PPOL 503-03, PPOL 503-04, Fall 2016
Course Notes #16: Difference-in-Differences Method

I. Overview

Suppose GPPI imposes a new requirement of a tough new Quant 4 course for all students who enter the program in Fall 2004 (who will graduate in 2006). **Research Question:** Is the new GPPI Quant 4 requirement effective, in terms of raising starting salaries of GPPI graduates?

Counterfactual Option #1: Use <u>GPPI students graduating in the previous year</u> who were not subject to the Quant 4 requirement. Then, compare the average starting salary of GPPI students who *were* subject to the new requirement (i.e., those graduating in 2006) to the average starting salary of GPPI students who *were not* subject to the new requirement (i.e., those graduating in 2005). The estimate of the mean effect of the new Quant 4 requirement is the difference between these averages.

- What is the problem with this strategy? [This strategy does not take into account the time effects (or shocks) that might have led to changes in salaries that had nothing to do with the new GPPI Quant 4 requirement. For example, the cohorts who entered GPPI in these years may be different; or something happened to the economy over these years. Therefore, any effect found cannot be solely attributed to the additional course requirement.]

Counterfactual Option #2: Use <u>Harvard JFK students graduating in the same year</u> as GPPI students subject to the new course requirement. Then, compare the average starting salary in 2006 of GPPI students who were subject to the Quant requirement to the average starting salary of Harvard JFK graduates in 2006. (Assume that the JFK students do not have the Quant 4 requirement.) The estimate of the average effect of GPPI's new Quant 4 requirement is the difference between these averages.

- What is the problem with this strategy? [This strategy does not take into account the possibility of school effects – differences in students or programs at the two schools that had nothing to do with the new GPPI course requirement – that might have led to differences in the starting salaries. For example, the students who go to GPPI and Harvard may be different, leading to different starting salaries.]

Counterfactual Option #3: Assume that without the new GPPI program, the difference between 2005 and 2014 GPPI graduating students' starting salaries would be the same as the difference between the 2005 and 2006 JFK students' starting salaries. That is, while GPPI and JFK may be different and may lead to different starting salaries, those differences remain constant over time. Any difference in the differences found from 2005 to 2006 can then be attributed to the effect of GPPI's new Quant 4 requirement. This strategy is the **difference-in-difference approach**: it first takes the difference within GPPI and within JFK across years, and it then takes the difference between these two differences. This strategy addresses the drawbacks of counterfactuals #1 and #2 above by accounting for both time and school effects.

- What is the problem with this strategy? [This strategy does not account for interactions between time and school; i.e., determinants of starting salaries that change over time differently for each school.]

Table: Difference-in-Difference Hypothetical Example: Average Starting Salaries of GPPI and JFK Students

Group:	GPPI Students	JFK Students	Difference (GPPI-JFK)
Graduates in 2006 (GPPI students subject to new Quant 4 requirement)	$60,000	$52,000	$8,000
Graduates in 2005 (GPPI students *not* subject to new Quant 4 requirement)	$50,000	$45,000	$5,000
Difference (2006-2005):	$10,000	$7,000	**$3,000**

Point estimate of GPPI new Quant 4 requirement's average effect using Counterfactual #1: $10,000

Point estimate of GPPI new Quant 4 requirement's average effect using Counterfactual #2: $8,000

Point estimate of GPPI new Quant 4 requirement's average effect using Counterfactual #3 (D-in-D): $3,000

[Note that a standard error should be tested for statistical significance. We omit that step here, because the main point is to convey the concept of difference-in-difference.]

II. Difference-in-Differences Estimate in Regression Analysis

- The difference-in-differences estimate in the table above could also be estimating using OLS. Using Meyer's (1995) notation, the OLS equation is as follows:
  - $Y_{it}^{j} = \alpha + \alpha_1 d_t + \alpha^1 d^j + \beta d_t^j + \varepsilon_{it}^{j}$,

  where:

  Y_{it}^{j} = outcome for individual *i* in time *t* in group *j* (treatment or control)
  d_t = 1 if the observation is measured in *t = 1* (post-program); zero otherwise
  d^j = 1 if the observation is in group *j = 1* (treatment group); zero otherwise
  d_t^j = 1 if the observation is in group *j = 1* in time *t = 1*; zero otherwise.

