temos:

$$\frac{\widehat{\beta}_2 + \widehat{\beta}_3 - 1}{d.p.\left(\widehat{\beta}_2 + \widehat{\beta}_3\right)} \sim t(n-k) \tag{17}$$

Caso Geral

- Generalizando, testamos a hipótese nula $H_0\colon \beta_2 = \beta_3 = \cdots = \beta_k = \mathbf{0}$.

- Assim,

$$\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c} = \begin{bmatrix} \widehat{\beta}_2 \\ \widehat{\beta}_3 \\ \vdots \\ \widehat{\beta}_k \end{bmatrix} = \widehat{\boldsymbol{\beta}} \tag{18}$$

- Desse modo,

$$
\begin{aligned}
& s^2 \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' = \\
& = s^2 \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} c_{11} & c_{21} & c_{31} & \cdots & c_{k1} \\ c_{12} & c_{22} & c_{32} & \cdots & c_{k2} \\ c_{13} & c_{23} & c_{33} & \cdots & c_{k3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{1k} & c_{2k} & c_{3k} & \cdots & c_{kk} \end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \\
& = s^2 \begin{bmatrix} c_{22} & \cdots & c_{k2} \\ \vdots & . & \vdots \\ c_{2k} & \vdots & c_{kk} \end{bmatrix} \\
& = s^2 \boldsymbol{C}
\end{aligned} \tag{19}
$$

18/40

- A estatística de teste é dada por:

$$F = \frac{\widehat{\boldsymbol{\beta}}'(s^2\boldsymbol{C})^{-1}\widehat{\boldsymbol{\beta}}}{k-1} \sim F(k-1, n-k) \qquad (20)$$

ou

$$F = \frac{\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}}}{(s^2\boldsymbol{C})(k-1)} \sim F(k-1, n-k) \qquad (21)$$

Omissão de Variáveis Relevantes

- Suponha que o modelo verdadeiro seja

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e} \tag{22}$$

- Porém o modelo estimado foi

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{e} \tag{23}$$

  ou seja, estima-se um modelo caracterizado pela omissão de um conjunto de variáveis relevantes $\boldsymbol{X}_2$.
- O estimador de mínimos quadrados para eq. (23) é dado por:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y} \\
&= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}) \\
&= \boldsymbol{\beta}_1 + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2 + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{e}
\end{aligned} \tag{24}$$

- Tomando o valor esperado, temos:

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2 \tag{25}$$

- Vemos que $\widehat{\boldsymbol{\beta}}_1$ será viesado. Isso só não ocorre se $\boldsymbol{X}_1\boldsymbol{X}_2 = \mathbf{0}$. Isto é, se os regressores incluídos no modelo sejam ortogonais aos regressores omitidos.
- **Se algumas variável relevante for omitida do modelo, e se a correlação dessa variável com as variáveis incluídas no modelo não for zero, então o estimador de Mínimos Quadrados será viesado**.
- Na prática, é improvável que os regressores sejam ortogonais, de forma que deve-se esperar que a omissão de variáveis relevantes gere estimativas viesadas.

- O sinal do viés será determinado pelo sinal da covariância entre $\boldsymbol{X}_1'\boldsymbol{X}_2$ e $\boldsymbol{\beta}_2$. Temos as seguintes possibilidades:
  1. Se $\boldsymbol{X}_1'\boldsymbol{X}_2 > 0$ e $\boldsymbol{\beta}_2 > 0$: o sinal do viés será positivo;
  2. Se $\boldsymbol{X}_1'\boldsymbol{X}_2 < 0$ e $\boldsymbol{\beta}_2 < 0$: o sinal do viés será positivo;
  3. Se $\boldsymbol{X}_1'\boldsymbol{X}_2 > 0$ e $\boldsymbol{\beta}_2 < 0$: o sinal do viés será negativo;
  4. Se $\boldsymbol{X}_1'\boldsymbol{X}_2 < 0$ e $\boldsymbol{\beta}_2 > 0$: o sinal do viés será negativo;

Viés de Omissão e Variância

- O que acontece com a **variância do estimador de mínimos quadrados** ao omitirmos variáveis relevantes?
- A variância do estimador de mínimos quadrados no modelo dado pela eq. (23) é dada por:

$$\text{var}\left(\widehat{\boldsymbol{\beta}}_1\right) = \sigma^2 \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \tag{26}$$

