POLS 207

Problem Set 4*

Villaseñor-Derbez, J.C.

Difference-in-Differences

Assume that the fast-food restaurants surveyed by Card and Krueger represent a random sample from a larger population of all fast-food restaurants in New Jersey and eastern Pennsylvania.

State	February	November
NJ	α	β
PA	γ	$delta$

Define the difference-in-differences (DID) estimand in terms of these values, and concisely explain what this represents.

Given values that represent group- and time-specific means α, β, γ, δ, the DID estimand is the difference in time trends observed for each state. We first calculate the temporal trend as the November value minus the February value, for each state. For NY, the temporal trend is given by $\beta - \alpha$ and for PA the temporal trend is given by $\delta - \gamma$. We assume NJ and PA behave similarly through time, and that any deviation from the expected trend is caused by the change in wages, which only affects NJ. In other words we can calculate the difference as what we observe (for NJ) minus what one would have expected, had the intervention not occurred (NJ in Nov + trend of PA). Therefore, the difference in the temporal trends is our DiD estimand: $(\beta - \alpha) - (\delta - \gamma)$.

Consider the eight potential quantities $\{\alpha(0), \alpha(1), \beta(0)...\}$. Let these represent the mean potential level of FTE employment levels that would have realized if the minimum wage had been raised in each state at each time. For example, $\alpha(0) = \mathbb{E}[Y_i(D = 0)|T = February, state = NJ]$. Define the causal quantity of interest, the ATT, in terms of these potential outcomes. Describe which of these are observed.

The ATT is given by:

$$
\begin{aligned}
ATT &= \beta(1) - \beta(0) \\
&= \mathbb{E}[Y_i(D = 1)|T = Feb, state = NJ] - \mathbb{E}[Y_i(D = 0)|T = Feb, state = NJ] \\
&= \mathbb{E}[Y_1(1) - Y_0(1)|D = 1]
\end{aligned}
$$

The potential untreated outcome for the treated or the FTE levels for NJ in February, had NJ not received the treatment ($\beta(0)$) is an unobserved quantity.

*Available on GitHub: https://github.com/jcvdav/POLS207/blob/master/ps4/ps4.Rmd

Show that the DID estimand (part (i)) can identify the ATT (part (ii)). Show all your work and be sure to clearly state the assumption(s) needed.

We do not observe what FTE in NJ would have been in NJ, had NJ not been subject to a change in minimum wage (*i.e.* treated). In part i we stated that the estimand of interest was given by $DiD = (\beta - \alpha) - (\delta - \gamma)$. Based on table 1, these observed quantities correspond to:

State	February	November
NJ	$\mathbb{E}[Y_0(0)\|D = 1]$	$\mathbb{E}[Y_1(1)\|D = 1]$
PA	$\mathbb{E}[Y_0(0)\|D = 0]$	$\mathbb{E}[Y_0(1)\|D = 0]$

If we are willing to assume that in the absence of treatment NJ would have had the same time trend as PA, we can leverage our counterfactual design to estimate the ATT. In other words, we assume that the observed value for NJ in February plus the time trend observed for PA gives us what NJ would have been had it not raised minimum salary. We can then use the difference between the expected and the realized value to estimate our ATT.

Taking the terms in the table above and rearranging them to match the DiD, we would obtain:

$$\begin{aligned} DiD &= (\beta - \alpha) - (\delta - \gamma) \\ &= [\mathbb{E}[Y_1(1)|D = 1] - \mathbb{E}[Y_0(0)|D = 1]] - [\mathbb{E}[Y_0(1)|D = 0] - \mathbb{E}[Y_0(0)|D = 0]] \end{aligned}$$

