_15

Empirical Analysis III 2019

Problem Set 5

Aditya Chaudhry

This exercise uses data from Feb-Mar and Nov-Dec 1992 on employment at fast food restaurants in the US States of New Jersey and Pennsylvania taken from Card and Krueger in The American Economic Review, Vol. 84(4). The data are described in the output below. We are interested in how the minimum wage affects employment decisions in these restaurants. In April 1992, New Jersey increased the minimum wage from $4.25 to $5.05. In Pennsylvania, the minimum wage was unchanged, and we assume it to be $3.80 for this exercise.

Question 1

Describe the data (-codebook- if you use Stata).

Table 1 displays the output of codebook, which summarizes the data. This data provides information on 408 fast food restaurants in New Jersey and Pennsylvania. We also have data on variables that might also affect employment and that we might want to condition on

Table 1: Summary of Data table

	(1) full-time employees	(2) full-time employees	(3) full-time employees	(4) full-time employees	(5) full-time employees	(6) full-time employees	(7) full-time employees	(8) full-time employees
1 if New Jersey; 0 if Pennsylvania	-2.341	-2.341	-2.341	-2.341	-2.341	-2.341	-2.341	-2.341
	(-2.37)	(-2.37)	(-2.37)	(-2.37)	(-2.37)	(-2.37)	(-2.37)	(-2.37)
1 if after the law; 0 if before the law	-2.889	-2.889	-2.889	-2.889	-2.889	-2.889	-2.889	-2.889
	(-2.28)	(-2.28)	(-2.28)	(-2.28)	(-2.28)	(-2.28)	(-2.28)	(-2.28)
state_post	3.666	3.666	3.666	3.666	3.666	3.666	3.666	3.666
	(2.60)	(2.60)	(2.60)	(2.60)	(2.60)	(2.60)	(2.60)	(2.60)
number of cash registers in store	0.447	0.447	0.447	0.447	0.447	0.447	0.447	0.447
	(1.44)	(1.44)	(1.44)	(1.44)	(1.44)	(1.44)	(1.44)	(1.44)
number hrs open per day	1.251	1.251	1.251	1.251	1.251	1.251	1.251	1.251
	(7.85)	(7.85)	(7.85)	(7.85)	(7.85)	(7.85)	(7.85)	(7.85)
1 if KFC; 0 otherwise	1.199	1.199	1.199	1.199	1.199	1.199	1.199	1.199
	(1.06)	(1.06)	(1.06)	(1.06)	(1.06)	(1.06)	(1.06)	(1.06)
1 if Roy Rogers; 0 otherwise	-1.878	-1.878	-1.878	-1.878	-1.878	-1.878	-1.878	-1.878
	(-2.22)	(-2.22)	(-2.22)	(-2.22)	(-2.22)	(-2.22)	(-2.22)	(-2.22)
1 if Wendy's; 0 otherwise	4.168	4.168	4.168	4.168	4.168	4.168	4.168	4.168
	(3.79)	(3.79)	(3.79)	(3.79)	(3.79)	(3.79)	(3.79)	(3.79)
Constant	-9.938	-9.938	-9.938	-9.938	-9.938	-9.938	-9.938	-9.938
	(-3.58)	(-3.58)	(-3.58)	(-3.58)	(-3.58)	(-3.58)	(-3.58)	(-3.58)
Observations	775	775	775	775	775	775	775	775

t statistics in parentheses

Question 2

Estimate the following regression on the sample of fast food restaurants in Feb-Mar 1992:

$$empft_{ikt} = \alpha + \gamma\,\text{minwage}_{kt} + \beta_1 nregs_{ikt} + \beta_2 hrsopen_{ikt} + \sum_{j=2}^{4} \eta_j d_j + \epsilon_{ikt}$$

where i denotes restaurant, k denotes state, and t = 0 if the observation is from Feb-Mar and t = 1 if the observation is from Nov-Dec.

Table 2 displays the output of this regression. We see that the coefficient on the minimum wage term is $\gamma = -5.175$.

