# Stacking: How to solve problem with TWFE when having already-treated?

Juhyun Bae (5.27.2022)

Sources: Prof. Scott Cunningham's workshop

# The problem with TWFE w. already treated

## Staggered DID (last time):

- Units, e.g., individuals, firms, or states, adopt the policy or treatment of interest at a particular point in time, and then **remain exposed to this treatment at all times afterwards**. The **adoption date** at which units are first exposed to the policy may **vary by unit** (Athey & Imbens, 2022).
- We concentrate our attention on DiD with staggered adoption, i.e., to DiD setups such that **once units are treated, they remain treated** in the following periods (Callaway & Sant’Anna, 2021).
- staggered DiD estimates are likely **biased: heterogenous treatment effect either (1) <u>over time</u> or (2) <u>across groups</u>** (page 374).

Can we use TWFE, but avoid the use of already-treated as controls?

# Three ways to stick with TWFE

1. Stacking: Cengiz et al (2019) create a series of smaller datasets recentered in relative event time, combine them, estimate TWFE on the combined datasets

2. Interactions: Wooldridge (2021) introduces the Mundlak estimator which is TWFE with interactions which addresses the flaw in the canonical specification

3. Imputation: Borusyak et al (2021) impute missing Y0 for all treated units using estimated fixed effects from an untreated sample, estimate unit level treatment effects than aggregate

# Steps to stacked regression

1. Define an event window

2. Create separate “event by cohort specific” datasets for each policy cohort

3. Append each dataset (“stacking”) to one another

4. Estimate a simple 2*2 model but include “cohort-by-state” FE so as to account for the multiple control groups

# 1. Define an event window

Pre-treatment - post-treatment

Choosing the event window is a research design decision

# 2. Create separate "event by cohort specific" datasets for each policy cohort

- Dataset will consist of the relevant policy cohort **plus** controls
- Data is structured in "event time" and will be balanced such that panel length is *h* (e.g. -3, +4)
- Each dataset will contain individuals untreated over the *h* period defined.

# 3. Append each dataset ("stacking") to one another

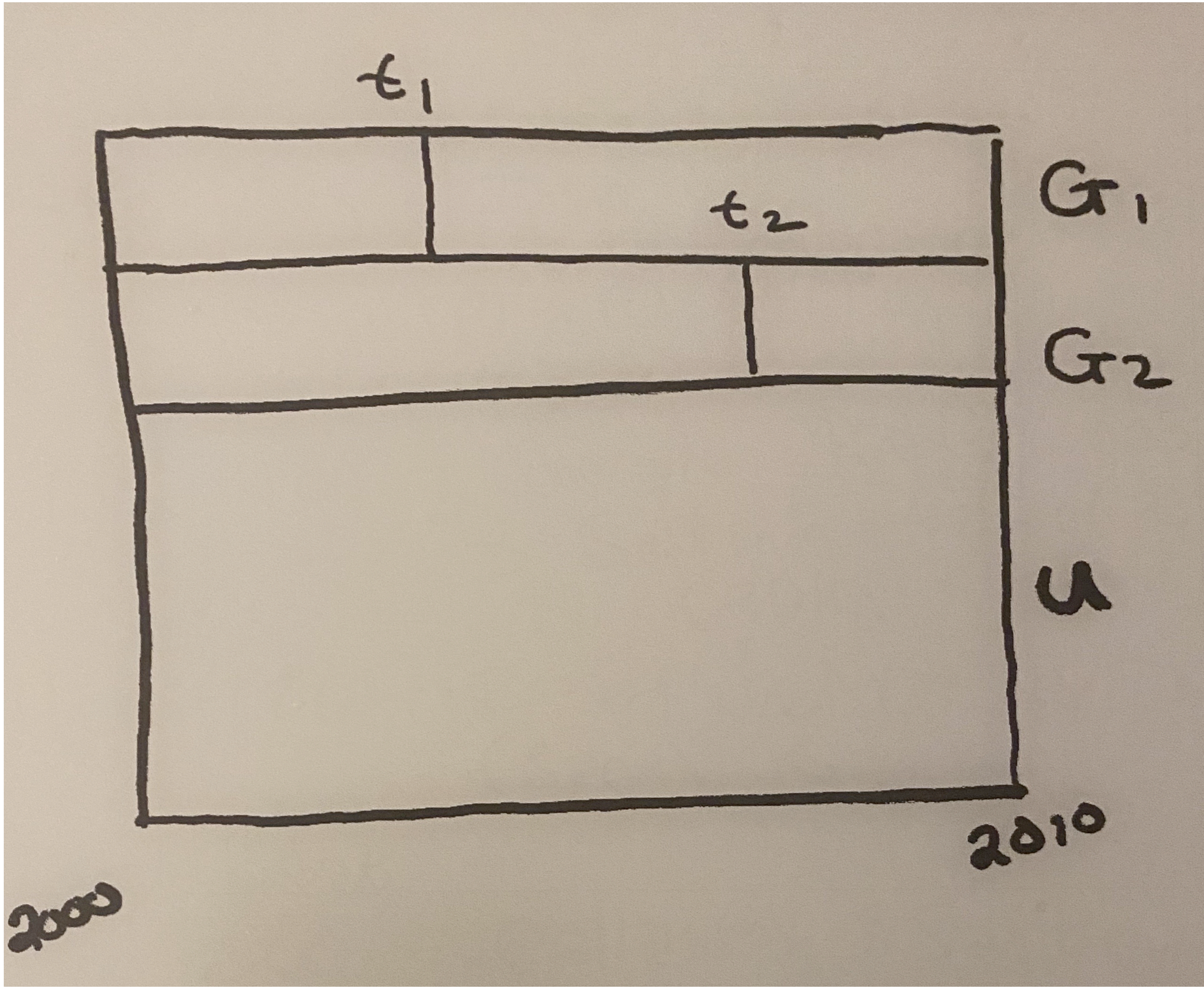
- This necessarily replicates control observations though as they are in each datasets.
- Since the same people are often appearing many times, you will correct for this in the regression model specification.