The assumption that we made above is that the different groups have parallel time trends:

$$\mathbb{E}[Y_0(1) - Y_0(0)|D = 1] = \mathbb{E}[Y_0(1) - Y_0(1)|D = 0]$$

```
# Setup
suppressPackageStartupMessages({
  library(here)
  library(magrittr)
  library(tidyverse)
})
```

```
# Load the data
load("card_krueger.RData")
d_long <- d %>%
  select(state, chain, emp_pre, emp_post) %>%
  gather(time, emp, -c(state, chain)) %>%
  mutate(time = str_remove(time, "emp_"),
         time = fct_relevel(time, "pre"))
```

Estimate all quantities in above and write them in a similar table

```
d_long %>%
  group_by(state, time) %>%
  summarize(emp = mean(emp, na.rm = T)) %>%
  spread(time, emp) %>%
  knitr::kable(booktabs = T,
               digits = 2,
               caption = "Estimates for quantities of interest in Table 1.")
```

Table 3: Estimates for quantities of interest in Table 1.

state	pre	post
nj	20.44	21.03
pa	23.33	21.17

Then in a second table, fill in all cells in the following variance-covariance matrix

Formulas

The variances in the diagonal are given by:

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2]$$

The covariances in all other cells are given by:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$$

Assumptions

We assume that PA and NJ are independent of each other, so $\mathrm{Cov}(\gamma,\alpha)=0$. The same is true for $\mathrm{Cov}(\delta,\alpha)$, $Cov(\gamma,\beta)$ $\mathrm{Cov}(\delta,\beta)$.

Assert equivalence

Finally, since $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$, corresponding cells have the same values ($i.e.$ $A_{1,2}=A_{2,1}$).

With the rules above, we can define our variance-covariance matrix as follows:

```r
# First, calculate the diagonal with variances
vars <- d_long %>%
  group_by(state, time) %>%
  summarize(emp = var(emp,
                      na.rm = T))

#Covariance between alpha and beta
nj_cov <- d %>%
  filter(state == "nj") %$%
  cov(emp_pre, emp_post,
      use = "pairwise.complete.obs")

# Covariances between gamma and delta
pa_cov <- d %>%
  filter(state == "pa") %$%
  cov(emp_pre, emp_post,
      use = "pairwise.complete.obs")
```

```r
# Create an empty matrix
mat <- matrix(data = 0, nrow = 4, ncol = 4)

# Fill-in the diagonal
diag(mat) <- vars$emp

# Fill in the covariances
mat[1, 2] <- nj_cov
mat[2, 1] <- nj_cov
mat[3, 4] <- pa_cov
mat[4, 3] <- pa_cov

# Assign columnames and create table
colnames(mat) <- c("alpha", "beta", "gamma", "delta")
rownames(mat) <- colnames(mat)

knitr::kable(mat,
             booktabs = T,
             digits = 2)
```

	alpha	beta	gamma	delta
alpha	82.92	50.72	0.00	0.0
beta	50.72	86.36	0.00	0.0
gamma	0.00	0.00	140.57	48.3
delta	0.00	0.00	48.30	68.5

Calculate (i) your DID estimate, and (ii) the estimated variance of the DID estimate. Calculate a two-sided p-value. Interpret your results.

Based on the quantities estimated in table 3, the DiD estimate is:

$$
\begin{aligned}
DiD &= (\beta - \alpha) - (\delta - \gamma) \\
&= (21.03 - 20.44) - (21.17 - 23.33) \\
&= (0.59) - (-2.16) \\
&= 0.59 + 2.16 \\
&= 2.75
\end{aligned}
$$

```r
# re-level PA as reference
d_long_levels <- d_long %>%
  mutate(state = fct_relevel(state, "pa"))

lm(emp ~ state * time, data = d_long_levels) %>%
  stargazer::stargazer(se = estimatr::starprep(., se_type = "HC1"),
                       t.auto = T,
                       p.auto = T,
                       single.row = T,
                       header = F,
                       covariate.labels = c("NJ", "Post", "DiD"),
                       dep.var.labels = "FTE",
                       title = "OLS estimates of the DiD.")
```

Table 5: OLS estimates of the DiD.

