Mestrado Profissional em Avaliação e Monitoramento de Políticas Públicas

Métodos Quantitativos I

Aula 8: Modelos em Painel: Efeitos Fixos

Professores: Daniel Grimaldi e Arthur Bragança

3º Trimestre - 2025

Contexto

Dados em *cross-section*

- Durante as aulas de OLS, estivemos trabalhando com dados em corte transversal (*cross-section*).
- Isso equivale a um contexto em que observamos *i* indivíduos, todos em um mesmo momento do tempo.

Formalmente, nesse contexto temos modelos de regressão que podem ser representados pela Equação (1)

$$y_i = \beta_0 + \beta_1 \, x_i^1 + ... + \beta_k \, x_i^k + u_i \tag{1}$$

Dados em painel

- Dados em painel são aqueles que combinam uma dimensão transversal (i) com outra longitudinal (t).
- Isso equivale a um contexto em que observamos i indivíduos, em múltiplos momentos de tempo.
  - tipicamente, em um modelo de painel, a dimensão transveral é maior que a longitudinal.

Formalmente, nesse contexo passamos a ter um modelo como o da Equação (2).

$$y_{it} = \beta_0 + \beta_1 \ x_{it}^1 + ... + \beta_k \ x_{it}^k + u_{it} \tag{2}$$

Como dados em painel podem ser úteis?

- Já vimos que em vários contextos não conseguimos obervar todas os controles que são necessários para garantir a hipótese de exogeneidade estrita (ou *backdoor path criteria* na literatura de DAGs).
- Sempre que $E[u_i|x_i] \neq 0$, teremos um $\hat{\beta}_{ols}$ viesado - que, portanto, não recupera o efeito causal de X em y.
- Com os dados em painel podemos usar um modelo de efeitos fixos (FE), que controla por qualquer variável (observável ou não) desde que ela permaneça constante ao longo do tempo dentro uma categoria específica.

Como isso acontece?

- Considere um modelo como da Equação (3)

$$y_{it} = \beta_0 + \beta_1 \, x^1_{it} + u_{it} \tag{3}$$

- E considere que: (i) $u_{it} = a_i + \epsilon_{it}$; e (ii) $E[\epsilon_{it}] = 0$.
- Neste caso, temos que: $E[u_{it} | x^1_{it}] = E[a_i + \epsilon_{it}] \equiv E[a_i]$.
  - Então, $\hat{\beta}_{ols}$ é viesado

Como FE resolve?

- Na prática, tratamos a_i como uma variável a ser estimada. Dessa forma, a Equação (3) pode ser reescrita:

$$\begin{aligned} y_{it} =& \beta_0 + a_i + \beta_1 \; x_{it}^1 + \epsilon it \\ =& \alpha_i + \beta_1 \; x_{it}^1 + \epsilon it \end{aligned} \tag{4}$$

- Agora, temos que: (i) $E[\epsilon_{it}|a_i, x_{it}^1] = E[\epsilon_{it}] \equiv 0$
- Ou seja, ao controlarmos pelo elemento não observável α_i, recuperamos a validada da hipótese de exogeneidade estrita.

FE e a variância dos dados

- Num contexto de painel, podemos decompor a variância total dos dados em variância *within* e *betwewen*.
  - **Variância *within* (ou intragrupo)**: variância longitudinal de uma variável para um mesmo indivíduo i
  - **Variância *between* (ou intergrupo)**: variância de uma variável entre diferentes indivíduos.
- Na prática, um modelo FE expurga dos dados toda a variância *between*. Isso tem uma implicação importante: em um modelo FE não é possível estimarmos coeficientes para variáveis que sem variância longitudinal (fixas no tempo).