# 4. Estimate a simple 2*2 model but include “cohort-by-state” FE so as to account for the multiple appearances of observations from NT states

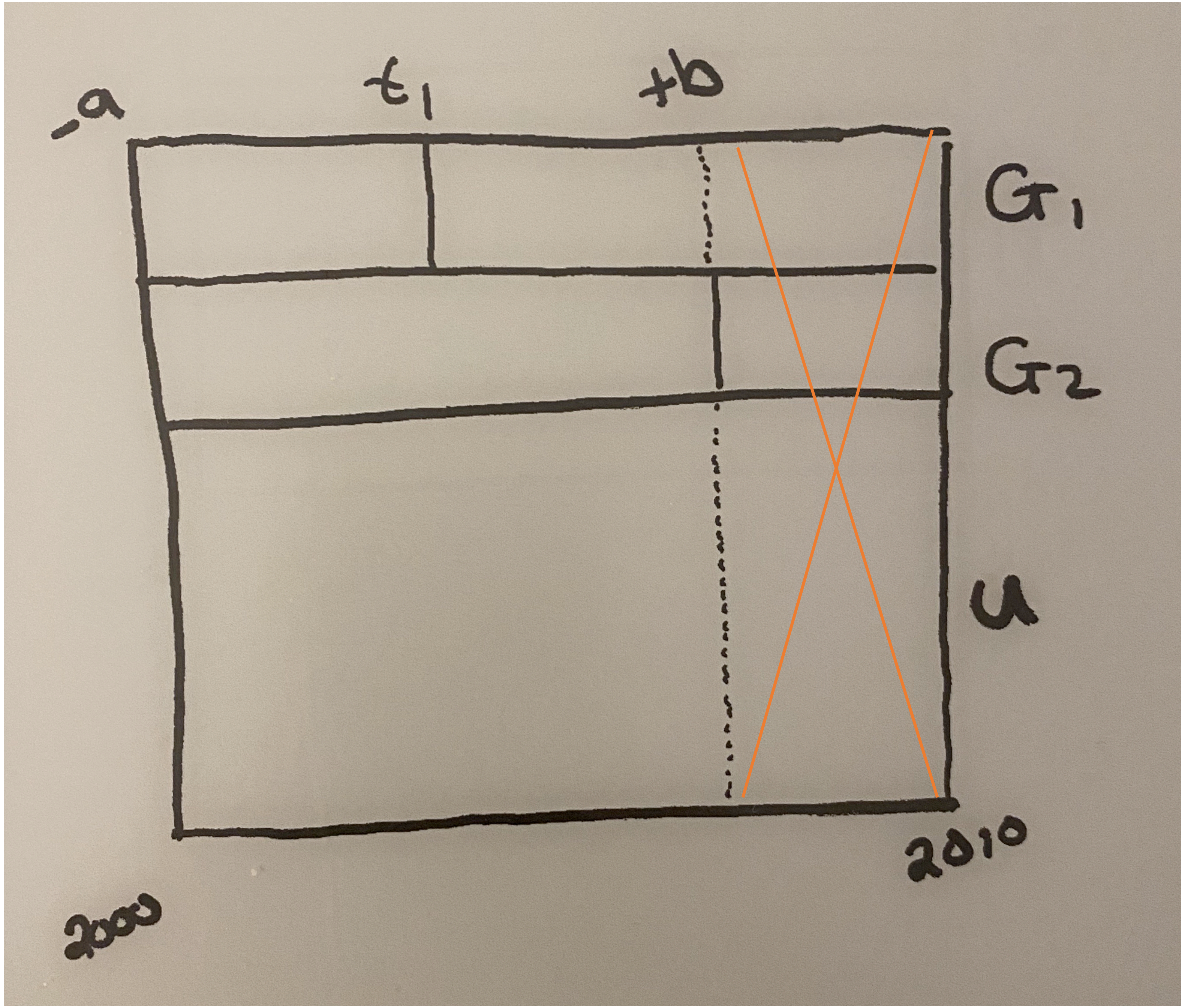
- Stacked DID with clustering at the Unit x Design level over rejects the null hypothesis. This makes sense because it does not account for duplication.
- Stacked DID with clustering at the unit level works pretty well.