Table 2: OLS Regression: full time employment on minimum wage, with controls table

	(1)
	full-time employees
minimum wage rate (USD)	-5.175*
	(-2.26)
number of cash registers in store	0.407
	(0.92)
number hrs open per day	1.247***
	(5.31)
1 if KFC; 0 otherwise	1.125
	(0.67)
1 if Roy Rogers; 0 otherwise	-1.178
	(-0.99)
1 if Wendy's; 0 otherwise	4.410**
	(2.79)
Constant	9.740
	(0.95)
Observations	397
R^2	0.136
df_m	6
df_r	390
rmse	8.041
mss	3963.8
rss	25218.5

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 3

Interpret the coefficient γ , and calculate a 90% confidence interval.

The estimate $\gamma = -5.1753$ (t-stat $= -2.26$) implies that a $1 increase in the minimum wage is associated with an 5.1753 fewer full-time employees per establishment, after controlling for restaurant size (nregisters), hours of operation (hrsopen), and restaurant type ($d_2 - d_4$). Of course, it's certainly the case that treatment (the minimum wage) has not been randomly assigned to restaurants, and the OLS regression here provides no identification. Thus, we cannot interpret this estimate as a causal effect.

The 90% confidence interval is $\gamma \in [-8.9473, -1.4034]$. This range, thought strictly less than zero, is very wide.

Question 4

Use the Sum of squares table from the regression output to calculate the R2 and the standard error of the regression (Root MSE).

Using the numbers from Table 2tab:part_2 the R^2 is given by

$$R^2 = \frac{MSS}{TSS} = \frac{3963.8072}{29182.3111} = .1358,$$

and the standard error is given by

$$SE = \sqrt{\frac{RSS}{Degrees\ of\ freedom}} = \sqrt{\frac{25218.5039}{390}} = 8.0413.$$

Question 5

Give an economic interpretation of the coefficients $\eta_2 - \eta_4$ What might explain the relatively large coefficient on -d_4-?

The interpretations of these coefficients (from Table 2) is that:

- The average KFC restaurant has $\eta_2 = 1.1254$ more employees than the average Burger King.

- The average Roy Rogers restaurant has $\eta_3 = 1.1777$ fewer employees than the average Burger King.

- The average Wendy's restaurant has $\eta_4 = 4.4104$ more employees than the average Burger King.

The large η_4 coefficient might represent that the Wendy's food preparation, serving, etc. processe is less automated than Burger King's.

Question 6

Test $H_0 : \eta_2 = \eta_3 = 0$.

We see in Figure 1 that, since the p-value $= .0180 < \alpha = .05$, we reject $H_0 : \eta_2 = \eta_3 = 0$ at the $\alpha = .05$ level.

```
. eststo:   test d2 = d3 = 0

 ( 1)  d2 - d3 = 0
 ( 2)  d2 = 0

       F(  2,   390) =    1.00
            Prob > F =    0.3673
```

Figure 1: Hypothesis test of $H_0 : \eta_2 = \eta_3 = 0$.

Question 7

Test the hypothesis $H_0 : \eta_2 = \eta_3$ using the estimated covariance matrix of the coefficients (-mat list e(V)- if you use Stata). Verify your answer by running the test in Stata using -test-, and/or by performing an F-test.

Let

$$\hat{\beta} = [\hat{\gamma}, \hat{\beta}_1, \hat{\beta}_2, \hat{\eta}_2, \hat{\eta}_3, \hat{\eta}_4, \hat{\alpha}]',$$

and

$$r = [0, 0, 0, 1, -1, 0, 0]'.$$

Then under $H_0 : \eta_2 = \eta_3$, we have that

$$r'\hat{\beta} = \hat{\eta_2} - \hat{\eta_3} \sim N(0, \sigma^2),$$

where

$$\sigma^2 = r'\Sigma r,$$

and Σ is the variance-covariance matrix of $\hat{\beta}$ (from e(V) in Stata). Thus, under H_0

$$\frac{\hat{\eta_2} - \hat{\eta_3}}{\sigma} \sim N(0, 1)$$

asymptotically. Thus,

$$\left(\frac{\hat{\eta_2} - \hat{\eta_3}}{\sigma}\right)^2 \sim F_{1, df_{resid}}$$