Transformação within

- Considere um modelo como o da Equação (4) e defina uma função $h(v)$ que calcula a média longitudinal de v para cada i. Então, temos que:

$$\begin{aligned} h(y_{it}) =& h(\alpha_i) + h(\beta_1 \; x_{it}^1) + h(\epsilon it) \\ =& \alpha_i + \beta_1 \; h(x_{it}^1) + h(\epsilon it) \end{aligned} \tag{5}$$

- Agora, podemos subtrair (5) dos dois lados de (4) e temos que:

$$\begin{aligned} y_{it} - h(y_{it}) =& \alpha_i - \alpha_i + \beta_1 \; (x_{it}^1 - h(x_{it}^1)) + \epsilon_{it} - h(\epsilon_{it}) \\ \ddot{y_{it}} =& \beta_1 \; (\ddot{X}_{it}^1) + \ddot{\epsilon}_{it} \end{aligned} \tag{6}$$

Usando dummies

- A Equação (6) pode ser estimada diretamente por meio de um MQO.
  - A vantagem é que essa estimação é computacionalmente mais simples (e rápida).
  - A desvantagem é que não estimamos *de facto* os efeitos fixos - apenas expurgamos os seus efeitos.
- Uma alternativa é estimar a Equação (4) diretamente por meio da inclusão de dummies de efeitos fixos (α_i)
  - A vantagem é que assim conseguimos estimativas interpretáveis para α_i.
  - Contudo, em casos onde existem muitas categorias, esse processo se torna computacionalmente muito demandante.

Autocorrelação e clusterização

- Uma das hipóteses para que o estimador MQO seja BLUE é que os resíduos devem ser homoscedásticos e não autocorrelacionaldos. Em um contexto de painel, isso implica que:

$$(i)\ E[\epsilon_{it}] = \sigma_\epsilon\ \forall\ i, t$$
$$(ii)\ E[\epsilon_{it}\epsilon ik] = 0\ \forall\ i, t, k$$

Autocorrelação e clusterização

Note que (ii) é altamente improvável - ambos os resíduos estão associados ao mesmo indivíduo. Por isso, usar estimativas de desvio-padrão que assumem homoscedasticidade e ausência de autocorrelação serial tendem a produzir intervalos de confiança subestimados.

- A regra geral é assumir que existe correlação entre diferentes observações de um mesmo indivíduo e corrigir o desvio-padrão estimado considerando isso. A isso chamamos de *clusterizar* o desvio-padrão (cada indivíduo passa a ser um cluster)

Painel não balanceado

- Painel não balanceado: um painel em que certos indivíduos estão ausentes em alguns períodos.
- A estimação de um modelo FE com dados não balanceados (seja por transformação within ou por inclusão de dummies) ocorre exatamente da mesma maneira de um painel balanceado, mas...
- é crucial entendermos o que gera o não balanceamento. Se a ausência de observações possuir um padrão claro, então há risco de vies e/ou perda de validade externa da estimação.

Quando FE não resolve?

- FE se dedica a controlar o viés decorrente da **omissão de variáveis fixas no tempo**.
  - Se a variável não-observável possuir tendência heterogênea (entre categorias) ao longo do tempo, mesmo o FE será viesado
  - Se o tratamento tiver efeitos heterogêneos associados à variável não heterogênea, o FE também não conseguirá recuperar causalidade...

Hands on

Autodefesa enquanto política de segurança

O exemplo de hoje é inspirado em *Cheng and Hoekstra. 2013. "Does Strengthening Self-Defense Law Deter Crime or Escalate Violence? Evidence from Expansions to Castle Doctrine." Journal of Human Resources.*

- Investigam se mudanças legislativas que facilitam autodefesa levam a redução de criminalidade
  - entre 2000 e 2010 há uma expansão de medidas estaduais flexibilizando *Castle Doctrine*;
  - mas essa expansão foi centrada em locais com características específicas

Aproximando do contexto brasileiro

- Temos um painel de municípios (i) ao longo dos anos (t) e o nível de segurança-pública é medido por um índice de mortes violentas anuais por 100 mil habitantes (y_{it})
- y_{it} depende de fatores observáveis, tais como:
  - taxa de desemprego médio (x^1_{it}); e
  - nível de desigualdade médio, medido pelo % da renda detido pelo 1% mais ricos (x^2_{it}).
- y_{it} também depende de um fator não observável (α_{it}): a intensidade de disputa territorial por parte de grupos criminosos armados
  - α_{it} causa aumento de y_{it}.

