

## Topic 6: DID and TWFE

by Sai Zhang

**Key points:** This note is on the causal panel data, building upon [Arkhangelsky and Imbens \(2023\)](#).

**Disclaimer:** *This note is compiled by Sai Zhang.*

## 6.1 Panel Data Configurations

### 6.1.1 Data Types

#### 6.1.1.1 Panel Data

For observations on  $N$  units, indexed by  $i = 1, \dots, N$ , over  $T$  periods, indexed by  $t = 1, \dots, T$ , the outcome of interest is denoted by  $Y_{it}$ , the treatment  $W_{it}$ . These observations may themselves consist of averages over more basic units:

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & \cdots & Y_{1T} \\ \vdots & \ddots & \vdots \\ Y_{N1} & \cdots & Y_{NT} \end{pmatrix} \quad \mathbf{W} = \begin{pmatrix} W_{11} & \cdots & W_{1T} \\ \vdots & \ddots & \vdots \\ W_{N1} & \cdots & W_{NT} \end{pmatrix}$$

we may also observe exogenous variables  $X_{it}$  or  $X_i$ . Typically, we focus on a balanced panel where for all units  $i = 1, \dots, N$  we observe outcomes for all  $t = 1, \dots, T$ .

#### 6.1.1.2 Grouped Repeated Cross-Section Data

In a GRCS data, we have observations on  $N$  units, each observed only once in period  $T_i$  for unit  $i$ . Different units may be observed at different points in time,  $T_i$  typically takes on only a few values, with many units sharing the same value for  $T_i$ .

### 6.1.2 Shapes of Data Frames

#### 6.1.3 Assignment Mechanisms

## References

Dmitry Arkhangelsky and Guido Imbens. Causal models for longitudinal and panel data: A survey. Technical report, National Bureau of Economic Research, 2023.