- Clustered at the unit level in stacked event studies. (Deshpande and Li, 2019)

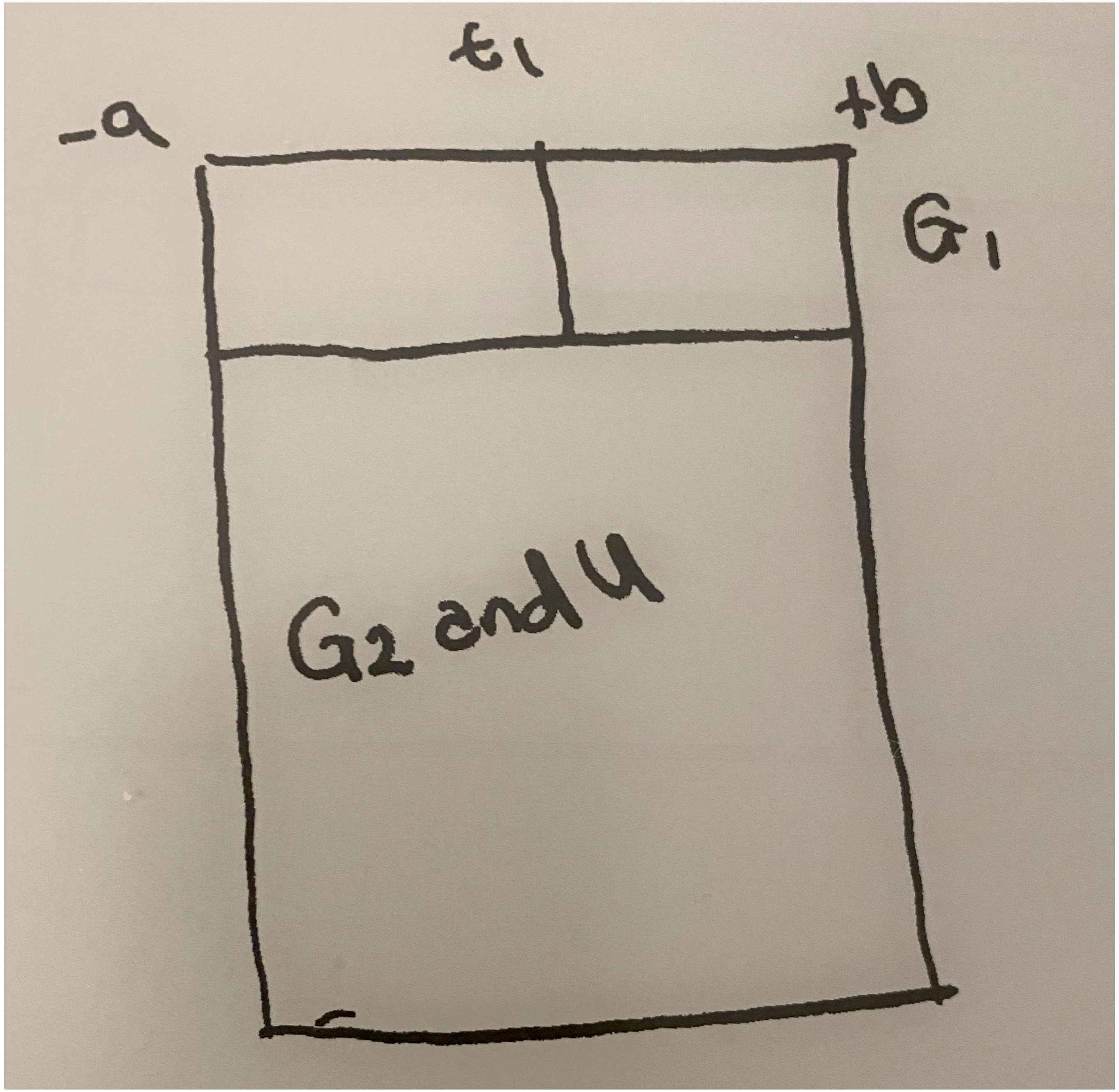
EX)
Imbalanced in relative time w. differential timing



