Econometria I

Regressão Particionada

Victor Oliveira
Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023

Sumário

Modelo de Projeção Linear

- O modelo de projeção linear pode ser representado pelas seguintes equações:

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon \tag{1}$$

$$\mathbb{E}(\boldsymbol{x}\varepsilon) = \boldsymbol{0} \tag{2}$$

$$\boldsymbol{\beta} = \left[\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')\right]^{-1}\mathbb{E}(\boldsymbol{x}y) \tag{3}$$

Preditor da Variância

- Definimos a variância do erro como

$$\sigma^2 = \mathbb{E}\left(y - \boldsymbol{x}'\boldsymbol{\beta}\right)^2 = \mathbb{E}\left(\varepsilon^2\right) \tag{4}$$

- Considerando $\boldsymbol{Q}_{yy} = \mathbb{E}\left(y^2\right)$ e $\boldsymbol{Q}_{y\boldsymbol{x}} = \mathbb{E}(y\boldsymbol{x}')$ podemos escrever σ^2 como:

$$\begin{aligned}
\sigma^2 &= \mathbb{E}(y - \boldsymbol{x}'\boldsymbol{\beta})^2 \\
&= \mathbb{E}\left(y^2\right) - 2\mathbb{E}\left(y\boldsymbol{x}'\right)\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\boldsymbol{\beta} \\
&= \boldsymbol{Q}_{yy} - 2\boldsymbol{Q}_{y\boldsymbol{x}}\boldsymbol{Q}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}y} + \boldsymbol{Q}_{y\boldsymbol{x}}\boldsymbol{Q}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{Q}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}y} \\
&= \boldsymbol{Q}_{yy} - \boldsymbol{Q}_{y\boldsymbol{x}}\boldsymbol{Q}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}y} \overset{\text{def}}{=} \boldsymbol{Q}_{yy.\boldsymbol{x}} \qquad (5)
\end{aligned}$$

em que $\boldsymbol{Q}_{yy.\boldsymbol{x}} = \boldsymbol{Q}_{yy} - \boldsymbol{Q}_{y\boldsymbol{x}}\boldsymbol{Q}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\boldsymbol{Q}_{\boldsymbol{x}y}$ é igual a variância do erro da projeção linear de y em $\boldsymbol{x}$.

Projeção

- Às vezes é útil e didático escrever a equação de projeção linear na forma:

$$y = \alpha + \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon \tag{6}$$

em que α é o intercepto e $\boldsymbol{x}$ não contém uma constante.

- Tirando a esperança dessa equação, temos:

$$\mathbb{E}[y] = \mathbb{E}[\alpha] + \mathbb{E}[\boldsymbol{x}'\boldsymbol{\beta}] + \mathbb{E}[\varepsilon] \tag{7}$$

ou

$$\mu_y = \alpha + \mu_{\boldsymbol{x}}'\boldsymbol{\beta} \tag{8}$$

em que $\mu_y = \mathbb{E}[y]$, $\mu_{\boldsymbol{x}} = \mathbb{E}[\boldsymbol{x}]$ e $\mathbb{E}[\varepsilon] = 0$.

- Rearranjando, obtemos:

$$\alpha = \mu_y - \boldsymbol{\mu}_{\boldsymbol{x}}'\boldsymbol{\beta} \tag{9}$$

- Subtraindo (6) de (8), temos:

$$\begin{aligned} y - \mu_y &= (\alpha - \alpha) + (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})'\boldsymbol{\beta} + \varepsilon \\ y - \mu_y &= (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})'\boldsymbol{\beta} + \varepsilon \end{aligned} \tag{10}$$

- A eq. (10) é uma equação linear entre as variáveis centradas $y - \mu_y$ e $(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})$. Como $(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})$ é não correlacionado com o termo de erro ε, a eq. (10) é também uma projeção linear.

- Pela fórmula da projeção linear, temos:

$$\boldsymbol{\beta} = \left(\mathbb{E}\left[(\boldsymbol{x} - \mu_{\boldsymbol{x}}) (\boldsymbol{x} - \mu_{\boldsymbol{x}})' \right]^{-1} \mathbb{E}\left[(\boldsymbol{x} - \mu_{\boldsymbol{x}}) (y - \mu_{\boldsymbol{y}}) \right] \right)$$

$$\boldsymbol{\beta} = \mathrm{var}[\boldsymbol{x}]^{-1} \mathrm{cov}(\boldsymbol{x}, y) \tag{11}$$

como uma função das covariâncias de $\boldsymbol{x}$ e y.