where $df_{resid} = 390$ from Table 2. Thus, the F-stat is v

$$\left(\frac{\hat{\eta_2} - \hat{\eta_3}}{\sigma}\right)^2 = \left(\frac{1.125 - (-1.178)}{1.7395}\right)^2 = \left(\frac{2.3030}{1.7395}\right)^2 = 2.3030,$$

which has p-value = $.1863 > .05$. Note that this F-stat and p-value match those of the test $d2 = d3$ output in Figure 2. Hence, we fail to reject H_0 at the .05 level.

```
. eststo:  test d2 = d3

 ( 1)  d2 - d3 = 0

       F(  1,   390) =    1.75
            Prob > F =    0.1863
```

Figure 2: Hypothesis test of $H_0 : \eta_2 = \eta_3$.

Question 8

Explain why the previous estimate of γ is likely to suffer from omitted variable bias.

The estimate of γ likely suffers from omitted variable bias because other variables in addition to those we have controlled for impact employment, for example non-wage state-level regulations and business-cycle factors in the time series. Specifically, it's possible that New Jersey was experiencing an economic downturn and in response to falling wages decided to raise its minimum wage, which would induces correlation between the omitted business cycle variable and the minimum wage regressor.

Question 9

Assume that $\epsilon_{ikt} = \mu_k + \zeta_t + u_{ikt}$ and that $\mathbb{E}[u_{ikt}|X_{ikt}] = 0$ (where X_{ikt} is the vector of RHS- variables in (2) except -minwage-). Explain how you can then use the increase in the minimum wage in New Jersey and a difference-in-differences (DD) model to identify the effect of the minimum wage on employment. Give an example where the necessary assumption(s) are violated.

Plugging this parametric specification of the error term into the original regression specification, we get

$$empft_{ikt} = \alpha + \gamma \,\text{minwage}_{kt} + \beta_1 nregs_{ikt} + \beta_2 hrsopen_{ikt} + \sum_{j=2}^{4} \eta_j d_j + \mu_k + \zeta_t + u_{ikt}.$$

We want to identify γ, but μ_k, ζ_t are unobserved. So we can do a DD. In the first stage, we hold state k fixed and difference across the two time periods to remove mu_k:

$$\begin{aligned} empft_{ik2} - empft_{ik1} &= \gamma(\text{minwage}_{k2} - \text{minwage}_{k1}) + \beta_1(nregs_{ik2} - nregs_{ik1}) + \beta_2(hrsopen_{ik2} - hrsopen_{ik1}) \\ &\quad + \mu_k - \mu_k + \zeta_2 - \zeta_1 + u_{ik2} - u_{ik1} \\ \leftrightarrow \Delta empft_{ik} &= \gamma\Delta\,\text{minwage}_k + \beta_1 \Delta nregs_{ik} + \beta_2 \Delta hrsopen_{ik} + \Delta\zeta + \Delta u_{ik} \\ \leftrightarrow \mathbb{E}[\Delta empft_{ik}] &= \mathbb{E}[\text{Causal effect of minimum wage increase}|\text{Observables}] + \Delta\zeta \end{aligned}$$

In the second difference we difference $\Delta empft_{ik}$ across k to remove Δ:

$$\begin{aligned} \Delta empft_{i1} - \Delta empft_{i0} &= \gamma(\Delta\,\text{minwage}_1 - \Delta\,\text{minwage}_0) + \beta_1(\Delta nregs_{i1} - \Delta nregs_{i0}) \\ &\quad + \beta_2(\Delta hrsopen_{i1} - \Delta hrsopen_{i0}) + \Delta u_{i1} - \Delta u_{i0} \\ \mathbb{E}[\Delta empft_{i1} - \Delta empft_{i0}] &= \mathbb{E}[\text{Causal effect of minimum wage increase}|\text{Observables}]. \end{aligned}$$

Note that $\Delta u_{i1} - \Delta u_{i0} = (u_{i12} - u_{i11}) - (u_{i02} - u_{i01})$ has conditional mean of zero given the covariates. So the DD has eliminated both the time and state fixed effects in the original error term. Hence, we can now use OLS to identify the causal effect (ATT) of the minimum wage increase on employment.