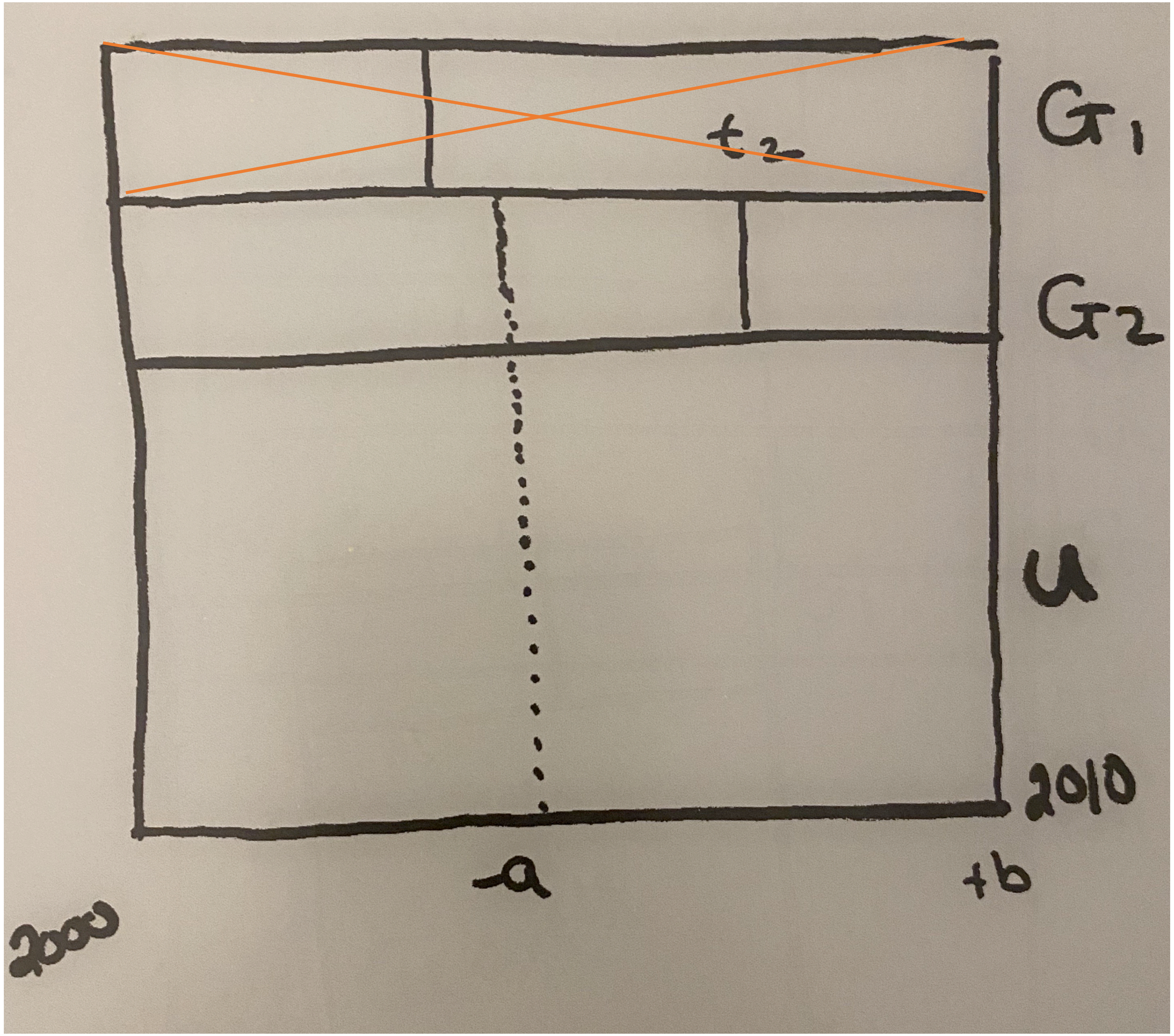
# Creating G1 dataset: choose max pre(-a) and post (+b) periods