Observação

A matriz de covariância entre vetores $\boldsymbol{x}$ *e* $\boldsymbol{z}$ *é:* $cov(\boldsymbol{x}, \boldsymbol{z}) = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})](\boldsymbol{z} - \mathbb{E}\boldsymbol{z})'$.

Observação

A matriz de covariância do vetor $\boldsymbol{x}$ *é:* $var(\boldsymbol{x}) = cov(\boldsymbol{x}, \boldsymbol{x}) = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})](\boldsymbol{x} - \mathbb{E}\boldsymbol{x})'$.

- No modelo de projeção linear

$$y = \alpha + \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon \tag{12}$$

temos

$$\alpha = \mu_y - \mu_{\boldsymbol{x}}^{'}\boldsymbol{\beta} \tag{13}$$

e

$$\boldsymbol{\beta} = \mathrm{var}[\boldsymbol{x}]^{-1}\mathrm{cov}(\boldsymbol{x}, y) \tag{14}$$

Regressão Particionada

- Suponha que os vetores são particionados como

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \tag{15}$$

- Podemos reescrever a projeção de y em $\boldsymbol{x}$ como

$$\begin{aligned} y &= \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon \\ &= \boldsymbol{x}_1'\beta_1 + \boldsymbol{x}_2'\beta_2 + \varepsilon \end{aligned} \tag{16}$$

em que

$$\mathbb{E}(\boldsymbol{x}\varepsilon) = \mathbf{0} \tag{17}$$

- Podemos particionar $\boldsymbol{Q_{xx}}$ em conformidade com $\boldsymbol{x}$, tal que

$$\boldsymbol{Q_{xx}} = \begin{bmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{bmatrix} = \begin{bmatrix} \mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{x}_1'\right) & \mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{x}_2'\right) \\ \\ \mathbb{E}\left(\boldsymbol{x}_2\boldsymbol{x}_1'\right) & \mathbb{E}\left(\boldsymbol{x}_2\boldsymbol{x}_2'\right) \end{bmatrix} \tag{18}$$

- De forma similar $\boldsymbol{Q_{xy}}$

$$\boldsymbol{Q_{xy}} = \begin{bmatrix} \boldsymbol{Q}_{1y} \\ \boldsymbol{Q}_{2y} \end{bmatrix} = \begin{bmatrix} \mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{y}\right) \\ \mathbb{E}\left(\boldsymbol{x}_2\boldsymbol{y}\right) \end{bmatrix} \tag{19}$$

- Pela fórmula da inversão de matriz particionada temos[1]:

$$
\boldsymbol{Q}_{\boldsymbol{xx}}^{-1} = \begin{bmatrix} \boldsymbol{Q}_{11.2}^{-1} & -\boldsymbol{Q}_{11.2}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1} \\ -\boldsymbol{Q}_{22.1}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{Q}_{22.1}^{-1} \end{bmatrix} \tag{20}
$$

em que $\boldsymbol{Q}_{11.2} \stackrel{\text{def}}{=} \boldsymbol{Q}_{11} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$ e $\boldsymbol{Q}_{22.1} \stackrel{\text{def}}{=} \boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}$.

- Assim,

$$
\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \tag{21}
$$

[1] Segue da fórmula de Sherman-Morrison.

- Usando a notação da matriz particionada, temos:

$$
\begin{aligned}
\boldsymbol{\beta} &= \begin{bmatrix} \boldsymbol{Q}_{11.2}^{-1} & -\boldsymbol{Q}_{11.2}^{-1}\boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1} \\ -\boldsymbol{Q}_{22.1}^{-1}\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} & \boldsymbol{Q}_{22.1}^{-1} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{Q}_{1y} \\ \boldsymbol{Q}_{2y} \end{bmatrix} \\
&= \begin{pmatrix} \boldsymbol{Q}_{11.2}^{-1}(\boldsymbol{Q}_{1y} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{2y}) \\ \boldsymbol{Q}_{22.1}^{-1}(\boldsymbol{Q}_{2y} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{1y}) \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{Q}_{11.2}^{-1}\boldsymbol{Q}_{1y.2} \\ \boldsymbol{Q}_{22.1}^{-1}\boldsymbol{Q}_{2y.1} \end{pmatrix} \qquad (22)
\end{aligned}
$$

