Applied Quantitative Methods for the Social Sciences II
MA in Social Sciences, UC3M, Spring 2026

Assignment 6: Panel Data II

Instructions:

- Deadline: **March 19, before class**
- Submit your work in a separate folder in your GitHub repository
  - You can include only the R file or additional ones (e.g. pdf with results)
- **Always use comments** in your R code – and use them to answer questions
- You are encouraged to work together, but each person must submit their own code
- Plan is to start Part 1 in class and complete Part 2 at home
- I'll upload a solution file to the website after next class

Contents

1 Part 1: In-Class (Card-Krueger Minimum Wage) 2

2 Part 2: Take-Home (Staggered DiD with the `did` Package) 5

3 Data Sources 8

4 Submission 8

1 Part 1: In-Class (Card-Krueger Minimum Wage)

In this lab, we replicate the classic Card and Krueger (1994) study on the employment effects of minimum wage increases. In February 1992, New Jersey announced it would raise its minimum wage from $4.25 to $5.05, effective April 1, 1992. Pennsylvania, a neighboring state, did not change its minimum wage. Card and Krueger surveyed fast food restaurants in both states before (February 1992) and after (November 1992) the policy change. This is a canonical difference-in-differences design: NJ is the treatment group, PA is the control group, and the policy change is the treatment.

Download the data here:

- github.com/franvillamil/AQM2/tree/master/datasets/other

Load it with `df = read.csv("minwage.csv")`. Key variables:

- `chain` — fast food chain (wendys, burgerking, kfc, royalrogers)
- `location` — state/region (PA, centralNJ, northNJ, shoreNJ, southNJ)
- `wageBefore` — starting wage before the minimum wage increase
- `wageAfter` — starting wage after the minimum wage increase
- `fullBefore` — number of full-time employees before
- `fullAfter` — number of full-time employees after
- `partBefore` — number of part-time employees before
- `partAfter` — number of part-time employees after

1.1 Data setup and exploration

a) Create a `NJ` dummy variable that equals 1 if `location` is not `"PA"` and 0 otherwise. Report the number of restaurants in NJ vs. PA using `table()`. Then compute the average `wageBefore` and `wageAfter` separately for NJ and PA restaurants using `group_by()` and `summarise()`. In a comment, note whether wages in NJ increased relative to PA after the policy change.

b) Compute the simple DiD estimate manually using the following steps:

   - For NJ: `mean(fullAfter) - mean(fullBefore)`
   - For PA: `mean(fullAfter) - mean(fullBefore)`
   - DiD = (NJ after − NJ before) − (PA after − PA before)

   In a comment, interpret the result in words. What does this number say about the effect of the minimum wage increase on employment?

c) To run regressions, reshape the data to long format (one row per restaurant-period) using the following code:

```
library(dplyr)
library(tidyr)

df_long = df %>%
  mutate(id = row_number()) %>%
```

```
  pivot_longer(
    cols = c(fullBefore, fullAfter),
    names_to = "period",
    values_to = "full_emp") %>%
  mutate(
    post = ifelse(period == "fullAfter", 1, 0),
    NJ = ifelse(location != "PA", 1, 0))
```

Check that the resulting dataset has the correct structure: `nrow(df_long)` should equal twice `nrow(df)`. In a comment, explain why the long format is needed for the DiD regression.

1.2 DiD regression

a) Estimate the DiD regression using `fixest`:

```
library(fixest)
m_did = feols(full_emp ~ post * NJ, data = df_long, cluster = ~id)
```

Report the results using `modelsummary()`. Identify and interpret the coefficient on the interaction term `post:NJ`: this is the DiD estimator. Compare it to your manual calculation from question 1.1b — they should match.

b) Add chain fixed effects to absorb time-invariant differences across fast food chains:

```
m_did_fe = feols(full_emp ~ post * NJ | chain, data = df_long, cluster = ~id)
```

Compare the two models in a single `modelsummary()` table. Does controlling for chain type change the DiD estimate noticeably? In a comment, explain what the chain fixed effects are absorbing and why controlling for them may or may not matter here.

c) In a comment, state the parallel trends assumption for this specific example. What would we need to observe about NJ and PA employment trends in the pre-period to be confident in the DiD estimate? Give one concrete example of something that could violate this assumption (i.e., something that would affect NJ but not PA employment independently of the minimum wage change).

