Econometrics **April 29, 2024**

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

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

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

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 diffrent points in time, T_i typically takes on only a few values, with many units sharing the same value for T_i. The outcome Y_i and treatment W_i are indexed by the unit index i. The set of units is **partitioned** into 2 or more groups, with the group that unit i belongs to denoted by $G_i \in \mathcal{G} = \{1, 2, \cdots, G\}$.

Define the average outcomes for each group-time-period pair:

$$\overline{Y}_{gt} \equiv \frac{\sum_{i=1}^{N} \mathbf{1}_{G_i=g, T_i=t} Y_i}{\sum_{i=1}^{N} \mathbf{1}_{G_i=g, T_i=t}}$$

for treatment

$$\overline{W}_{gt} \equiv \frac{\sum_{i=1}^{N} \mathbf{1}_{G_i=g, T_i=t} W_i}{\sum_{i=1}^{N} \mathbf{1}_{G_i=g, T_i=t}}$$

then treat the $G \times T$ group averages $\overline{Y}_{gt}$ and $\overline{W}_{gt}$ as the unit of observation, then the grouped data is just a panel. The major issue in practice is that the number of groups is very small comparing to proper panel data.

6.1.1.3 Row and Column Exchangeable Data

The data are doubly indexed by $i = 1, \cdots, N$ and $j = 1, \cdots, J$, with outcomes Y_{ij}. They are different from panel data in that there is **no time ordering** for the second index. Many methods developed for panel data are also applicable here.

6.1.2 Shapes of Data Frames

Panel data can also be loosely classified by the shape:

- **Thin Frames** ($N \gg T$), where the number of cross-section units is large relative to the number of time periods:
  - unit-specific parameters (individual FEs) **can not be estimated consistently** due to the short time series
  - REs might be more suitable since they place a stocahstic structure on the individual components
- **Fat Frames** ($N \ll T$), where the number of cross-section units is large relative to the number of time periods.
- **Square** $N \simeq T$, where the number of units and time periods is comparable.

6.1.3 Assignment Mechanisms

6.1.3.1 The General Case

In the most general case, the treatment may vary both across units and over time, with units **switching** in and out of the treatment group:

$$\mathbf{W}^{\text{general}} = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$$

This is more relevant for the RCED configurations, and for panel data of products and promotions as treatments. The assumption on the absence/presence of **dynamic treatment** effects is very important.

6.1.3.2 Single Treated Period

One special case arises when a substantial number of units is treated, but these units are only treated **in the last period**

$$\mathbf{W}^{\text{last}} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

If T is relatively small, this case is often analyzed as a cross-section problem, the lagged outcomes are used as exogenous covariates or pre-treatment variables to be adjusted. Here, dynamic effects are not testable,

nor do they matter since the shortness of the panel.

$$\mathbf{W}^{\text{last}} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 1 & \cdots & 0 \end{pmatrix}$$

this setting is prominent in the original applications of the synthetic control literature, here T is usually small.

6.1.3.3 Single Treated Unit and Single Treated Period

An extreme case is where only a single unit is treated, and it is only treated in a single period (typically the last).

$$\mathbf{W}^{\text{block}} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Normally, we focus on the effect for the single treated/time-period pair and construct prediction intervals.

6.1.3.4 Block Assignment

The case of block assignment is where a subset of units is treated every period after a common starting date:

$$\mathbf{W}^{\text{block}} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

There is typically a sufficient number of treated unit/time-period pairs to allow for reasonable approximations. The presence of dynamic effects change the interpretation of the average effect of the treated: the average effect for the treated now is an average over short **and** medium term effects during different periods.

Staggered Adoption (a.k.a. absorbing treatment setting)

The staggered adoption is the case where units adopt the treatment at various period, and remain in the treatment group once they adopt the treatment:

$$\mathbf{W}^{\text{block}} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

Here, with some assumptions, we can separate dynamic effects from heterogeneity across calendar time.

6.1.3.5 Event Study Designs

In the event-study design, units are exposed to the treatment in at most one period:

$$\mathbf{W}^{\text{block}} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & \cdots & 0 \end{pmatrix}$$

There are often dynamic effects of the treatment past hte time of initial treatment, however, the effects might be changing over time.