- Caso fosse estimado o modelo verdadeiro dado pela eq. (22), a variância seria:

$$\begin{aligned} \text{var}\left(\widehat{\boldsymbol{\beta}}_{1,2}\right) &= \sigma^2 \left(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\right)^{-1} \\ &= \sigma^2 \left[\boldsymbol{X}_1'\boldsymbol{X}_1 - \boldsymbol{X}_1'\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1\right]^{-1} \end{aligned} \tag{27}$$

em que $\boldsymbol{M}_2 = \boldsymbol{I} - \boldsymbol{X}_2\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_2'$.

- Podemos comparar as duas matrizes de variância comparando as suas inversas.

$$\text{var}\left(\widehat{\boldsymbol{\beta}}_1\right) - \text{var}\left(\widehat{\boldsymbol{\beta}}_{1.2}\right) = \left(\frac{1}{\sigma^2}\right) \boldsymbol{X}_1' \boldsymbol{X}_2 \left(\boldsymbol{X}_2' \boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2' \boldsymbol{X}_1 \quad (28)$$

  que é uma matriz positiva definida. O $\widehat{\boldsymbol{\beta}}_1$ é viesado, porém ele possui menor variância do que $\widehat{\boldsymbol{\beta}}_{1.2}$.
- A inversa da variância de $\widehat{\boldsymbol{\beta}}_1$ é maior que a inversa da variância de $\widehat{\boldsymbol{\beta}}_{1.2}$. Ou seja, a variância de $\widehat{\boldsymbol{\beta}}_1$ é menor que a variância de $\widehat{\boldsymbol{\beta}}_{1.2}$.
- Outro problema diz respeito à estimação de σ^2 que é necessária para realização de inferências (teste de hipótese e intervalos de confiança). O estimador usual é dado por:

$$s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k} \quad (29)$$

- Podemos mostrar que esse estimador também é viesado. Note que

$$\boldsymbol{e}_1 = \boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}) = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{M}_1\boldsymbol{e} \tag{30}$$

- Para obter o valor esperado de $\boldsymbol{e}_1'\boldsymbol{e}_1$, usamos a mesma abordagem anterior e consideramos $\mathbb{E}[\boldsymbol{X}_1'\boldsymbol{e}] = 0$. Assim obtemos:

$$\begin{aligned} \mathbb{E}[\boldsymbol{e}_1'\boldsymbol{e}_1] &= \boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \sigma^2 \operatorname{tr}(\boldsymbol{M}_1) \\ &= \boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + (n-k)\sigma^2 \end{aligned} \tag{31}$$

- O primeiro termo é positivo, tal que s^2 é viesado para cima, não sendo possível estimar σ^2. Portanto, não é possível testar hipóteses sobre o vetor de coeficientes $\boldsymbol{\beta}_1$.
- **Conclusão**: se omitirmos variáveis relevantes da regressão, então obtemos estimativas para $\boldsymbol{\beta}_1$ e σ^2 viesadas!

- Suponha que o modelo verdadeiro seja

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{e} \tag{32}$$

- Porém o modelo estimado foi

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e} \\ &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}\end{aligned} \tag{33}$$

em que $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_1 & \boldsymbol{X}_2\end{bmatrix}$ e $\boldsymbol{\beta} = \begin{bmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{bmatrix}$.

- Nesse caso, o estimador de mínimos quadrados é dado pela fórmula usual, e **não é viesado**:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e} \end{aligned} \tag{34}$$

- Note que

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix} \tag{35}$$

- Da mesma forma, a variância do estimador de mínimos quadrados é dada pela fórmula usual e **também não é viesada**.
- Tais resultados parecem indicar que a inclusão de variáveis irrelevantes não causa nenhum problema de estimação. Porém, **essa conclusão está errada**!!

- É importante pensar que temos duas situações:
    1. Estimar um modelo omitindo uma variável relevante equivale a impor uma restrição falsa (restrição de que o coeficiente da variável é zero).
    2. Estimar um modelo incluindo uma variável irrelevante equivale a deixar de impor uma restrição verdadeira (restrição de que o coeficiente da variável é zero).
- O custo da segunda situação é a **perda de precisão da estimação**. Conforme vimos, a variância do estimador de mínimos quadrados aumenta com a inclusão de novas variáveis explicativas.