- Assim, mostramos que

$$\boldsymbol{\beta}_1 = \boldsymbol{Q}_{11.2}^{-1}\boldsymbol{Q}_{1y.2} \qquad (23)$$

$$\boldsymbol{\beta}_2 = \boldsymbol{Q}_{22.1}^{-1}\boldsymbol{Q}_{2y.1} \qquad (24)$$

em que $(\boldsymbol{Q}_{1y} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{2y}) = \boldsymbol{Q}_{1y.2}$ e $(\boldsymbol{Q}_{2y} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{1y}) = \boldsymbol{Q}_{2y.1}$

Teorema de Frisch-Waugh-Lovell

- Como fazer análise *ceteris paribus* num contexto de regressão múltipla?
- Considere o caso

$$y = x_1\beta_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + e \tag{25}$$

- Considere a projeção de x_1 em $\boldsymbol{x}_2$:

$$x_1 = \boldsymbol{x}_2'\boldsymbol{\gamma}_2 + u_1 \tag{26}$$

$$\mathbb{E}(\boldsymbol{x}_2 u_1) = \boldsymbol{0} \tag{27}$$

em que $\gamma = \boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$ e $\mathbb{E}[u_1^2] = \boldsymbol{Q}_{11.2} = \boldsymbol{Q}_{11} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$.

- Podemos calcular

$$\begin{aligned}\mathbb{E}(u_1 y) &= \mathbb{E}((x_1 - \boldsymbol{\gamma}_2' \boldsymbol{x}_2) y) \\ &= \mathbb{E}(x_1 y) - \boldsymbol{\gamma}_2' \mathbb{E}(\boldsymbol{x}_2 y) \\ &= \boldsymbol{Q}_{1y} - \boldsymbol{Q}_{11} - \boldsymbol{Q}_{12} \boldsymbol{Q}_{22}^{-1} \boldsymbol{Q}_{21} \\ &= \boldsymbol{Q}_{1y.2} \end{aligned} \tag{28}$$

- De modo que

$$\beta_1 = \boldsymbol{Q}_{11.2}^{-1} \boldsymbol{Q}_{1y.2} = \frac{\mathbb{E}(u_1 y)}{\mathbb{E}(u_1^2)} \tag{29}$$

é o coeficiente da regressão de y em u_1. Em uma projeção múltipla, o coeficiente β_1 é igual a coeficiente da projeção de uma regressão de y em u_1. Sendo que u_1 é o erro da regressão de x_1 em x_2.

- As aplicações deste resultado incluem o seguinte:
  - Se X_2 é uma matriz de variáveis *dummy* sazonais, vemos que usar FWL equivale a ajustar sazonalmente y e os demais regressores.
    - Uma crítica a isso é que esse tipo de ajuste sazonal é muito grosseiro, pois não presta atenção à tendência e aos componentes cíclicos dos dados.
  - Se X_2 tiver apenas uma coluna, e essa variável for uma variável linear de tendência temporal, então sua inclusão equivale a remoção de tendência de todos os outros regressores, e y da mesma maneira.
  - Se X_2 tem apenas uma coluna, e essa coluna é o intercepto, então sua inclusão é equivalente a expressar y e as colunas de X_1 em termos de desvios sobre suas médias amostrais, e então estimar o modelo sem o intercepto.

Análise de Sensibilidade

- Cinelli e Hazlett (2020) desenvolvem uma série de ferramentas para os pesquisadores realizarem análises de sensibilidade em modelos de regressão, usando uma extensão da estrutura de viés de variável omitida. Para fazer isso, eles usam o FWL para motivar esse viés. Suponha que o modelo de regressão completo seja especificado como:

$$Y = \tau D + X\beta + \gamma Z + \varepsilon \tag{30}$$

em que τ, β e γ são os coeficientes de regressão estimados, D é a variável de tratamento, X são covariáveis observadas e Z são covariáveis não observadas.

- Como Z não é observado, os pesquisadores mensuram:

$$Y = \tau_{obs} D + X\beta_{obs} + \varepsilon_{obs} \tag{31}$$

- Por FWL, sabemos que τ_{obs} é equivalente ao coeficiente de regressão do resultado residualizado (em relação a X), no resultado residualizado de D (novamente em relação a X). Chame esses dois resíduos de Y_r e D_r.
- E lembre-se que o modelo de regressão para o estágio final das regressões parciais é bivariado ($Y_r \sim D_r$). Convenientemente, um coeficiente de regressão bivariada pode ser expresso em termos da covariância entre as variáveis do lado esquerdo e do lado direito:

$$\tau_{obs} = \frac{\text{cov}(Y_r, D_r)}{\text{var}(D_r)} \tag{32}$$

- Observe que dado o modelo de regressão completo na eq (30), o resultado parcial Y_r na verdade é composto pelos elementos $\tau D_r + \gamma Z_r$, e assim:

$$\tau_{obs} = \frac{\text{cov}(\tau D_r + \gamma Z_r, D_r)}{\text{var}(D_r)} \tag{33}$$