	Dependent variable:
	FTE
NJ	-2.892^{**} (1.439)
Post	-2.166 (1.641)
DiD	2.754 (1.795)
Constant	23.331^{***} (1.346)
Observations	794
R^2	0.007
Adjusted R^2	0.004
Residual Std. Error	9.406 (df = 790)
F Statistic	1.964 (df = 3; 790)
Note:	*p<0.1; **p<0.05; ***p<0.01

Based on the DID estimate and DID_OLS estimate (Table 5), the interpretation is that an increase in wages in NJ caused a 2.75 increase in FTE in NJ, relative to the changes observed in PA. This estimate, however, is not significant.

Instrumental Variables

In class we talked about instrument variables in the potential outcomes framework. Besides the monotonicity ("no defiers") assumption and relevance ("first stage") assumption we required an ignorability assumption in the form:

$$\{Y_{1i}, Y_{0i}, D_{1i}, D_{0i}\} \perp Zi$$

where Y_{1i}, Y_{0i} are treatment and non-treatment potential outcome, D_{1i}, D_{0i} are the potential treatment status under encouragement or non-encouragement values of the instrument, and Z_i is the observed assignment of the encouragement (instrument).

Explain which parts of this assumption are required for you identify the first-stage effect. Prove that you can identify the first stage using the assumption you state.

The first stage is given by:

$$D = \tau + \rho Z + \eta$$

which is the effect of the instrument Z_i to the treatment uptake. Here, the ignorability assumption combines the independence and exclusion assumptions. The first part of this (independence) is needed to identify the first stage effect.

We know that the Wald estimator is given by

$$\frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

With the denominator being our first-stage estimation (effect of Z on D).

The independence assumption allows us to write the denominator as $\mathbb{E}[D_1] - \mathbb{E}[D_0]$, so the above becomes:

$$\frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D_1 - D_0]}$$

And since we know that $\mathbb{E}[Y|Z=0]\mathbb{E}[D_1 - D_0] = p(D_1 > D_0)$, we have essentially obtained the probability of treatment needed in the Wald estimator.

Explain which parts of this assumption are needed to identify the reduced form or intent-to-treat estimate. Again prove that you can identify the reduced form estimate using the stated assumption.

The estimate is given by $ATE_z = \mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]$, which can be identified through the exclusion restriction assumption. Therefore, we can remove the conditional statements in the numerator.

Explain what the exclusion restriction means, and why it too is implied by the ignorability assumption above.

A good instrument can affect our outcome of interest Y only through a dependent variable, but not directly. The exclusion restriction assumption implies that there are no other causal pathways in which the instrument may affect our outcome other than through a dependent variable.

If the exclusion restriction is violated, is the intent-to-treat effect still valid? Why or why not?

The ATT_z is still valid if Z is exogenous, but the interpretation of the causal quantity becomes complicated. We will be able to say something about the direction and magnitude in which the outcome of interest changed, but it will be harder to say anything about the process that generated the causality.

Suppose a colleague who has not taken PS200D asks you to explain why the Wald estimate – the reduced form divided by the first stage effect – identifies the ATE among the compliers. Provide a clear and intuitive explanation.

Normally we would want to randomly assign a treatment in order to estimate the effect that the treatment has. However, we can not always randomly assign it. A way around this is to use instrumental variables. A good instrumental variable randomly assigns units into treatment but does not directly modify the outcome. Now we have two steps: the process by which an instrumental variable assigns treatment at random, and the effect that being assigned to treatment has on the outcome, given potential outcomes.

Explain why we do not need to write the potential outcomes with two arguments, i.e. as $Y_i(D, Z)$ for treatment D and instrument Z.

Our ignorability and exclusion restriction assumptions states that Y is only affected by D and independent on Z. The effect of Z is already accounted for in D.

Suppose you have knowledge that Z_i is randomly assigned. Does this ensure you satisfy the ignorability assumption? Why or why not?

It is sufficcient to estimate the ATT_z, but not the ITT. We would also need to know that the only causal pathway of Z on Y is through D.