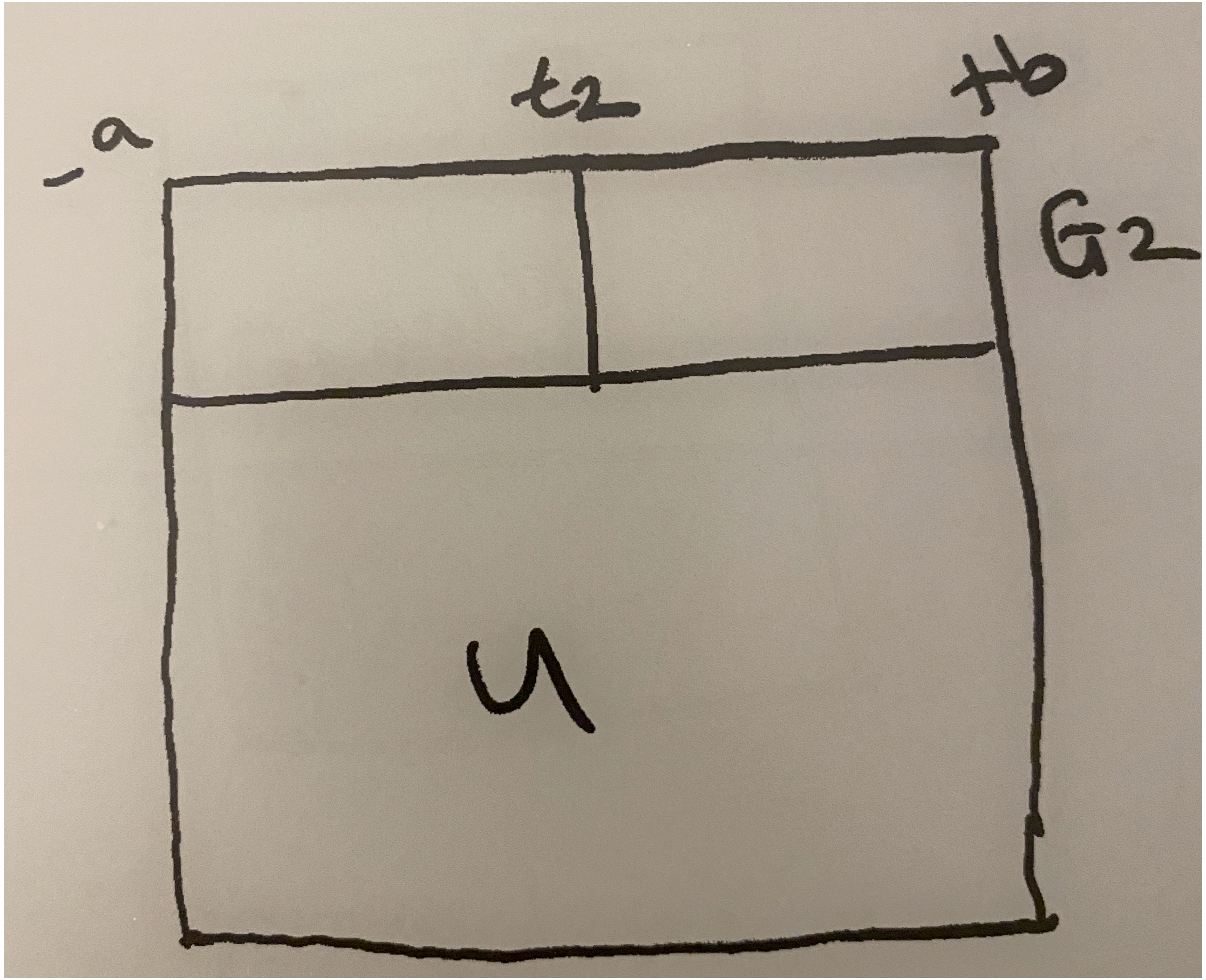
Creating G1 dataset: keep untreated units on [-a, +b]