The key assumption here is the common trend assumption: that both states k have the same ${}_{kt} ={}_{t}$ term for each t. If instead we have ${}_{1t} \neq{}_{2t}$, then the $\Delta Zeta$ term will not cancel in the second difference and we will be unable to identify the causal effect. The common trend assumption would be violated if the time trends in restaurant employment are different between New Jersey and Pennsylvania, for example due a non-wage policy change (e.g. one state implements healthcare reform that raises costs to employers while the other state does not).

Question 10

Generate a table of means, a table of standard errors and a table of frequencies for -empft- in each state and each time period (post = 1 and post = 0).

Figure 3 displays the aggregated table of full time employment mean, standard deviation, and restaurant frequency, all broken down by time period and state.

```
. eststo:  tabulate state post, summarize( empft)

    Means, Standard Deviations and Frequencies of # full-time employees

  1 if New  |
 Jersey; 0  |
        if  |  1 if after the law;
Pennsylvan  |  0 if before the law
        ia  |          0          1 |     Total
------------+-----------------------+----------
         0  |  10.311688  7.6513158 | 8.9901961
            |  10.805103  8.5143094 | 9.7941763
            |         77         76 |       153
------------+-----------------------+----------
         1  |  7.7323077   8.446875 | 8.0868217
            |  7.9747307  7.8571886 | 7.9185624
            |        325        320 |       645
------------+-----------------------+----------
     Total  |  8.2263682  8.2941919 | 8.2600251
            |  8.6339238   7.982493 | 8.3118934
            |        402        396 |       798
```

Figure 3: Table of full time employment mean, standard deviation, and restaurant frequency, in that order, broken down by time period and state.

Question 11

Using these statistics, calculate a DD estimate of the impact of the minimum wage law on employment.

We seek to estimate the ATT, which is given by

$$\begin{aligned} ATT &= (\mathbb{E}[empft|NJ, t=2] - \mathbb{E}[empft|NJ, t=1]) - (\mathbb{E}[empft|Penn, t=2] - \mathbb{E}[empft|Penn, t=1]) \\ &= (8.446875 - 7.732308) - (7.651316 - 10.31169) \\ &= 3.374941. \end{aligned}$$

Thus, the increase in minimum wage in New Jersey appears to cause an increase in employment among New Jersey restaurants of about 3.3749 full-time employees.

Question 12

Specify and estimate the corresponding regression.

I estimate the following regression for the DD:

$$empft_{itk} = \beta_0 + \beta_1 state_{itk} + \beta_2 post_{itk} + \beta_3 state_{itk} \times post_{itk} + \nu_{itk}.$$

As displayed in Table 3, $beta_3 = 3.37494$ (t-stat $= 2.26$), which exactly matches the DD estiate from question 11.

Table 3: Difference-in-difference regression for employment on minimum wage increase table

	(1)
	full-time employees
1 if New Jersey; 0 if Pennsylvania	-2.579*
	(-2.45)
1 if after the law; 0 if before the law	-2.660*
	(-1.98)
state_post	3.375*
	(2.26)
Constant	10.31***
	(10.91)
Observations	798
R^2	0.008
df_m	3
df_r	794
rmse	8.293
mss	454.0
rss	54608.8

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 13

How much does this suggest that the minimum wage affects full time employment in fast food restaurants?

The DD suggests that the $\$0.80$ raise in minimum wage in New Jersey increased full-time employment among these fast-food restaurants by 3.374941 employees.

Question 14

Explain why the t-test from the regression above may understate the uncertainty in the effect of the minimum wage on full time employment. How could you correct the standard error? Compare the t-values with and without this correction.

The issue is that we We might have serially correlated errors ν_{ikt}

- Within state k,
- Within time period t, or
- Within state x time period.

In any of these cases, the serial correlation will deflate (downwardly bias versus the true standard errors) the OLS standard errors leading to an unwanted decrease in size and increase in power. The issue is analogous to 2SLS where the second stage standard errors underestimate the true uncertainty in parameter estimates because they don't recognize that the fitted values from the first stage are estimated with noise. Here, the issue is that common within group or group x time period shocks don't get averaged out by adding more observations, and the standard errors don't account for this problem.

