

# Assignment 6: Panel Data II – Part 1 (Card-Krueger Minimum Wage)

Applied Quantitative Methods for the Social Sciences II

Spring 2026

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(fixest)
library(modelsummary)

df = read.csv("https://raw.githubusercontent.com/franvillamil/AQM2/refs/heads/master/datasets/other/minwage")
```

## 1. Data setup and exploration

a) Create an NJ dummy and summarise wages by state:

```
df = df %>% mutate(NJ = ifelse(location != "PA", 1, 0))
table(df$NJ)
```

```
##
##    0    1
## 67 291
```

NJ restaurants (coded 1) outnumber PA restaurants (coded 0) because the NJ sample spans four sub-regions (centralNJ, northNJ, shoreNJ, southNJ) while PA is treated as a single region.

```
df %>%
  group_by(NJ) %>%
  summarise(
    mean_wage_before = mean(wageBefore, na.rm = TRUE),
    mean_wage_after  = mean(wageAfter,  na.rm = TRUE))
```

```
## # A tibble: 2 x 3
##       NJ mean_wage_before mean_wage_after
##   <dbl>          <dbl>          <dbl>
## 1     0             4.65             4.61
## 2     1             4.61             5.08
```

Before the policy change, average starting wages in NJ and PA were nearly identical (both close to the federal minimum of \$4.25). After the NJ minimum wage rose to \$5.05, NJ wages increased noticeably while PA wages remained flat, confirming the policy raised wages in the treated state.

b) Manual DiD estimate for full-time employment:

```
means = df %>%
  group_by(NJ) %>%
  summarise(
    before = mean(fullBefore, na.rm = TRUE),
    after = mean(fullAfter, na.rm = TRUE),
    change = after - before)
means
```

```
## # A tibble: 2 x 4
##   NJ before after change
##   <dbl> <dbl> <dbl> <dbl>
## 1     0  10.7   8.17 -2.49
## 2     1   7.97   8.41  0.435
```

```
nj_change = means$change[means$NJ == 1]
pa_change = means$change[means$NJ == 0]
did_est = nj_change - pa_change
cat("DiD estimate:", round(did_est, 3), "\n")
```

```
## DiD estimate: 2.927
```

The DiD estimate is the difference in within-group changes. A positive value means full-time employment grew more (or fell less) in NJ than in PA after the minimum wage increase, which contradicts the standard prediction that higher minimum wages reduce employment.

c) Reshape to long format for regression:

```
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))
```

```
nrow(df_long)
```

```
## [1] 716
```

```
nrow(df)
```

```
## [1] 358
```

The long-format dataset has exactly twice as many rows as the original. The DiD regression requires long format because the interaction `post × NJ` is the DiD estimator: it captures how the within-NJ change in employment (`post - pre`) differs from the corresponding within-PA change.

## 2. DiD regression

a) Basic DiD regression:

```
m_did = feols(full_emp ~ post * NJ, data = df_long, cluster = ~id)
modelsummary(m_did, stars = TRUE, gof_map = c("nobs", "r.squared"),
             output = "markdown")
```

	Model 1
(Intercept)	10.664*** (1.350)
post	-2.493 (1.639)
NJ	-2.693+ (1.429)
post × NJ	2.927+ (1.724)
Num.Obs.	716
R2	0.008

**Note:** + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

The coefficient on post:NJ is the DiD estimator and should match the manual calculation from 1.1b. The post coefficient captures the pre–post change in PA (the counterfactual trend), the NJ coefficient captures the baseline NJ–PA gap, and the interaction captures the additional change in NJ relative to that trend.

**b) Adding chain fixed effects:**

```
m_did_fe = feols(full_emp ~ post * NJ | chain, data = df_long, cluster = ~id)
modelsummary(
  list("DiD" = m_did, "DiD + Chain FE" = m_did_fe),
  stars = TRUE, gof_map = c("nobs", "r.squared"),
  output = "markdown")
```

	DiD	DiD + Chain FE
(Intercept)	10.664*** (1.350)	
post	-2.493 (1.639)	-2.493 (1.642)
NJ	-2.693+ (1.429)	-2.321+ (1.362)
post × NJ	2.927+ (1.724)	2.927+ (1.728)
Num.Obs.	716	716
R2	0.008	0.066

**Note:** + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

The DiD estimate does not change when chain fixed effects are added. Chain FEs absorb baseline differences in staffing levels across fast-food chains (e.g., Wendy’s may have structurally different employment levels than KFC), but since chain type is roughly balanced across states, controlling for it has little impact on the DiD coefficient.

c) The parallel trends assumption here requires that, absent the NJ minimum wage increase, employment trends in NJ and PA fast-food restaurants would have been the same from February to November 1992. This is plausible because both states share a similar economic environment and the two surveys were close together in time, limiting opportunities for diverging trends. A concrete violation would occur if NJ experienced an independent economic shock during this period — for instance, if a major employer opened or closed plants in NJ between the two survey waves, this would change NJ employment for reasons unrelated to the minimum wage, biasing the DiD estimate.

### 3. Wages as a validation check

a) DiD for wages (first-stage check):

```
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)
modelsummary(m_wage, stars = TRUE, gof_map = c("nobs", "r.squared"),
  output = "markdown")
```

	Model 1
(Intercept)	4.651*** (0.044)
post	-0.038 (0.047)
NJ	-0.041 (0.048)
post × NJ	0.510*** (0.051)
Num.Obs.	716
R2	0.407

**Note:** +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

The interaction coefficient `post:NJ` is positive and statistically significant: wages rose substantially in NJ relative to PA after the policy change, and the magnitude is consistent with the \$0.80 minimum wage increase (\$5.05 - \$4.25). This is precisely the sign and magnitude one would expect if the law was actually binding.

b) The wage DiD serves as a “first stage” or manipulation check. If wages had not risen in NJ after the minimum wage increase, it would be unclear whether the study is truly estimating the effect of a minimum wage change at all — the law might not have been binding, or firms might have already been paying above the new minimum. The fact that wages did rise in NJ gives us confidence that the treatment actually occurred as intended, so the employment

DiD can be credibly interpreted as a causal response to the minimum wage increase rather than a spurious or null comparison.