- β is the difference-in-difference impact estimate (assuming that the model assumptions hold)

- One can (and should) also include other observable characteristics in the OLS model above. This may help control for determinants that change by school by time. Including these variables also improves the efficiency of the estimates.

EXAMPLE 1:

- We have data for North Andover on *real* housing prices and whether the house is close to (within 3 miles) of the garbage incinerator. Construction of the incinerator began in 1981.

- We have data for 1978 (before) and 1981 (after). Note that this is a *pooled cross section*, not a *panel* (or *longitudinal*) data set.

  - Independent pooled cross section: Random samples drawn at different times. We do not follow the same individuals through time. Likely to lead to observations that are not identically distributed.

  - Panel data: We follow the same observations through time. We cannot assume that the observations are independently distributed. Important aspect of panel data is the ability to remove the time constant or fixed unobservable attributes.

- The simplest model uses 1981 data to get estimates:

- $$rprice = 101{,}307.5 - 30{,}688.27 nearinc$$
  $$(3{,}093.0) \quad (5{,}827.71)$$
  $$n = 142, R^2 = .165$$

  Where *nearinc* is a dummy variable equal to one if the house is near the incinerator, and zero otherwise.

  - Can we infer that the incinerator caused the average price of a house near the incinerator to go down by $30,688.27 relative to a house far from the incinerator?

- If we run the same regression with the 1978 data, we get:
  - $$rprice = 82{,}517.23 - 18{,}824.37 nearinc$$
    $$(2{,}653.79) \quad (5{,}827.71)$$
    $$n = 179, R^2 = .082$$
  - [This implies that the incinerator was built in an area with lower housing values, so we can't make a causal inference using the 1981 estimates.]

- A more credible estimator would look at how the coefficient on *nearinc* changed between 1978 and 1981. That is, we want to examine how the average price difference between far houses and close houses changed after the beginning of construction of the incinerator. This is the difference-in-differences estimator, computed as:

  - $$\hat{\delta}_1 = (\overline{rprice}_{81,nr} - \overline{rprice}_{81,fr}) - (\overline{rprice}_{78,nr} - \overline{rprice}_{78,fr})$$

- The difference in average housing price was much larger in 1981 than in 1978:
- $$\hat{\delta}_1 = -30{,}688.27 - (-18{,}824.37) = -11{,}863.9$$

- We can estimate this using a regression format, which also gives us the standard error with which to test for significance.
  - $$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 \cdot nearinc + u$$

β_0 = average price of a home not near the incinerator in 1978

δ_0 = capture the change in all housing prices in Andover from 1978 to 1981

β_1 = coefficient on nearinc; it measures the location effect that is not due to the presence of the incinerator

δ_1 = coefficient on the interaction term; it measures the decline in housing values due to the new incinerator

Dependent Variable = *rprice*

Independent Variable	Model 1	Model 2	Model 3
Constant	82,517.23** (2,726.91)	89,116.54** (2,406.05)	13,807.67 (11,166.59)
Y81	18,790.29** (4,050.07)	21,321.04** (3,443.63)	13,928.48** (2,798.75)
nearinc	-18,824.37** (4,875.32)	9,397.94* 4,812.22	3,780.34 (4,453.42)
Y81 · nearinc	-11,863.90 (7,456.65)	-21,920.27** (6,359.75)	-14,177.93** (4,987.27)
Other controls	NO	Age, Age^2	Full set
Observations *R-squared*	321 .174	321 .414	321 .660

Model 3 controls for distance to the interstate, land area, house are, number of room, and number of baths

δ_1 = -14,178 but has t-statistic = -2.84

Model 3 is preferred because it controls for most factors and has the smallest errors. The fact that the variable "nearinc" is not significant in model 3 indicates the other controls capture the housing characteristics that are most important in determining housing prices.