If we had more groups we could cluster or do block bootstrap. However, neither of these options is feasible with only two groups and two time periods. Instead, I run the DD regression from question 12 with heteroskedastic robust standard errors (, r in Stata). Note that heteroskedasticity is likely an issue because there's no reason to expect that the residual variance is constant across groups or time periods. As shown in Tabble 4, the t-statistic on state_post is now 2.00, less than the non-robust t-statistic of 2.26 in Table 3 in Question 12.

Table 4: Difference-in-difference regression for employment on minimum wage increase table

	(1)
	full-time employees
1 if New Jersey; 0 if Pennsylvania	-2.579*
	(-1.98)
1 if after the law; 0 if before the law	-2.660
	(-1.70)
state_post	3.375*
	(2.00)
Constant	10.31***
	(8.41)
Observations	798
R^2	0.008
df_m	3
df_r	794
rmse	8.293
mss	454.0
rss	54608.8

Robust t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 15

What regression would you run to estimate the DD model including control variables? Run the regression using robust standard errors.

We can add controls by running the following regression

$$empft_{itk} = \beta_0 + \beta_1 state_{itk} + \beta_2 post_{itk} + \beta_3 state_{itk} \times post_{itk} + \beta_4' X_{ikt} + \beta_5' X_{ikt} \times state_{itk} + v_{itk},$$

where here X_{ikt} includes number of registers and hours of operation. As exhibited in Table **??** we find that here $\beta_3 = 3.651$, which is very similar to the DD regression estimate form question 12 (3.37494).

Table 5: Difference-in-difference regression for employment on minimum wage increase with controls table

	(1)
	full-time employees
1 if New Jersey; 0 if Pennsylvania	-1.447
	(-0.17)
1 if after the law; 0 if before the law	-2.873
	(-1.93)
state_post	3.651*
	(2.27)
number of cash registers in store	0.0792
	(0.11)
number hrs open per day	1.385**
	(3.25)
1 if KFC; 0 otherwise	-0.651
	(-0.24)
1 if Roy Rogers; 0 otherwise	-3.012
	(-1.73)
1 if Wendy's; 0 otherwise	4.177
	(1.37)
nregs_state	0.423
	(0.52)
hrsopen_state	-0.203
	(-0.42)
d2_state	2.046
	(0.68)
d3_state	1.513
	(0.79)
d4_state	-0.165
	(-0.05)
Constant	-10.14
	(-1.40)
Observations	775
R^2	0.139
df_m	13
df_r	761
rmse	7.768
mss	7411.7
rss	45924.1

Robust *t* statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 16

How might you test the key identifying assumptions underlying your DiD-estimation in this application, and in general?

The key assumption of a common trend between the treatment and control groups is inherently untestable, because we don't observe the counterfactual no-treatment potenetial outcome of the treatment group after treatment has been administered (i.e. here we don't see what employment in New Jersey would have been in the second period if the minimum wage had not been raised). However, here and in general, we can look at the pre-treatment values of the outcome variable over time (employment here) to see if both groups display similar time trends. If so, then we are more confident, though of course we can't be certain, that the two groups will have similar non-treated potential outcomes after the treatment date as well. If you have the data, you can also examine the post trends to see if several years after the treatment the employment trends in the two states is similar. We only observe two periods in this data set, so we can examine neither the pre- nor post trends. However, Card Krueger 2000 examines the post trends of this data set (shown in Figure 4) and concludes that the assumption of common trends is violated.