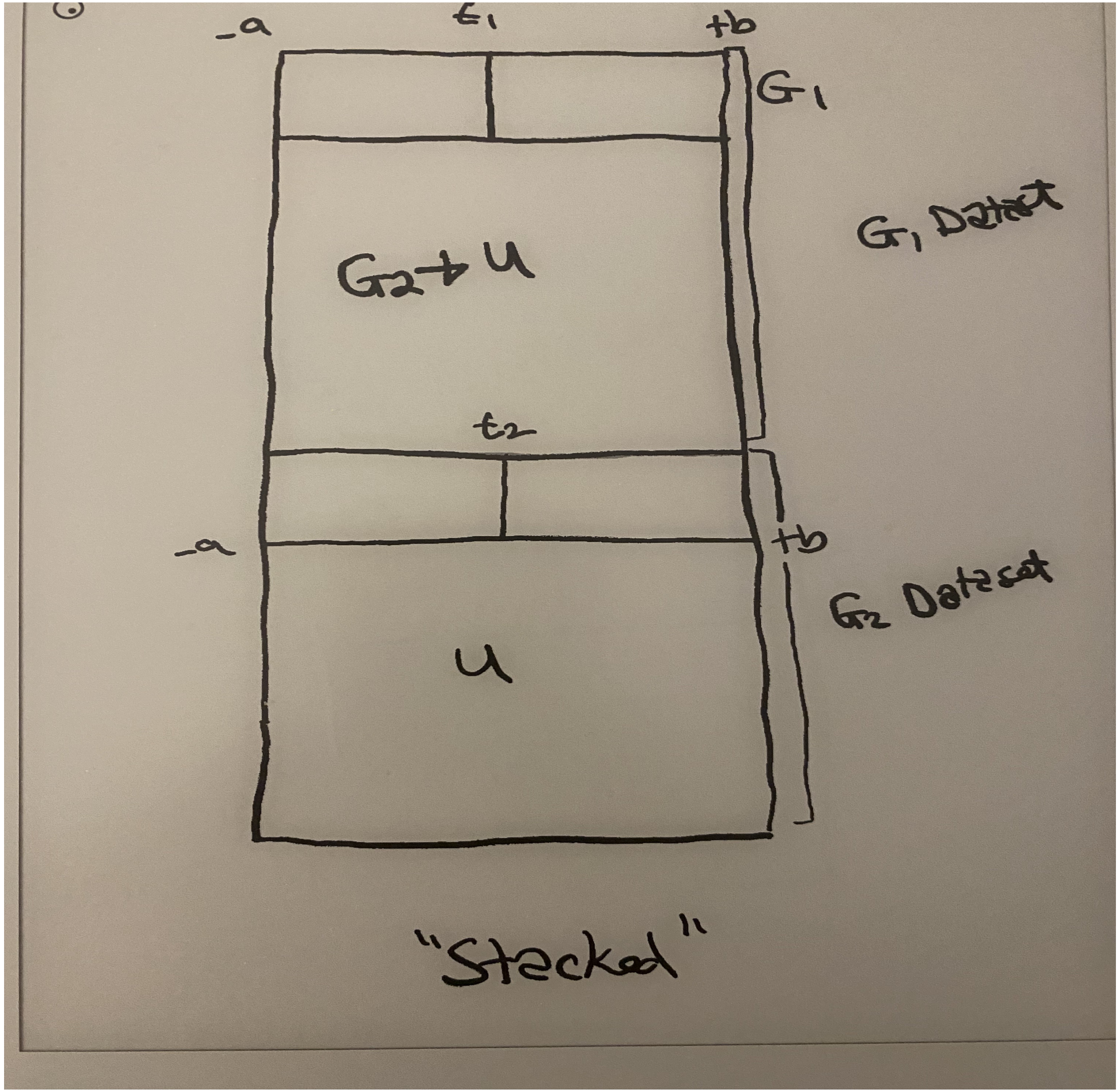
Creating G2 dataset: keep only untreated units on [-a, +b] intervals as controls