Variável Proxy

- Considere um modelo que assume um efeito aditivo da variável omitida dado por:

$$\mathbb{E}\left(y|x_1, x_2, \ldots, x_k, q\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \gamma q \tag{36}$$

em que q é um fator omitido. Estamos interessados em β_j que são os efeitos parciais das variáveis explicativas observáveis.

- Observando a eq. (36) como um modelo estrutural, podemos escrever na forma de erro como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \gamma q + \nu \tag{37}$$

$$\mathbb{E}(\nu|x_1, x_2, ..., x_k, q) = 0 \tag{38}$$

em que ν é o **erro estrutural**.

- Uma forma de manusear o aspecto não observável de q é incorporar ele dentro do termo de erro. Assim, podemos reescrever a eq. (37) como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{39}$$
$$u = \gamma q + \nu \tag{40}$$

- O termo de erro u consiste de duas partes. Sob a eq. (38), ν tem média zero e é não correlacionado com $x_1, x_2, ..., x_k, q$. Pela normalização, q também tem média zero.
- Assim, $\mathbb{E}(u) = 0$. Mas u é não correlacionado com x_j apenas se q é não correlacionado com x_j. Se q for correlacionado com algum dos regressores, o erro u também será, e teremos um problema de endogeneidade. Não podemos esperar estimativas consistentes de mínimos quadrados para β_j.
- Assim, o estimador de mínimos quadrados na presença de variável omitida é inconsistente ou estimador de mínimos quadrados é viesado.

- O viés de variável omitida pode ser eliminado ou no mínimo mitigado se é possível usar uma **variável *proxy*** par a variável não observada q.
- Para o uso de uma **variável *proxy***, duas condições devem ser cumpridas:
  - A **variável *proxy*** deveria ser redundante (irrelevante) na equação estrutural;
  - A correlação entre a variável omitida q e cada $\boldsymbol{x}$ deve ser zero depois de controlarmos para z.

Condição (1)

Se z é uma variável proxy para q, então devemos esperar a redundância de z na eq. (41)*:*

$$\mathbb{E}(y|\boldsymbol{x}, q, z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \gamma q + \delta z \tag{41}$$

isto é,

$$\mathbb{E}(y|\boldsymbol{x}, q, z) = \mathbb{E}(y|\boldsymbol{x}, q) \tag{42}$$

- Essa condição é fácil de interpretar: z é irrelevante para explicar y no sentido de média condicional, uma vez que $\boldsymbol{x}$ e q tenham sido utilizados. Embora seja difícil de verificar isso, não podemos ir muito longe sem essa condição!

Condição (2)

É uma condição mais complicada. Requer que a correlação entre a variável omitida q e cada x_j seja zero uma vez que retiramos z. Podemos ver isso em termos da projeção linear:

$$\mathbb{P}(q|1, x_1, x_2, ..., x_k, z) = \mathbb{P}(q|z) \tag{43}$$

É útil ver essa relação em termos de uma equação com o componente de erro não observado. Considere q como uma função linear de z e um termo de erro como:

$$q = \theta_0 + \theta_1 z + r, \tag{44}$$

*em que, por definição, $\mathbb{E}(r) = 0$ e a $\text{cov}(z, r) = 0$ por conta de que $\theta_0 + \theta_1 z$ é uma projeção linear de **q** em **1**, z.*

- Se z é uma *proxy* razoável para q, $\theta_1 \neq 0$. Porém, a condição é mais forte: ela é equivalente a

$$\operatorname{cov}(x_j, r) = 0 \tag{45}$$

para $j = 1, 2, .., k$.

- Esta condição requer que z seja próximo o suficiente e relacionado a q tal que, uma vez ele incluído, os x_j não são parcialmente correlacionados com q.
- A definição de **variável *proxy*** aqui não é universal. Enquanto assumimos que a **variável *proxy*** está satisfazendo a condição de redundância (41), nem sempre assumimos satisfazer a segunda condição.