<u>EXAMPLE 2</u>

- In 1980, Kentucky raised the cap on weekly earnings that were covered by workers' compensation. An increase in the cap has no effect on the benefit for low-income workers, but it makes it less costly for a high-income worker to stay on workers' compensation. The control group is low-income workers and the treatment group is high-income workers; high-income workers are defined as those who are subject to the prepolicy change cap. Does having more generous workers' compensation benefits cause people to stay out of work longer?

- $$\underset{}{\log(durat)} = \underset{(.031)}{1.126} + \underset{(.0447)}{.0077}afchnge + \underset{(.047)}{.256}highearn$$

  $$+\underset{(.069)}{.191}afchnge \cdot highearn$$

  $$n = 5,626, R^2 = .021$$

  Afchange = 1 if the period after the benefit increase; = 0 for the period before the benefit increase

  Highearn = 1 if high income worker; = 0 if low income worker

Afchange Highearn* = 1 if high income worker in the period after the change; = 0 if low income worker in the period before the change

The estimate of δ = .191 (t = 2.77), which implies that the average length of time on workers' compensation for high income workers increased by about 19% due to the increased earnings cap.

Afchange = is small and insignificant; the increase in earnings cap has no effect on duration for low-income workers.

EXAMPLE 3

To assess the impact of changing taxes on cigarette consumption, we can obtain random samples from 2 states for 2 years.
In state A, the control group, there was no change in the cigarette tax. In state B, the tax increased between the two years.

Outcome = cigarette consumption of the ith individual in year t

State B = 1 if person resides in state B; = 0 if state A

Year92 = 1 if year is 1992; = 0 if year is 1990

StateB Year92 = 1 if person reside in State B and year is 1992

$$C_{it} = \beta_0 + \beta_1 \text{ State B} + \beta_2 \text{ Year92} + \beta_3 \text{ StateBYear92}$$

III. Limitations of the Difference-in-Differences Method

The two main assumptions of the difference-in-differences method are parallel trends and common shocks.

1. The parallel trends assumption states that the trends in outcomes between the treated and comparison groups are the same prior to the intervention. If true, it is reasonable to assume that these parallel trends would continue for both groups even if the program was not implemented. This assumption is tested empirically by examining the trend in both groups before the policy was implemented. In a regression model, this is evaluated by assessing the significance of the interaction term between time and policy exposure in the pre-intervention period. If the trends are significantly different prior to the intervention, a difference-in-differences analysis will be biased and a different comparison group should be sought.

2 In economics a shock is defined as an unexpected or unpredictable event (unrelated to policy) that affects a system. The common shocks assumption state that any events occurring during or after the time the policy changes will equally affect the treatment and control groups. A key limitation of the D-in-D design is finding a control group for which these assumptions are met. Ideally, the only difference between the treatment and control groups should be exposure to the policy change. In practice such a control group may be difficult to identify.

IV. Summary: Difference-in-Differences Estimation

1. The difference-in-difference approach can be applied to numerous applications especially those where the data arise from a natural experiment.

2. A natural experiment occurs when some exogenous event—often a change in government policy—changes the environment in which individuals, families, firms or cities operate. A natural experiment always has a control group, which is not affected by the policy change, and a treatment group, which is affected by the policy change.

3. In a true experiment the treatment and control groups are randomly assigned and explicitly chosen. The control and treatment groups in natural experiments arise from the policy change. To control for systematic differences between the treatment and control groups, one needs two years of data, one before the policy change and one after the policy change.

4. Sample is comprised of four groups: 1) control group before the change, 2) control group after the change, 3) treatment group before the change, and 4) treatment group after the change.