1.3 Wages as a validation check

a) Repeat the DiD analysis using wages as the outcome instead of employment. Reshape the data for wages and estimate the model:

```
df_long_wage = df %>%
  mutate(id = row_number()) %>%
  pivot_longer(
    cols = c(wageBefore, wageAfter),
    names_to = "period",
```

```
    values_to = "wage") %>%
  mutate(
    post = ifelse(period == "wageAfter", 1, 0),
    NJ = ifelse(location != "PA", 1, 0))

m_wage = feols(wage ~ post * NJ, data = df_long_wage, cluster = ~id)
```

Report the results. Did the minimum wage increase actually raise wages in NJ relative to PA? Is the sign and magnitude of the DiD coefficient what you would expect?

b) In a comment, explain why the wage result is important for interpreting the employment DiD. If wages had *not* risen in NJ after the law change, what would that imply about the employment result? Why is it reassuring (or not surprising) that wages did rise in NJ?

2 Part 2: Take-Home (Staggered DiD with the `did` Package)

In the Card-Krueger design, all treated units (NJ restaurants) received treatment at exactly the same time. In many real-world settings, however, units adopt policies at different times — this is called *staggered treatment adoption*. Recall from lecture that the standard two-way fixed effects (TWFE) estimator can produce misleading results in staggered settings because it uses already-treated units as implicit controls for newly treated units. In this part, we work through these issues using the `mpdta` dataset built into the `did` package.

Load the data with:

```
library(did)
data(mpdta)
```

Key variables:

- `countyreal` — county identifier
- `year` — year (2003–2007)
- `lemp` — log teen employment (outcome)
- `lpop` — log population (control variable)
- `treat` — whether the county was ever treated (1 = yes, 0 = no)
- `first.treat` — year when the county first received treatment (0 if never treated); counties adopt in 2004, 2006, or 2007

2.1 Data structure and visualization

a) How many counties are in the data? How many unique treatment cohorts (distinct values of `first.treat`) are there? Use `table(mpdta$first.treat)` to see how many counties adopted treatment in each year. In a comment, explain what "staggered treatment adoption" means in this context: why is it a problem to simply compare treated vs. untreated counties?

b) Plot average log teen employment (`lemp`) over years, separately for each treatment cohort. Use the following code:

```
library(dplyr)
library(ggplot2)

mpdta_avg = mpdta %>%
  mutate(cohort = factor(first.treat,
    levels = c(0, 2004, 2006, 2007),
    labels = c("Never treated", "Adopted 2004",
               "Adopted 2006", "Adopted 2007"))) %>%
  group_by(year, cohort) %>%
  summarise(mean_lemp = mean(lemp, na.rm = TRUE))

ggplot(mpdta_avg, aes(x = year, y = mean_lemp, color = cohort)) +
  geom_line() +
```

```
geom_point() +
theme_minimal() +
labs(x = "Year", y = "Log teen employment", color = "Treatment cohort")
```

Save the plot with `ggsave()`. In a comment, describe the patterns: do the cohorts appear to follow similar trends before their respective treatment years? What happens after treatment? Are there any cohorts whose pre-trends look problematic?