- Para obter uma equação estimável, substituímos q na eq. (44) com a eq. (41) para obter:

$$y = (\beta_0 + \gamma\theta_0) + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \gamma\theta_1 z + (\gamma r + \nu) \tag{46}$$

- Sob o pressuposto feito, o termo de erro composto $u = \gamma r + \nu$ é não correlacionado com x_j para todo j; redundância de z significa que z é não correlacionado com ν e, por definição, z é não correlacionado com r.
- Disso teremos que: a regressão de mínimos quadrados de y em $1, x_1, x_2, \ldots, x_k, z$ produz estimadores consistentes de $(\beta_0+\gamma\theta_0)$, $\beta_1, \beta_2, \ldots, \beta_k$ e $\gamma\theta_1$.
- Assim, podemos estimar os efeitos parciais de cada x_j sob o pressuposto da **variável *proxy***.

Variável *Proxy* Imperfeita

- Quando z é uma variável *proxy* imperfeita, então r na eq. (46) é correlacionada com um ou mais dos x_j. Assim, a projeção linear passa a ser:

$$q = \theta_0 + \rho_1 x_1 + \rho_2 x_2 + \cdots + \rho_k x_k + \theta_1 z + r, \tag{47}$$

  a regressão da variável *proxy* resulta em $\operatorname{plim} \widehat{\beta}_j = \beta_j + \gamma\rho_j$. O estimador de mínimos quadrados com uma variável *proxy* imperfeita é inconsistente.

- A esperança é que ρ_j seja menor do que se z fosse omitido da projeção linear.

Variável *proxy* imperfeita

- Se a inclusão da variável *proxy* z induz a uma substancial colinearidade, deve ser melhor usar o estimador de mínimos quadrados sem a variável *proxy*.
- É importante considerar que incluir z reduz a variância do erro se $\theta_1 \neq 0$: $\text{var}(\gamma r + \nu) < \text{var}(\gamma q + \nu)$ porque $\text{var}(r) < \text{var}(q)$, e ν é não correlacionado com r e q.
- **Incluir uma variável *proxy* pode reduzir a variância assintótica bem como mitigar o viés.**

Exemplo

- Vamos aplicar o método da variável *proxy* para estimar o modelo estrutural

$$\begin{aligned}\log(\text{wage}) =& \beta_0 + \beta_1\text{exper} + \beta_2\text{tenure} + \beta_3\text{casado} + \beta_4\text{sul} + \\ & \beta_5\text{urbana} + \beta_6\text{black} + \beta_7\text{escolaridade} + \gamma\text{habil} + \nu\end{aligned}$$

- Assumimos que QI satisfaz o pressuposto de variável *proxy*: em um modelo de projeção linear

$$\text{habil} = \theta_0 + \theta_1\text{QI} + r$$

em que $\mathbb{E}(r) = 0$ e $\mathbb{E}(\text{QI}, r) = 0$. Assumimos que r seja não correlacionado com todos os regressores da eq. de salários, $\mathbb{E}(r|\boldsymbol{x}) = 0$.

- A equação estimada **sem a variável *proxy* QI** é dada por

$$
\begin{aligned}
\log wage = & \underset{(0.11)}{5.40} + \underset{(0.003)}{0.014}\text{exper} + \underset{(0.002)}{0.012}\text{tenure} + \underset{(0.039)}{0.199}\text{casado} \\
& - \underset{(0.026)}{0.091}\text{sul} + \underset{(0.027)}{0.184}\text{urbana} - \underset{(0.38)}{0.188}\text{black} \\
& + \underset{(0.008)}{0.065}\text{escolaridade}
\end{aligned}
$$

- A equação estimada **com a variável *proxy* QI** é dada por

$$
\begin{aligned}
\log wage = & \underset{(0.13)}{5.18} + \underset{(0.003)}{0.014}\text{exper} + \underset{(0.002)}{0.011}\text{tenure} + \underset{(0.039)}{0.200}\text{casado} \\
& - \underset{(0.026)}{0.080}\text{sul} + \underset{(0.027)}{0.182}\text{urbana} - \underset{(0.39)}{0.143}\text{black} \\
& + \underset{(0.007)}{0.054}\text{escolaridade} + \underset{(0.0010)}{0.0036}\text{QI}
\end{aligned}
$$

Econometria I

Teste de Hipóteses Lineares

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024