Center for Effective Global Action
University of California, Berkeley

Module 2.5: Difference-in-Differences Designs

Contents

List of Figures

1. INTRODUCTION

In previous modules, we have argued that Randomized Control Trials (RCT) are a gold standard because they make a minimal set of assumptions to infer causality: namely, under the randomization assumption, there is no selection bias (which arises from pre-existing differences between the treatment and control groups). However, randomization does not always result in balanced groups, and without balance in observed covariates it is also less likely that unobserved covariates are balanced. Later, we explored Regression Discontinuity Designs (RDD) as a quasi-experimental approach when randomization is not feasible, allowing us to use a forcing variable to estimate the (local) causal effects around the discontinuity in eligibility for study participation. In RDD, we use our knowledge of the assignment rule to estimate causal effects.

In this module, we cover the popular quasi- or non-experimental method of Difference-in-Differences (DID) regression, which is used to estimate causal effect – under certain assumptions – through the analysis of panel data. DID is typically used when randomization is not feasible. However, DID can also be used in analyzing RCT data, especially when we believe that randomization fails to balance the treatment and control groups at the baseline (particularly in observed or unobserved effect modifiers and confounders). DID approaches can be used with multi-period panel data and data with multiple treatment groups, but we will demonstrate a typical two-period and two-group DID design in this module.

We present analytical methods to estimate causal effects using DID designs and introduce you to extensions to improve the precision and reduce the bias of such designs. We conclude the module with a discussion of Triple-Differences Designs (DDD) to introduce analysis allowing more than two groups or periods to be analyzed in DID designs.

The learning objectives of this module are:

- ✓ Understanding the basics of DID designs
- ✓ Estimating causal effects using regression analysis
- ✓ Incorporating "matching" techniques to improve precision and reduce bias in DID designs
- ✓ Introducing Triple-Differences Designs.

2. BASICS OF DID DESIGNS

Imagine that we have data from a treatment groups and a control group at the baseline and endline. If we conduct a simple before-and-after comparison using the treatment group alone, then we likely cannot "attribute" the outcomes or impacts to the intervention. For example, if income from agricultural activities increases at the endline, then is this change attributable to the agriculture-based intervention or to a better market (higher demand and price), season, or something else that the intervention did not impact? If children's health improved over time, is it simply because they are getting older and having improved immune system or because of the intervention? In many cases, such baseline-endline comparison can be highly biased when evaluating causal effects on outcomes affected over time by factors other than the intervention.

A comparison at the endline between the treatment and control groups, on the other hand, may also be biased if these groups are unbalanced at the baseline. DID designs compare ***changes over time*** in treatment and control outcomes. Even under these circumstances, there often exist plausible assumptions under which we can control for time-invariant differences in the treatment and control groups and estimate the causal effects of the intervention. Consider the following math to better understand the DID design concept.

- ✓ The outcome Y_{igt} for an individual i at time t in group g (treatment or control) can be written as a function of:

$$Y_{igt} = \alpha_g + \theta_t + \beta_1 G + \beta_2 t + \beta_3 G.t + U_{igt} + \varepsilon_{igt}$$

where α_g captures group-level time-invariant (not changing over time) "fixed effects" (think of these as distinct Y-intercepts of the baseline outcome for each group); θ_t captures period time-invariant fixed effects (e.g., election effects if the baseline was an election year); G is an indicator variable for treatment (=1) or control (=0) groups; t is an indicator variable for baseline (=0) or endline/ (=1) measurements, the βs are the regression coefficients to be estimated; U_{igt} captures individual-level factors that vary across groups and over time; and ε_{igt} captures random error. Let's denote the outcomes for the following four conditions as,

- ✓ At baseline in treatment group:

$$Y_{i10} = \alpha_1 + \theta_0 + \beta_1.1 + \beta_2.0 + \beta_3.1.0 + U_{i10} + \varepsilon_{i10}$$

- ✓ Individual at baseline in control group:

$$Y_{i00} = \alpha_0 + \theta_0 + \beta_1.0 + \beta_2.0 + \beta_3.0.0 + U_{i00} + \varepsilon_{i00}$$

- ✓ Individual at follow-up in treatment group:

$$Y_{i11} = \alpha_1 + \theta_1 + \beta_1.1 + \beta_2.1 + \beta_3.1.1 + U_{i11} + \varepsilon_{i11}$$

- ✓ Individual at follow-up in control group:

$$Y_{i01} = \alpha_0 + \theta_1 + \beta_1.0 + \beta_2.1 + \beta_3.0.1 + U_{i01} + \varepsilon_{i01}$$