- Em seguida, podemos expandir a covariância usando a regra de expectativa que $\text{cov}(A, B + C) = \text{cov}(A, B) + \text{cov}(A, C)$ e dado que τ e γ são escalares, podemos movê-los para fora das funções de covariância:

$$\begin{aligned}
\tau_{obs} &= \frac{\tau\text{cov}(D_r, D_r) + \gamma\text{cov}(Z_r, D_r)}{\text{var}(D_r)} \\
&= \frac{\tau\text{var}(D_r) + \gamma\text{cov}(Z_r, D_r)}{\text{var}(D_r)} \\
&= \tau + \gamma\frac{\text{cov}(Z_r, D_r)}{\text{var}(D_r)}
\end{aligned} \tag{34}$$

- Frisch-Waugh é tão útil porque simplifica uma equação multivariada em uma bivariada.
- Embora computacionalmente isso não faça diferença, aqui nos permite usar uma expressão conveniente do coeficiente bivariado para mostrar e quantificar o viés quando você executa uma regressão na presença de um fator de confusão não observado.

Viés de Variável Omitida

- Considere que o modelo correto possui os vetores particionados na equação de regressão que é dada por:

$$y = \boldsymbol{x}_1'\beta_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + e \tag{35}$$

- Mas ao invés de estimarmos a eq (35), estamos estimando a projeção de y apenas em $\boldsymbol{x}_1$. Isso pode feito se considerarmos que a variável $\boldsymbol{x}_2$ não foi observada.

$$y = \boldsymbol{x}_1'\gamma_1 + u \tag{36}$$

$$\mathbb{E}(\boldsymbol{x}_1 u) = \boldsymbol{0} \tag{37}$$

- Observe que na eq (36) escrevemos o coeficiente γ_1 ao invés de β_1 e o termo de erro u ao invés de e. Isso porque a eq (35) é diferente da eq. (36).

- Para vermos que $\beta_1 \neq \gamma_1$, calculamos:

$$
\begin{aligned}
\gamma_1 &= \left[\mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{x}_1'\right)\right]^{-1}\mathbb{E}(\boldsymbol{x}_1 y) \\
&= \left[\mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{x}_1'\right)\right]^{-1}\mathbb{E}\left(\boldsymbol{x}_1\left(\boldsymbol{x}_1'\beta_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + e\right)\right) \\
&= \boldsymbol{\beta}_1 + \left[\mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{x}_1'\right)\right]^{-1}\mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{x}_2'\right)\boldsymbol{\beta}_2 \\
&= \boldsymbol{\beta}_1 + \boldsymbol{\Gamma}\boldsymbol{\beta}_2
\end{aligned}
\tag{38}
$$

  em que $\boldsymbol{\Gamma} = \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}$ é o coeficiente da matriz de projeção de $\boldsymbol{x}_2$ em $\boldsymbol{x}_1$.

- Note que $\gamma_1 = \boldsymbol{\beta}_1 + \boldsymbol{\Gamma}\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$ a menos que $\boldsymbol{\Gamma} = \mathbf{0}$ ou $\boldsymbol{\beta}_2 = 0$.

- A diferença $\boldsymbol{\Gamma}\boldsymbol{\beta}_2$ entre γ_1 e β_1 é conhecida como **viés de variável omitida**. Uma forma de evitar esse viés é incluir no modelo todas as variáveis relevantes.

- Considere a equação estrutural de salários como

$$\log(\text{salário}) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + \gamma habil + u \tag{39}$$

- Como a variável *habil* não é observada, estimamos o modelo:

$$\log(\text{salário}) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + v \tag{40}$$

em que $v = \gamma habil + u$. O estimador de β_3 da regressão de salários sobre escolaridade estamos chamando de $\tilde{\beta}_3$.

- Em termos algébricos $\tilde{\beta}_3 = \widehat{\beta}_3 + \widehat{\gamma}\tilde{\delta}_1$, em que $\widehat{\beta}_3$ e $\widehat{\gamma}$ serão os estimadores de inclinação da regressão.

Econometria I

Regressão Particionada

Victor Oliveira
Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023