6.1.3.6 Clustered Assignment

In many applications, units are grouped together in clusters. Units within the same clusters are always assigned to the treatment:

$$\mathbf{W}^{\text{cluster}} = \begin{pmatrix} & & & & & & & \text{cluster} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & \cdots & 1 & 1 & C \\ 0 & 0 & 0 & 1 & \cdots & 1 & 1 & C \end{pmatrix}$$

Clustering creates complications for inference.

6.1.4 Outcomes, Assumptions and Estimands

For a treatment assignment matrix $\mathbf{W}$, denote:

- the full $T-$component column vector of treatment assignments as

$$\underline{\mathbf{w}} \equiv (w_1, \cdots, w_T)'$$

- the t-component column vector of treatment assignments **up to** time t as

$$\underline{\mathbf{w}}^t \equiv (w_1, \cdots, w_t)'$$

  hence $\underline{\mathbf{w}}^T = \underline{\mathbf{w}}$
- the row vector of treatment values for unit i as $\underline{\mathbf{W}}_i$

Then in general, we can index the potential outcomes for unit i in period t by the full $T-$component vector of assignments $\underline{\mathbf{w}}$

$$Y_{it}\left(\underline{\mathbf{w}}\right)$$

A key underlying assumption is the **Stable Unit Treatment Value Assumption (SUTVA)**, which requires that there is no interference or spillovers between units[1].

[1] SUTVA can hold on a cluster/group level, where the spillover effects are within clusters/groups.

In this setup, there are 2^T potential outcomes for each unit and each time period, as a function of multi-valued treatment $\underline{\mathbf{w}}$. Then, define for each t-unit treatment effects for each pair of assignment vectors $\underline{\mathbf{w}}$ and $\underline{\mathbf{w}}'$:

$$\tau_{it}^{\underline{\mathbf{w}},\underline{\mathbf{w}}'} \equiv Y_{it}\left(\underline{\mathbf{w}}'\right) - Y_{it}\left(\underline{\mathbf{w}}\right)$$

and the corresponding population average effect

$$\tau_{t}^{\underline{\mathbf{w}},\underline{\mathbf{w}}'} \equiv \mathbb{E}\left[Y_{it}\left(\underline{\mathbf{w}}'\right) - Y_{it}\left(\underline{\mathbf{w}}\right)\right]$$

where the expectation is implicitly assumed to be taken over a **large** population.

Under completely random assignment, all $\tau_{t}^{\underline{\mathbf{w}},\underline{\mathbf{w}}'}$ are identified, and are **just-identified**, given sufficient variation in the treatment paths. Dynamic treatment effects can also be identified[2]. However, we have $2^{T-1} \times \left(2^T - 1\right)$ distinct average effects of the form $\tau_{t}^{\underline{\mathbf{w}},\underline{\mathbf{w}}'}$, in practice, we often need to focus on summary measures of these causal effects, which requires some addition assumptions:

Assumption 6.1.1: No Anticipation

The potential outcomes satisfy

$$Y_{it}\left(\underline{\mathbf{w}}\right) = Y_{it}\left(\underline{\mathbf{w}}'\right)$$

for all i, and for all combinations of t, $\underline{\mathbf{w}}$ and $\underline{\mathbf{w}}'$ such that $\underline{\mathbf{w}}t = \underline{\mathbf{w}}'^t$.

This is a testable assumption with experimental data and sufficient variation in treatment paths, by comparing units that have the same treatment path up to and including t and diverge after t.

- Units **are not** active decision-makers: the assumption can be guaranteed by design (random treatment assignment each period, or staggered adoption with randomly assigned adoption date)
- **Limited** anticipation: assuming the treatment can be anticipated for a **fixed number** of periods, which shifts $\underline{\mathbf{w}}$ by that number of periods.
- Units **are** active decision-makers: potential outcomes are functions of $\underline{\mathbf{w}}$ and the distribution of $\underline{\mathbf{w}}$ (experimental design itself):
  - one can define potential outcomes for a given randomized experimental design: the beliefs about the future treatment paths are incorporated in the definition of the potential outcomes, the actual values are by construction unknown. This does change the interpretation of the casual effects[3].
  - In obserational studies, one cannot directly control the information about the future treatment paths. In this case, different units need to be gauranteed to face the **same information environment** for Assumption 6.1.1 to hold.