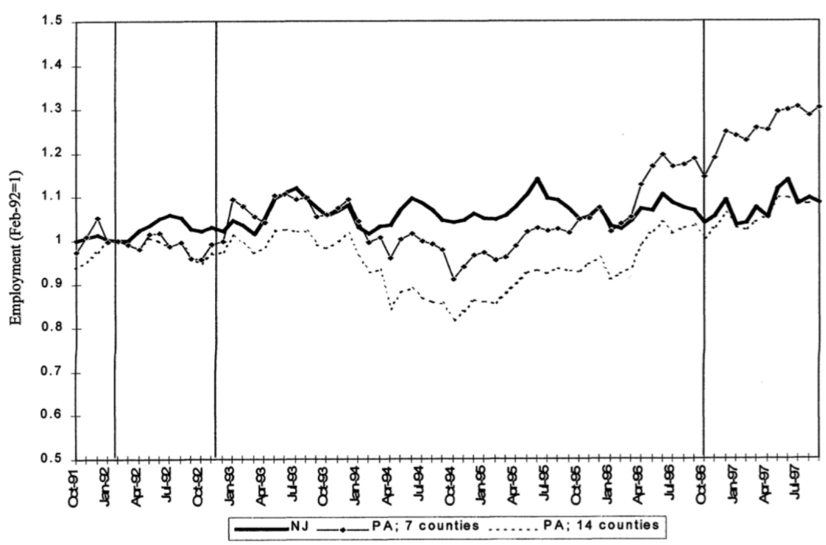

Figure 4: Caption

Stata Code for all Parts

```
* Aditya Chaudhry
set more off
clear

use "PS5 data"

* Part 1 - describe data
eststo: codebook
esttab using ps5_part1.tex, replace label nostar ///
        title(Summary of Data table\label{tab:part_1})
eststo clear

* Part 2,3 - regress empft on constant, minwage, nregs, hrsopen, restaurant dummies
set level 90
eststo:  regress empft minwage nregs hrsopen d2 d3 d4 if post == 0
esttab using ps5_part2.tex, r2 scalars(N df_m df_r r2 rmse mss rss) replace label
///
        title(OLS Regression: full time employment on minimum wage, with controls t


eststo clear



* Part 6, test equality of coefficients on d2, d3
eststo:   test d2 = d3 = 0
estout using ps5_part6.tex, replace label  ///
        title(Hypothesis Test of $\eta_1 = \eta_2 = 0$ table\label{tab:part_6})
eststo clear

* Part 7
eststo:  test d2 = d3
esttab using ps5_part7.tex, replace label nostar ///
        title(Hypothesis Test of $\eta_1 = \eta_2$ table\label{tab:part_7})
eststo clear

matrix r = (0, 0, 0, 1, -1, 0, 0)
```

```
matrix Sigma = e(V)
matrix standard_variance = r * Sigma * r'

gen standard_error = sqrt(standard_variance[1,1])
matrix point_estimate_matrix = r * e(b)'
gen point_estimate = point_estimate_matrix[1,1]
gen test_stat  = (point_estimate /  standard_error)^2 //Asymptotically stadnard nor
gen p_value =  Ftail( 1, e(df_r), test_stat)

disp test_stat
disp p_value

* Part 10

eststo:  tabulate state post, summarize( empft)
esttab using ps5_part10.tex, replace label nostar ///
        title(Full time employment mean, standard deviation, frequency broken down
table\label{tab:part_10})
eststo clear

* Part 11

disp (8.446875 -  7.732308 ) - (7.651316 -  10.31169 )


* Part 12
gen state_post = state * post
eststo: regress empft state post state_post
esttab using ps5_part12.tex, r2 scalars(N df_m df_r r2 rmse mss rss) replace label
///
        title(Difference-in-difference regression for employment on minimum wage ir
eststo clear

*Part 14
```

```
eststo: regress empft state post state_post, r
esttab using ps5_part14.tex, r2 scalars(N df_m df_r r2 rmse mss rss) replace label
///
        title(Difference-in-difference regression for employment on minimum wage ir
eststo clear


*Part 15, add controls
gen nregs_state = nregs * state
gen hrsopen_state = hrsopen * state
gen d2_state = d2 * state
gen d3_state = d3 * state
gen d4_state = d4 * state

eststo:  regress empft state post state_post nregs hrsopen d2 d3 d4 nregs_state
hrsopen_state d2_state d3_state d4_state, r
esttab using ps5_part15.tex, r2 scalars(N df_m df_r r2 rmse mss rss) replace label
///
        title(Difference-in-difference regression for employment on minimum wage ir
eststo clear
```