5. The two main assumptions of the D-in-D method are the parallel trends and common shocks. If either or both of these conditions are not satisfied then the D-in-D method will yield biased results.

IV. Difference-in-Difference-in Differences Estimator

- As Meyer (1995) points out, in some instances the treatment effects only impact a sub-group within the designated treatment group. This has the advantage of helping to control for other changes that are occurring over time within across groups.

EXAMPLE 4

- Ludwig (1998) tests whether a state passing a permissive handgun carrying law reduces homicide rates within the state. Some previous research claimed that being more permissive about allowing people to carry handguns decreases the likelihood that people will be murdered.

- A D-in-D approach could compare changes in homicide rate over time in states that passed the law to changes in homicide rate over time in states that did not pass the law. The problem with the D-in-D approach is that it does not control for unobserved factors that influence homicide rates that vary over time. For example, if crack use (which is unmeasurable) differs over time across states that did and did not pass the handgun law (e.g., if states with low crack rates were more likely to pass the permissive handgun law), then the D-in-D estimate will conflate these effects of crack on homicides.

- Ludwig uses *a D-in-D-in-D estimation* strategy that exploits the fact that juveniles living in the states that pass the law are unaffected by the law, since it only applies to adults. That is, the juvenile homicide victimization rate should be unaffected by the new law, since juveniles are not allowed to carry handguns, regardless of the permissiveness of the adult law. This allows him to use changes in juvenile homicide victimization rate in the treatment states as a control for changes that might affect homicide rates. In other words, one can test the treatment (permissive gun law) as follows:

Difference-in-Difference-in-Difference Model of Effect of Permissive Gun Laws on Homicide Rates

Group:	Post-Law	Pre-Law	Difference
States with Permissive Law:			
Adults	a	b	(a-b)
Juveniles	c	d	(c-d)
Difference	(a-c)	(b-d)	*(a-b)-(c-d)*
States without Permissive Law:			
Adults	e	f	(e-f)
Juveniles	g	h	(g-h)
Difference	(e-g)	(f-h)	*(e-f)-(g-h)*
D-in-D-in-D:			**[(a-b)-(c-d)]-[(e-f)-(g-h)]**

- This can be estimated using OLS:

$$
\begin{aligned}
Y_{it} = \beta_0 &+ \beta_1(Exper_i) + \beta_2(Adult_i) + \beta_3(Post_t) + \beta_4(Exper_i \times Adult_i) + \\
&\beta_5(Adult_i \times Post_t) + \beta_6(Exper_i \times Post_t) + \beta_7(Exper_i \times Post_t \times Adult_i) + u_{it},
\end{aligned}
$$

- where,

  Y_{it} = homicide victimization rate for person (either adult or juvenile) in state i, year t

  $Exper_i$ = 1 if state i enacts the permissive law during sample period, 0 otherwise

  $Adult_i$ = 1 if observation corresponds to adult victimization rate, 0 if juvenile rate

  $Post_t$ = 1 if observation occurs after law was passed, 0 otherwise.

EXAMPLE 5: Gavin, Farrelly and Simpson (1998) "Children's Use of Primary and Preventive care Under Medicaid Managed Care"

1. This study evaluates the impact of two mandatory primary care case management (PCCM) programs in Florida and New Mexico.

2. Florida: compared the experience of Medi-pass eligible Medicaid children in the four county pilot area around Tampa-St. Pete with the experience of children enrolled in Medicaid under the same eligibility groups in four comparable counties around Orlando. Pre-implementation period FY 1991(July 1990-June 1991) and the post-period was FY 1993 (July 1992 – July 1993). We estimated the impact of implementing a mandatory PCCM program in place of traditional FFS among Medicaid children who decline voluntary HMO coverage.

3. New Mexico: compared the experience of children in PCN-eligible eligibility categories in 19 non-metropolitan counties that had implemented in the program prior to 1993 with the experience of children in the same eligibility groups in 10 non-metropolitan that had implemented the program after 1993. Pre-period was calendar year 1990, the year prior to the implementation of PCN in the state and the post-period was 1993. They examined disabled and non-disabled categories separately.

