

Panel Data

Manoel Galdino

2024-05-21

Dados em Painel - TSCS - Longitudinais

- Dados em painel, também conhecidos como dados longitudinais ou Time Series Cross Section, são conjuntos de dados que contêm observações de múltiplas entidades (como indivíduos, empresas, países, etc.) ao longo de vários períodos de tempo.

Características dos Dados em Painel

- **Dimensão transversal (Cross-sectional):** Observações de várias entidades em um único ponto no tempo.
- **Dimensão temporal (Time-series):** Observações da mesma entidade ao longo do tempo.

Vantagens dos Dados em Painel (antigo)

- 1 **Heterogeneidade Individual:** Permitem controlar a heterogeneidade não observada que varia entre as entidades, mas é constante ao longo do tempo para cada entidade.
- 2 **Dinâmica Temporal:** Capturam a dinâmica temporal e permitem estudar mudanças ao longo do tempo.
- 3 **Maior Variabilidade:** A combinação das dimensões temporal e transversal oferece mais variabilidade.
- 4 **Informação Mais Completa:** Proporcionam uma quantidade maior de dados, levando a estimativas mais precisas e robustas.

Modelos Comuns para Dados em Painel

- **Modelos de Efeitos Fixos (FE):** Consideram que as diferenças entre entidades são capturadas por interceptos específicos para cada entidade.
- **Modelos de Efeitos Aleatórios (RE):** Assumem que as diferenças entre entidades são aleatórias e não correlacionadas com as variáveis explicativas.
- **Modelos Dinâmicos:** Incorporam variáveis defasadas para capturar a

Exemplo de Código R para Análise de Dados em Paineis

```
# Pacote necessário para análise de dados em painel
```

```
library(fixest)
```

```
library(peacesciencer)
```

```
## Warning: package 'peacesciencer' was built under R version
```

```
## {peacesciencer} includes additional remote data for separat
```

```
## This message disappears on load when these data are downloa
```

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages -----
```

```
## v dplyr      1.1.2      v readr      2.1.4
```

```
## v forcats   1.0.0      v stringr   1.5.0
```

```
## v ggplot2    3.4.4      v tibble    3.2.1
```

```
## v lubridate  1.9.2      v tidyr     1.3.0
```

```
## v purrr      1.0.2
```

DiD com k períodos

- O modelo de efeitos fixos pode ser pensado como uma generalização do DID com k períodos (tratamento binário).

Modelo básico

- A política pública (tratamento) ocorre integralmente no períodos t_o ¹.
- Se tivermos múltiplos períodos pré tratamento, podemos testar algumas suposições (pré-tendências)
- Se tivermos múltiplos períodos pós-tratamento, podemos examinar o timing do efeito
 - Efeito imediato t ou $t+1$?
 - Efeito persiste no tempo ou morre?
 - Ponderar todos os períodos em uma únviavariável post estima o efeito médio. Cuidado se não houver balanceamento.

¹cf. Notas do Paul GP https://nbviewer.org/github/paulgp/applied-methods-phd/blob/main/lectures/13_dind.pdf

O que two-way fixed effect está estimando?

- One-way fixed effect por unidade: “How much does a case change in relation to itself over time?” (Kropko & Kubinec, 2020)
- One-way fixed effect por tempo: “How much does a case change relative to other cases?” (Kropko & Kubinec, 2020)

How much does a case change relative to other cases?

Ignorability versus parallel trends

- DiD depende da suposição de tendências paralelas, não de ignorability.
- Modelos de efeitos fixos podem ser defendidos como garantindo CIA
- O que está sendo suposto para termos CIA? Vamos seguir a notação e formulação de Myint (2024).

Time varying treatment

- Quando o tratamento varia no tempo de 0 até o período k , falamos em história do tratamento $\bar{T}_k = \{T_0, T_1, \dots, T_k\}$.
- Se a resposta também é medida em k períodos, um estimando de interesse (com abuso de notação) pode ser: $\mathbb{E}[Y^{t_k=1}] - \mathbb{E}[Y^{t_k=0}]$, que seria a diferença nos resultados potenciais no período t_k entre o tratamento e controle no mesmo período.
- Contudo, o que precisamos supor sobre os tratamentos em períodos anteriores e posteriores para estimar esse efeito causal?

Sequências de tratamento

A sequência de tratamento \bar{T}_k pode ser estática, no sentido de a regra atribuindo tratamento ou controle se definida antes da sequência de tratamento (digamos, sempre tratar), ou dinâmica, a depender de outra variável que muda no tempo.

Histórias

- Tipicamente, estamos interessados não em comparações em um período de tempo, mas para toda uma dada história de tratamento. Sejam \bar{d} e \bar{d}^* duas histórias distintas de tratamento.
- Então, queremos estimar $\mathbb{E}[Y_t(\bar{d})] - \mathbb{E}[Y_t(\bar{d}^*)]$.
- Se o tratamento é binário e temos k períodos, existem 2^k histórias distintas e, portanto, 2^k resultados potenciais, dos quais dois podem ser escolhidos para definir o estimando causal de interesse.

O que é o DID estimando nesse framework?

- Seja $\bar{d} = (0, 1)$ a história do grupo de tratamento e $\bar{d}^* = (0, 0)$ a história do grupo controle. O ATT é:
- $ATT = \mathbb{E}[Y_2(0, 1)] - \mathbb{E}[Y_2(0, 0) | \bar{D} = (0, 1)]$

staggered rollout (implementação escalonada)

Acontece quando o tratamento ocorre em diferentes momentos para grupos distintos. Não há “on-off”.

Sequential Ignorability

- Sequential ignorability em um experimento seria se a alocação do tratamento ou controle dependesse de valores passados de covariáveis e da resposta.
- Sequential ignorability implica exogenidade sequencial do termo de erro, que é um pressuposto de regressão com dados de painel. E o que é exogeneidade sequencial?
- Que o termo de erro é não correlacionado com o tratamento e com valores passados da resposta e do tratamento e de covariáveis. Formalmente, seja \bar{D}_k a história até k e analogamente para outras variáveis. Seja $v_{i,k-1} = g(\bar{D}_{k-1}, \bar{Y}_{k-1}, \bar{X}_{k-1})$. Então, $\mathbb{E}[e_{ik} | d_{ik}, v_{i,k-1}] = 0$.

Myint, L. (2024). Controlling time-varying confounding in difference-in-differences studies using the time-varying treatments framework. *Health Services and Outcomes Research Methodology*, 24(1), 95-111.

https://nbviewer.org/github/paulgp/applied-methods-phd/blob/main/lectures/13_dind.pdf