Under Assumption 6.1.1, the total number of potential treatment effects is reduced from $2^{T-1} \times \left(2^T - 1\right)$ to $\left(\sum_{t=1}^{T} 2^{t-1}\right)\left(\sum_{t=1}^{T} 2^t - 1\right)$. The unit-period specific treatment effects are now of the type

$$\tau_{it}^{\underline{\mathbf{w}}^t,\underline{\mathbf{w}}^{t'}} \equiv Y_{it}\left(\underline{\mathbf{w}}^{t'}\right) - Y_{it}\left(\underline{\mathbf{w}}^t\right)$$

with the potential outcomes for period t indexed by treatments up to period t only. Here, one can still distinguish

[2]For example, consider that in the 2-period case

$$\tau_{2}^{(1,1),(0,1)}$$

is the average effect in the second period of being exposed to $(1,1)$, *treated in both period*, rather than $(0,1)$, *treated only in the second period*.

[3]Think about the differences between a surprise deviation from a given policy rule versus the effect of a permanent chagne int he policy rule itself.

- static treatment effects: $\tau_{it}^{(\underline{\mathbf{w}}^{t-1},0),(\underline{\mathbf{w}}^{t-1},1)}$, which measures the response of current outcome to the current treatment, holding the past ones fixed.
- dynamic treatment effects: $\tau_{it}^{(\underline{\mathbf{w}}^{t-1},w^t),(\underline{\mathbf{w}}^{t-1},w^t)}$, which does the opposite.

Assumption 6.1.2: No Dynamic/Carry-Over Effects

The potential outcomes satisfy

$$Y_{it}\left(\underline{\mathbf{w}}\right) = Y_{it}\left(\underline{\mathbf{w}}'\right)$$

for all i and for all combinations of t, $\underline{\mathbf{w}}$ and $\underline{\mathbf{w}}'$ such that $w_{it} = w'_{it}$.

This assumption is **not** guaranteed by randomization. It restricts the treatment effects and the potential outcomes for the **post**-treatment periods. It has testable restrictions given the random assignment of the treatment and sufficient variation in the treatment paths. It does **not** restrict the time path of the potential outcomes in the absence of any treatment $Y_{it}(\mathbf{0})$.

This assumption greatly reduce the total number of treatment effects for each unit to T:

$$\tau_{it} \equiv Y_{it}(1) - Y_{it}(0)$$

where τ_{it} has no superscripts because there are only 2 possible arguments of the potential outcomes $w \in \{0, 1\}$.

Assumption 6.1.3: Staggered Adoption

In staggered adoption,

$$W_{it} \leq W_{it-1},\ \forall t = 2, \cdots, T$$

define the adoption date A_i as the date of the first treatment, $A_i \equiv T + 1 - \sum_{t=1}^{T} W_{it}$ for treated units, and $A_i \equiv \infty$ for never-treated units.

Under Assumption 6.1.3, the potential outcomes can be written in terms of the adoption date as $Y_{it}(a)$, for $a = 1, \cdots, T, \infty$, and the realized outcome as $Y_{it} = Y_{it}(A_i)$. There are 2 broad classes of settings that are viewed as staggered adoption designs:

- interventions adopted and remain in place
- one-time interventions with a long-term, or even permanent, impact (where the post-intervention period effects are dynamic effects)

Under Assumption 6.1.3, but **not** Assumption 6.1.1 and 6.1.2, we can write

$$\tau_{it}^{a,a'} \equiv Y_{it}(a') - Y_{it}(a)$$

with the corresponding population average

$$\tau_{t}^{a,a'} \equiv \mathbb{E}\left[\equiv Y_{it}(a') - Y_{it}(a)\right]$$

References

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