Política de armamento

- Nesse contexto, imaginemos que o governo decide criar uma política pública que permite aos municípios encorajar autodefesa por meio da facilitação no acesso a armas.
  - adesão está associada a α, de forma que populações de municípios com confrontos territoriais mais deflagrados têm medo de portar armas;
  - Tal qual em Cheng, Cheng e Mark Hoekstra (2013) política tem impacto negativo na violência, aumentando y_{it}, em $\delta_p = 1$

PGD: Formação de y_{it}

$$\alpha_i \sim N(2,2);\ \alpha_{it} = \alpha_i \forall t \tag{7}$$
$$x_i^1 \sim N(7,0.5);\ \ x_{it}^1 = x_i^1 + \upsilon_{it}^1 \tag{8}$$
$$x_i^2 \sim N(40,10);\ \ x_{it}^2 = x_i^2 + \upsilon_{it}^2 \tag{9}$$
$$y_{it} = \alpha_{it} + \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \delta P_{it}\epsilon_{it} \tag{10}$$
$$\upsilon_{it}^1 \sim N(0,1);\ \upsilon_{it}^2 \sim N(0,5);\ \epsilon_{it} \sim N(0,1) \tag{11}$$
$$\beta_1 = 0.5;\ \beta_2 = 0.1;\ \delta = 5 \tag{12}$$

PGD: Adesão a P_{it}

O acesso à política P é definido pelas equações (13) e (14).