2.2 Naive TWFE vs. Callaway-Santánna estimator

a) Estimate a naive TWFE model treating all treated counties as a single group.
(Check the data, maybe you need to create a time-varying treatment indicator first, call it `treated_post`. It should indicate treatment units after treatment comes into effect.)
Hint: `(first.treat > 0 & year >= first.treat)`
Report and interpret the coefficient on `treated_post`. In a comment, note that this model pools all treatment cohorts together — what implicit assumption is it making about the treatment effect across cohorts and over time?

b) Now use the Callaway-Santánna (2021) estimator, which estimates group-time average treatment effects separately for each cohort and time period, using never-treated counties as the control group. Check: https://bcallaway11.github.io/did/
Report the overall ATT estimate. Is it similar to or different from the naive TWFE estimate?

c) Examine the event-study version of the Callaway-Santánna results (check the documentation for the 'aggte' function). Save the event-study plot with `ggsave()`. In a comment, describe what the plot shows. Are the pre-treatment estimates (leads, i.e. periods before treatment) statistically distinguishable from zero? What does this tell us about the parallel trends assumption? What do the post-treatment estimates (lags) show about the dynamic effects of treatment?

2.3 Pre-testing the parallel trends assumption

The group-time ATT estimates from `att_gt()` include pre-treatment periods — specifically, ATT(g, t) for $t < g$ — which can be used to construct a formal joint test of the parallel trends assumption.

a) Re-run the CS estimator with bootstrapped standard errors to obtain valid uniform confidence bands and a joint pre-test. (See options 'bstrap' and 'cband' to 'TRUE'). The `summary()` output includes a p-value for the pre-test of the parallel trends assumption. Report this p-value. In a comment, explain what the test is doing: what is the null hypothesis, and what does a large p-value tell us?

b) Visualize all group-time ATT estimates — both pre- and post-treatment — with "ggdid". Save the plot. Each panel corresponds to a treatment cohort; negative event-time

values are pre-treatment periods. In a comment, describe what you see: are the pre-treatment ATT(g, t) estimates close to zero and statistically indistinguishable from zero across all cohorts?

c) In a comment (2–3 sentences), reflect on the limitations of pre-testing. Even if we cannot reject parallel trends in the pre-period, can we be certain the assumption holds during the post-treatment period? What is the pre-test actually telling us, and what is it *not* telling us?

2.4 Comparing control group specifications

By default, the CS estimator uses *never-treated* units as the control group. An alternative is *not-yet-treated* units — counties that will eventually receive treatment but have not yet been treated at time t. This expands the control group (more observations, potentially more precision) but introduces a different assumption: that outcomes for not-yet-treated units are unaffected by anticipation of their own future treatment.

a) Re-estimate the CS model using not-yet-treated counties as the control group. Report the overall ATT. Compare it to the never-treated estimate from Section 2.2b. Are they similar or different in sign and magnitude?

b) Produce and save an event-study plot for this specification. In a comment, compare the pre-trends and post-treatment patterns to the never-treated event study from Section 2.2c. Does using the broader control group change the conclusions?

c) In a comment (2–3 sentences), discuss the trade-off between the two control group choices. Under what conditions would you prefer never-treated as the control group? When might not-yet-treated be preferable despite the additional assumption it requires?

2.5 Discussion: why does TWFE fail in staggered settings?

a) In a comment (3–5 sentences), explain intuitively why the naive TWFE estimator can produce misleading results in staggered DiD settings. What is the "forbidden comparison" problem? Which units get used as the control group in a way that is problematic, and why is that a problem if treatment effects are heterogeneous across cohorts or over time?

b) Compare the TWFE estimate from question 2.2a to the Callaway-Santánna estimate from question 2.2b. Are they similar or different? In a comment, based on the event-study pre-trends from question 2.2c, which estimate do you find more credible and why?

3 Data Sources

- Minimum wage (Card and Krueger): github.com/franvillamil/AQM2/tree/master/datasets/other
- Teen employment (Callaway-Santánna): `mpdta` dataset in the `did` R package (`library(did); data(mpdta)`)

4 Submission

Commit your file to your GitHub repository before the deadline. Put it in a separate folder, e.g. `assignment6`. Make sure your repository is public so I can access it.

Your R script should:

- Be well-organized with clear section headers (using comments)
- Include all code needed to reproduce your analysis
- Include your answers and interpretations as comments
- Save any plots to files (e.g., using `ggsave()`)
- Run without errors from top to bottom