4. Descriptive analysis: Difference-in-Difference

The program impact is measured as the difference in the changes from the pre- to the post-implementation periods between the PCCM waiver and non-waiver counties.

The DD is measured by subtracting the change in the measure of interest from the pre- to the post-period in the comparison counties

(C) from the change in the measure from pre- to post-period in the waiver counties (W);

$$DD = (Y_{post^*W} - Y_{pre^*W}) - (Y_{post^*C} - Y_{pre^*C})$$

A positive sign indicates that the measure increased more (or decreased less) in the waiver counties than in the non-waiver counties.

A negative sign indicates that it decreased more (or increased less) in the waiver counties compared with the non-waiver counties.

If an increase in the measure is desirable (ie preventive care) then we expect to find a positive sign on DD. If the desirable program effect is a decrease in the measure, as is the case with ER visits and ACSC hospitalizations, we hope for a negative sign on the DD.

5. Multivariate Analysis: Difference-in-Differences

A limitation of the tabular analysis is that it fails to control for other factors that may influence service use and costs (age, race, gender, enrollment duration). Both states experienced significant growth and changes in the composition of the child populations enrolled in Medicaid under the categories eligible for PCCM programs.

Basic analytical model:

$$Y_{it} = f\,(\alpha = \gamma_T Tt + \gamma_w W_{it} + \gamma_{TW} TW_{it} + \beta X_{it} + u_{it})$$

Where Y is the dependent variable, Y = 1 indicates the child was in compliance with the periodicity schedule; = 0 otherwise.

W = 1 if individual lived in a waiver county; = 0 OW

T = 1 if observation is post-period (T = 1) or pre-period (T = 0)

The program effect is estimated by the coefficient on the indicator variable TW that represents the interaction of the pre/post indicator T and the waiver/comparison group indicator W. This coefficient measures the difference between the waiver and non-waiver groups in the change in the outcome measure over time, holding X constant.

$$\gamma_{TW} = [\,(Y_{T=1, W=1} - Y_{T=0, W=1}) - (Y_{T=1,W=0} - Y_{T=0,W=0})\,]$$

6. Researchers found a significant number of eligible Medicaid beneficiaries in both states covered under Medicaid FFS for all or part of the post-period analysis years. Since the program's impact is likely to vary by level of participation (exposure) to the program, they reran each equation but replaced the TW variable with dichotomous variables for four mutually exclusive categories: 1) delayed enrollees—covered under FFS prior to their PCCM enrollment in the analysis year; 2) full-period enrollees were in PCCM for all months during the analysis year; 3) disenrollees-left PCCM and had at least 1 month in Medicaid FFS before the end of the analysis year; 4) non-participants-enrolled in FFS throughout.

Other controls include age, gender, race; eligibility category, enrollment duration county fixed effects. The results show the aggregate impact of the waiver and the differential impacts by level of participation after controlling for other confounding factors.

7. Florida Population Characteristics (Table 4): Children in waiver and non-waiver counties were comparable with respect to age, gender, Medicaid enrollment duration, and eligibility categories.

Waiver counties had proportionately more whites and proportionately fewer black and Hispanic children.

8. Florida Program Impacts (Tables 5 & 6):
Ambulatory Care Days: The percent of children with any ambulatory care days increased in both waiver and non-waiver counties, but the increase was nearly 2 percentage points higher in waiver counties. Multivariate results in Table 6 corroborate the descriptive findings in Table 5. After controlling for other confounding factors—there was a 1.9 percentage point increase in ambulatory care days for PCCM participants in waiver counties. Full PCCM participants were 21.1 % points more likely to have had any care, whereas non-participants were 28.6% points less likely to have had any care.