$$S_i = \frac{1}{1 + e^{g(\alpha_{it})}} \tag{13}$$

$$g(\alpha_{it}) = \sum_T \alpha_{it} - \mu_\alpha \tag{14}$$

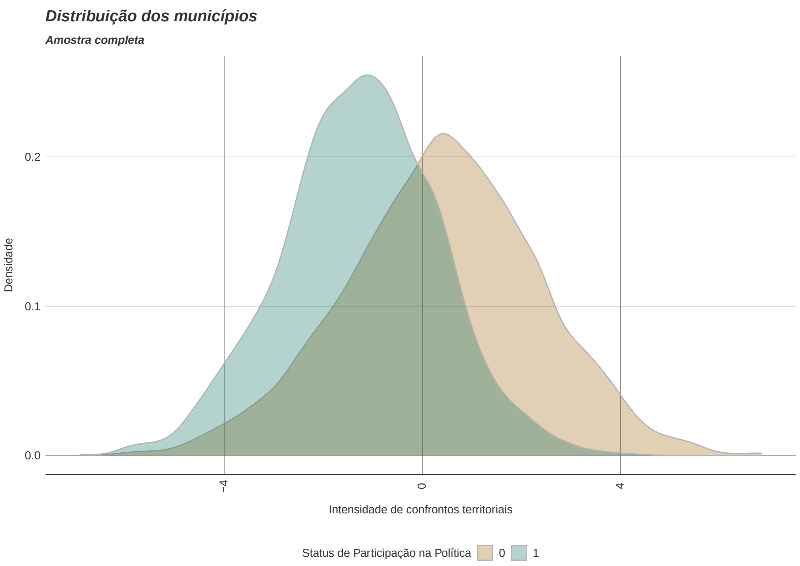
$$P_i \sim Bern(S_i) \tag{15}$$

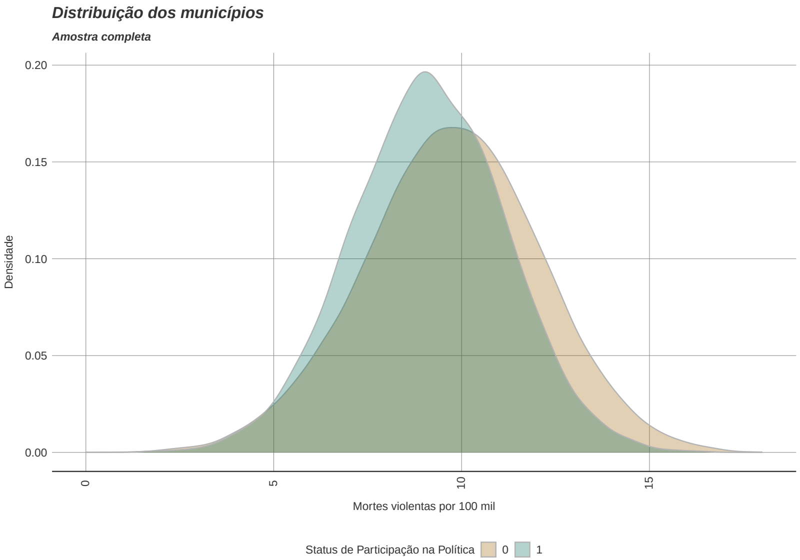
P_{it} assume valores 0 ou 1, a dependender da realização de uma Bernoulli com probabilidade de sucesso S_i. **Por simplificação, todos os indivíduos que têm acesso à política, começam a receber o tratamento no mesmo período $t_p = \frac{T}{2}$.**

Implementando PGD

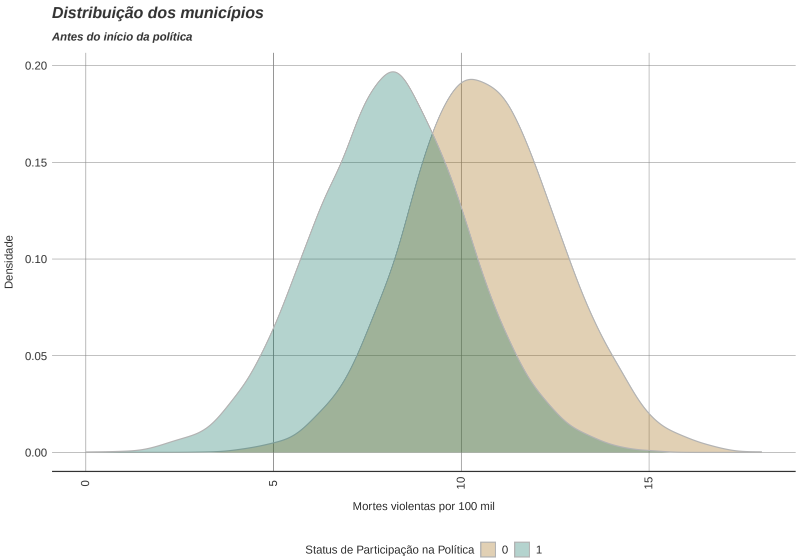
Como de costume, escrevemos uma função (denominada pgd.R) para simular esse PGD.

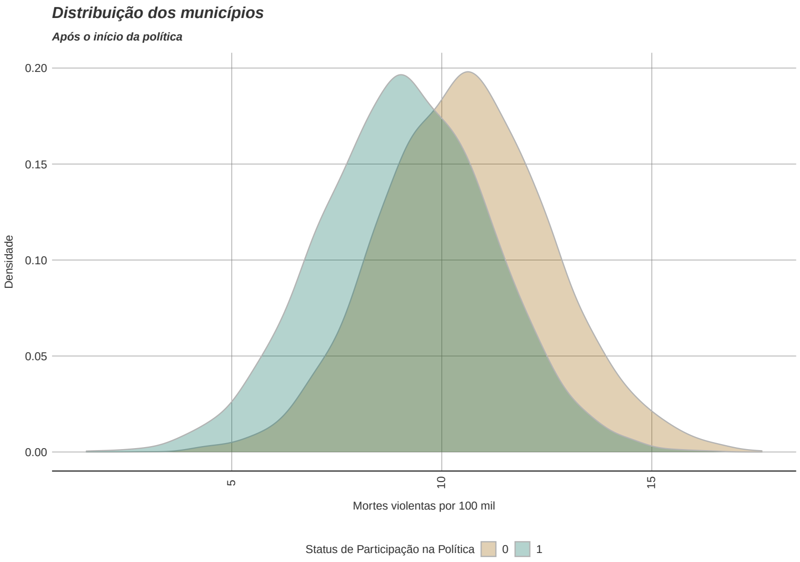
```r
source("pgd.R")
data <- pgd(n.id=5570, n.t=6, beta=c(0.5, 0.1),
            delta=1, mu_x=c(7, 40),
            sigma_x=c(0.5, 10), mu_alpha=0,
            sigma_alpha=2)
```