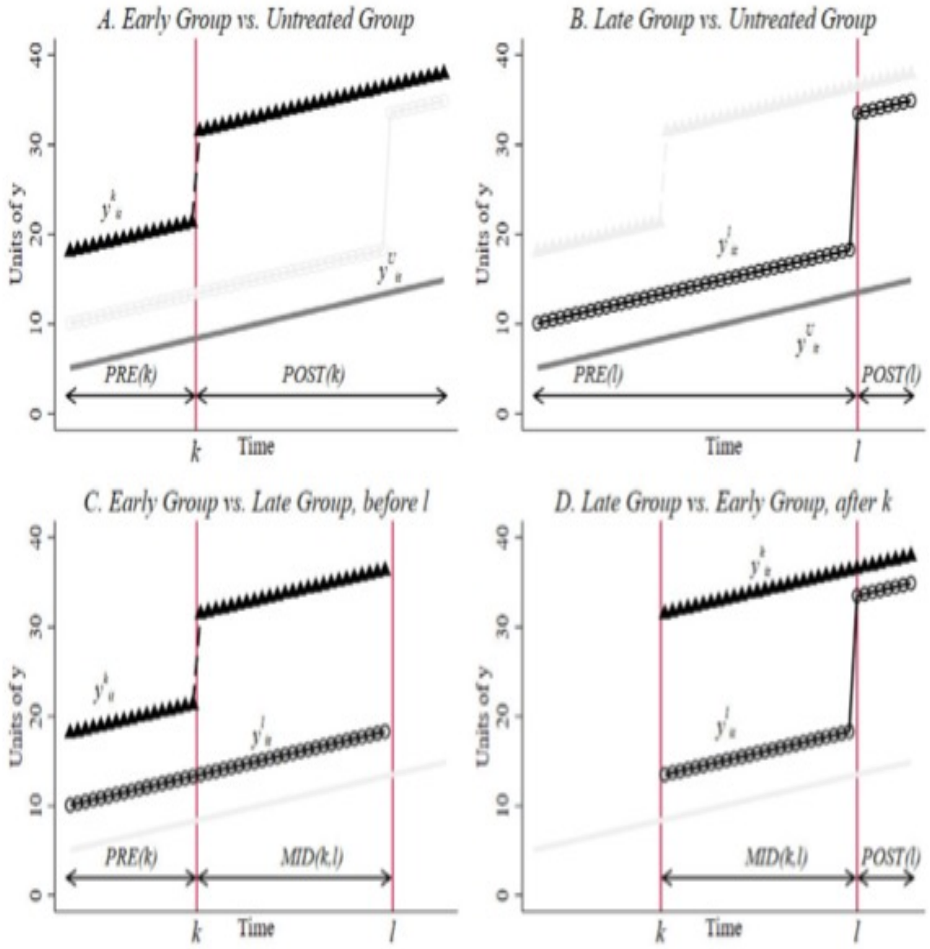
Creating G2 dataset:
save the dataset

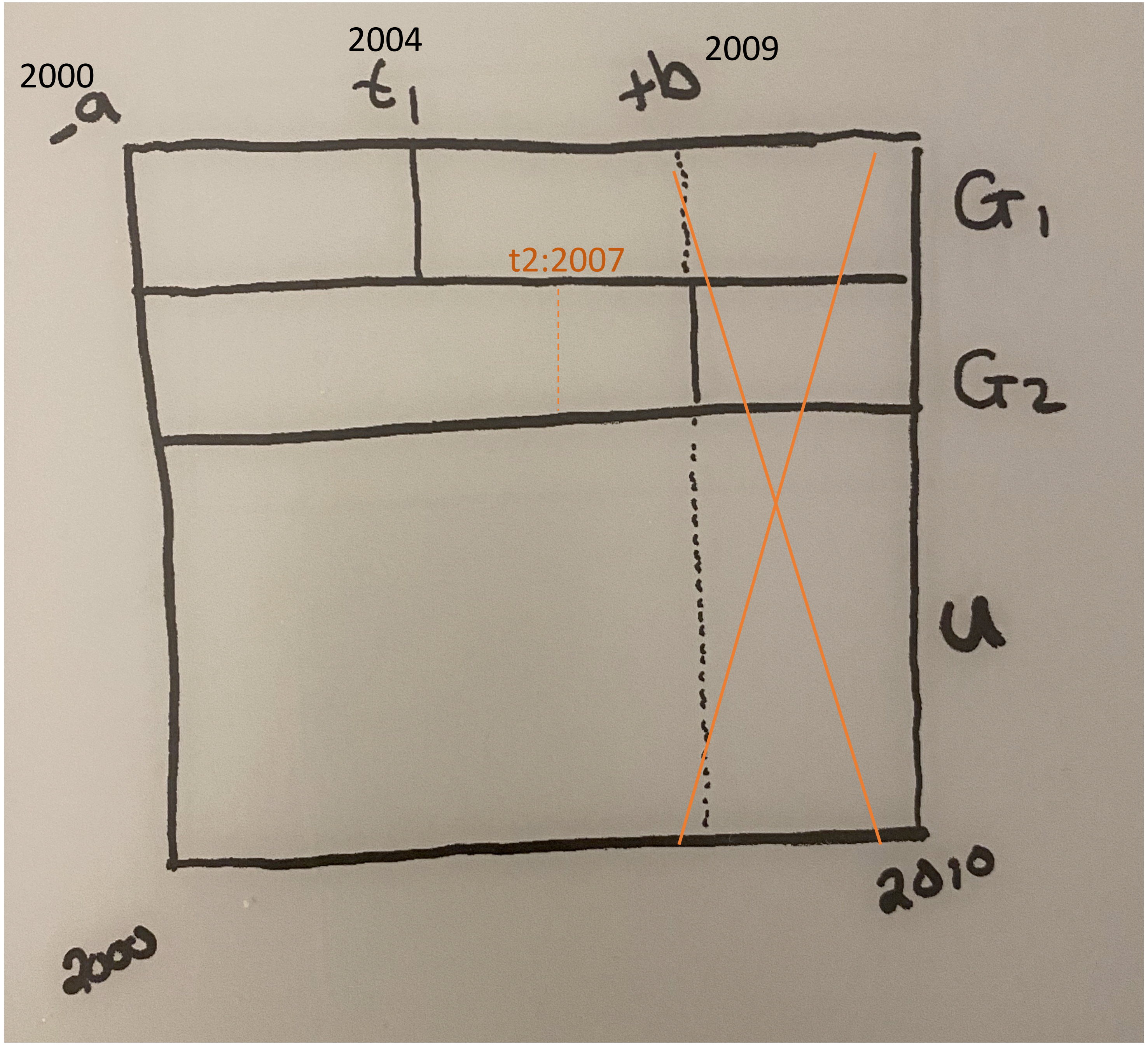


# Creating new dataset: stack the dataset



# **Q**: Creating G1 dataset: what if t2 is earlier than b?



# Discussion of the stacked data

- Why doesn't G1 appear in G2?
- Notice that U appears in both G1 and G2 datasets.
- Unclear to me exactly what Cengiz, et al. (2019) run in their stacked regression, but we probably need to say it now that we want to have a control for "dataset-by-state" fixed effects some units appear more than once (e.g., U).

# Pros and Cons of Stacking

It ensures that you don’t have any problematic comparisons and so it is robust to biases.

BUT!

Stacking is also not perfect! It doesn’t provide the perfect way to “weight and sum” event treatment effects (ref. Gardner 2021)

It still won’t actually “reveal” the time varying effects

(B.c. the stacked DID specification averages all of the time-varying effects into a single averaged effect.)

# Summary

- TWFE is biased w. differential timing and heterogenous treatment effects by cohort.
- Staking is a way of deconstructing and reconstructing a dataset w. differential timing.
- Alternative way to “clean controls”, stacking= re-center each treatment date
- A stacked data is balanced in **relative “event time”** -> no longer be described by differential timing
- W/O differential timing, canonical static specification can be estimated with unbiased TWFE
- Several units will appear more than once.

• Hard part of this is the looping so that each time you create a new dataset, you drop the group that had been treated in the previous one

# Next plan

To discuss the next study group plan:

- Replication project: to choose which paper to replicate for practice