9. Florida Program Impacts (Tables 5 and 6):
Emergency Room Visits: An equal percent of children in Florida waiver and non-waiver counties had ER visits during 1991. While both groups experienced declines, the decline in the waiver counties exceeded the decline in the non-waiver counties by 7.5 % points. Multivariate results show an 8.8 % point greater decline in the % of waiver-county children with ER visits. The decline in ER was evident among both participants and non-participants in the PCCM program in the waiver counties. Full PCCM participants had an 11.2 % point decline in ER visits, while non-participants had a 6.2% point decline.

10. Florida Program Impacts (Tables 5 and 6):
Hospital Stays for ACSCs: Only 1.9% of PCCM eligible children in waiver counties and 1.7% of Medicaid children enrolled in eligible categories in non-waiver counties had any hospitalizations for ACSCs in FY91. The % of children with ACSC hospitalization only declined in waiver counties. The DD estimator is -.5 % points. Multivariate analyses show a -.1% point DD estimate between waiver and non-waiver groups.

11. Florida Program Impacts (Tables 5 and 6):
EPSDT Screening Visits: In FY91, 21.4% of children in waiver counties received recommended visits, compared to 15.8% of children in non-waiver counties. Because the percent of children who received such visits increased by about the same amount in both waiver and non-waiver counties, the DD estimate of the program impact is small (.8). The multivariate results show a .8 % point decline in the % of recommended screening visits completed among preschoolers in the waiver counties. The decline was concentrated among non-participants; full participants had a 1.7% point greater increase in the % of recommended visits.

12. New Mexico Population Characteristics (Table 7):
The population of non-disabled children in the waiver counties grew by 67%, while the increase in the non-waiver counties was 75%. Waiver county children were younger and they were comprised of proportionately more Hispanics and fewer Native Americans compared to non-waiver counties.

13. New Mexico Program Impacts (Table 8 and 9):
Ambulatory Care Days: A higher percentage of children in waiver counties had ambulatory care days compared to non-waiver counties, but the DD estimator of the program effects is small and not significant.

14. New Mexico Program Impacts (Tables 8 and 9):
Emergency Room Visits: In 1990 less than 25% of children in waiver counties and 22% of children in non-waiver counties had any ER visits. The %point increase in ER visits was higher for non-waiver counties so the DD estimate of the program effect was a decline of more than 3% points. Probit results indicate the program effect was 3.5% point decline after controlling for other confounding factors.

15. New Mexico Program Impacts (Tables 8 and 9):
Hospital Stays for ACSCs: 1.7% of children in waiver counties and 1.4% in non-waiver counties had a preventable hospitalization in 1990. Both groups had increases in the % of children with a preventable hospitalization but the increase was higher in non-waiver counties. Probit results indicate a -0.3 % point smaller increase in the % of children with ACSC hospitalization. Note, however, for both groups the % with ACSC hospitalization were lower prior to the implementation of the PCCM program.

16. New Mexico Program Impacts (Table 8 and 9):
EPSDT Screening Visits: Preschool children in waiver counties had higher EPSDT visit completion rates than preschoolers in non-waiver counties in both 1990 (44.7%) and 1993 (54.4%). The rates for non-waiver counties were 30.9% (1990) and 36.6% (1993). The DD estimator is 3.9% points ($p<.01$). Multivariate results show no program effect at the aggregate level. Recognizing differences in the level of participation shows that full PCCM participants completed 4.6% points more of recommended screening visits, whereas non-participants had 3.8% points fewer visits.

17. New Mexico Program Impacts (Table 8 and 9):
EPSDT Referrals: In 1990, referrals were made for further diagnosis for 10.5% of those residing in waiver counties and 12.8% of those residing in non-waiver counties. By 1993, the referral rate was 24.3% in waiver counties and 19.9% in non-waiver counties. The DD estimator is 6.7% points ($p<.01$). Multivariate probit results show the program effect is 6.9% points ($p<.01$). Significant increases in referrals were experienced by participants and non-participants.