Distribuição dos municípios
Amostra completa
Densidade
0.2
0.1
0.0
−4
0
4
Intensidade de confrontos territoriais
Status de Participação na Política
0
1

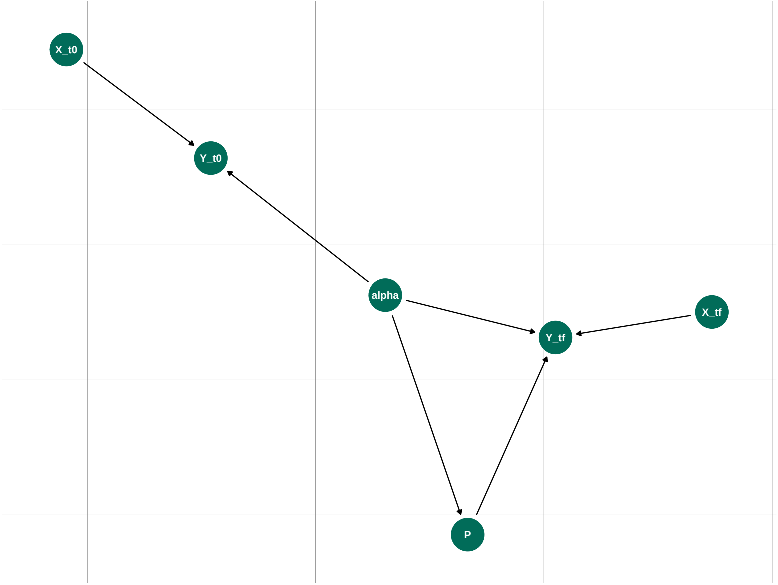


Distribuição dos municípios
Amostra completa
Densidade
0.20
0.15
0.10
0.05
0.00
0
5
10
15
Mortes violentas por 100 mil
Status de Participação na Política
0
1


Distribuição dos municípios
Antes do início da política
0.20
0.15
0.10
0.05
0.00
Densidade
0
5
10
15
Mortes violentas por 100 mil
Status de Participação na Política
0
1


Distribuição dos municípios
Após o início da política
0.20
0.15
0.10
0.05
0.00
Densidade
5
10
15
Mortes violentas por 100 mil
Status de Participação na Política
0
1

Teoria causal do Programa



MQO: Estimação

```
require(estimatr)
reg1 <- lm_robust(y ~ x1 + x2 + P, data=data)
reg2 <- lm_robust(y ~ x1 + x2 + P, data=filter(data, pos==1))
reg3 <- lm_robust(y ~ x1 + x2 + P + alpha, data=data)
```

MQO: Resultados

```r
summary(reg1)
```

```
##
## Call:
## lm_robust(formula = y ~ x1 + x2 + P, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   2.1427   0.104178   20.57  2.045e-93  1.93846   2.3468 33416
## x1            0.5058   0.009462   53.45  0.000e+00  0.48722   0.5243 33416
## x2            0.1012   0.001964   51.52  0.000e+00  0.09733   0.1050 33416
## P            -0.6319   0.024611  -25.67 5.594e-144 -0.68012  -0.5836 33416
##
## Multiple R-squared:  0.1521 ,    Adjusted R-squared:  0.1521
## F-statistic:  2044 on 3 and 33416 DF,  p-value: < 2.2e-16
```

MQO: Resultados

```r
summary(reg2)
```

```
##
## Call:
## lm_robust(formula = y ~ x1 + x2 + P, data = filter(data, pos ==
##     1))
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   2.9902   0.130353   22.94 1.107e-114  2.73471   3.2457 16706
## x1            0.5114   0.011793   43.37  0.000e+00  0.48833   0.5346 16706
## x2            0.0995   0.002432   40.91  0.000e+00  0.09473   0.1043 16706
## P            -1.4521   0.028694  -50.61  0.000e+00 -1.50831  -1.3958 16706
##
## Multiple R-squared:  0.2674 ,    Adjusted R-squared:  0.2673
## F-statistic:  2007 on 3 and 16706 DF,  p-value: < 2.2e-16
```

MQO: Resultados

```r
summary(reg3)
```

```
##
## Call:
## lm_robust(formula = y ~ x1 + x2 + P + alpha, data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   1.7624  0.0494536   35.64 4.893e-273  1.66546   1.8593 33415
## x1            0.5042  0.0044799  112.55  0.000e+00  0.49541   0.5130 33415
## x2            0.1010  0.0009361  107.88  0.000e+00  0.09915   0.1028 33415
## P             1.0147  0.0135621   74.82  0.000e+00  0.98810   1.0413 33415
## alpha         0.9965  0.0029226  340.95  0.000e+00  0.99075   1.0022 33415
##
## Multiple R-squared:  0.8103 ,    Adjusted R-squared:  0.8103
## F-statistic: 3.614e+04 on 4 and 33415 DF,  p-value: < 2.2e-16
```

FE: Estimação por dummies

```r
require(fixest)
reg1 <- feols(y ~ x1 + x2 + P | id, data)
summary(reg1)
```

```
## OLS estimation, Dep. Var.: y
## Observations: 33,420
## Fixed-effects: id: 5,570
## Standard-errors: Clustered (id)
##    Estimate Std. Error t value  Pr(>|t|)
## x1 0.499676   0.005958 83.8620 < 2.2e-16 ***
## x2 0.100114   0.001215 82.4091 < 2.2e-16 ***
## P  1.020072   0.015368 66.3751 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.916542     Adj. R2: 0.809948
##                  Within R2: 0.395514
```

FE: Estimação por transformação within

```r
data %<>%
  group_by(id) %>%
  mutate(y_mean_i = mean(y),
         x1_mean_i = mean(x1),
         x2_mean_i = mean(x2),
         P_mean_i = mean(P)) %>%
  ungroup() %>%
  mutate(y_within = y - y_mean_i,
         x1_within = x1 - x1_mean_i,
         x2_within = x2 - x2_mean_i,
         P_within = P - P_mean_i)
reg2 <- lm_robust(y_within ~ x1_within + x2_within + P_within, data=data)
```

FE: Estimação por transformação within

```r
summary(reg2)
```

```
##
## Call:
## lm_robust(formula = y_within ~ x1_within + x2_within + P_within,
##     data = data)
##
## Standard error type:  HC2
##
## Coefficients:
##              Estimate Std. Error   t value Pr(>|t|)  CI Lower CI Upper    DF
## (Intercept) 2.072e-16   0.005014 4.133e-14        1 -0.009827 0.009827 33416
## x1_within   4.997e-01   0.005471 9.133e+01        0  0.488953 0.510400 33416
## x2_within   1.001e-01   0.001108 9.033e+01        0  0.097942 0.102286 33416
## P_within    1.020e+00   0.014166 7.201e+01        0  0.992307 1.047838 33416
##
## Multiple R-squared:  0.3955 ,    Adjusted R-squared:  0.3955
## F-statistic:  7332 on 3 and 33416 DF,  p-value: < 2.2e-16
```

Obrigado!