- ✓ Change over time in outcome in treatment group = (4) – (2):

$$\begin{aligned} Y_{i11} - Y_{i10} &= (\alpha_1 + \theta_1 + \beta_1.1 + \beta_2.1 + \beta_3.1.1 + U_{i11} + \varepsilon_{i11}) \\ &\quad - (\alpha_1 + \theta_0 + \beta_1.1 + \beta_2.0 + \beta_3.1.0 + U_{i10} + \varepsilon_{i10}) \\ &= (\theta_1 - \theta_0) + \beta_2 + \beta_3 + (U_{i11} - U_{i10}) + (\varepsilon_{i11} - \varepsilon_{i10}) \end{aligned}$$

- ✓ Change over time in outcome in control group = (5) – (3):

$$Y_{i01} - Y_{i00} = (\theta_1 - \theta_0) + \beta_2 + (U_{i01} - U_{i00}) + (\varepsilon_{i01} - \varepsilon_{i00})$$

- ✓ The average treatment effect (or the DID impact) = (6) – (7)

$$(Y_{i11} - Y_{i10}) - (Y_{i01} - Y_{i00}) = \beta_3 + (U_{i11} - U_{i10} - U_{i01} + U_{i00}) + (\varepsilon_{i11} - \varepsilon_{i10} - \varepsilon_{i01} + \varepsilon_{i00})$$

- ✓ $\boldsymbol{DID\ Impact\ or\ ATE = \beta_3 + (U_*) + (\varepsilon_*)}$

The final equation specified clarifies the assumptions needed in order to infer causality from DID designs. First, we expect that the regression error term has a distribution with mean 0, so that ε_* is also distributed with mean 0. Second, we assume that the time-variant differences over time in the treatment and control groups are equal, thus cancelling each other out ($U_* = 0$). This is a critical assumption made in DID analysis, allowing for causal analysis despite the absence of randomization, and in some cases we may not believe it to be true.

The concept of DID is displayed in Figure 1. The solid red line shows how the outcome (some outcome of interest, measured in percentages) would change over time without the treatment (as measured in the control group), while the solid blue line displays the change over time in the treatment group. By shifting the red dotted line upwards from the solid red line, we remove the change over time attributable to other-than-treatment factors. Therefore, DID design estimates the outcome attributable to the intervention. However, if the assumption that the **changes in time-variant** factors in treatment and control groups are equal does not hold (known as the Parallel Trend Assumption), then the true control outcome could track the red dashed line. As the figure demonstrates, we could overestimate (or underestimate) the causal effect using DID if the above assumption is violated.

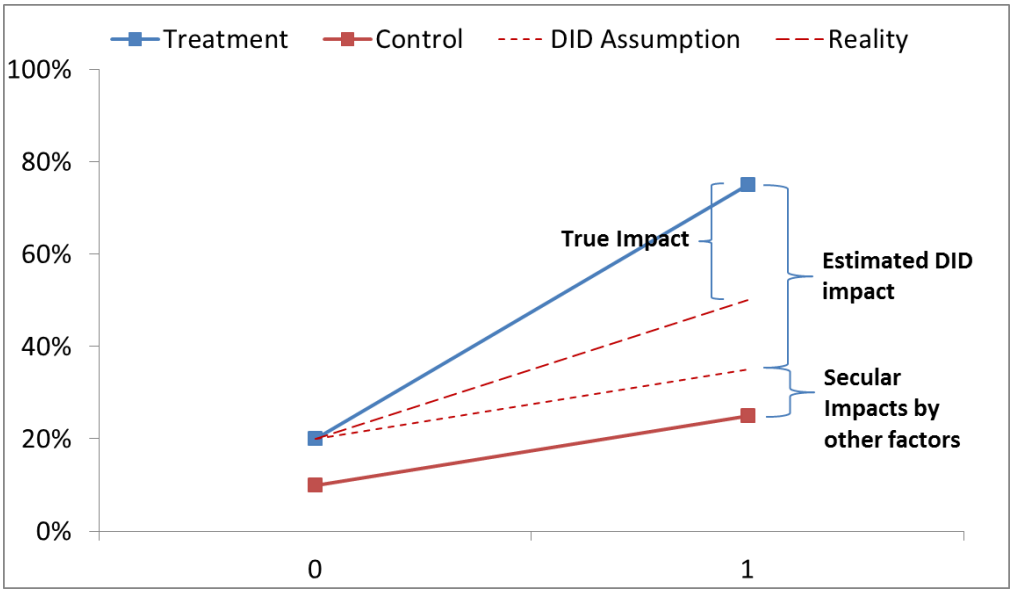

Figure 1. Graphical demonstration of difference-in-difference

It is possible to "control" for factors that may vary or change over time differently between the treatment and control groups in regression analysis but one can always be concerned about immeasurable or unmeasured factors causing time variant changes. Also, mathematically, DID can also be shown as subtracting from the mean difference at the endline between treatment and control groups the pre-existing differences in these groups at the baseline.

3. DEMONSTRATION: DID IN OPORTUNIDADES

We will demonstrate application of DID with dataset for OPORTUNIDADES (`Panel_OPORTUNIDADES_00_07_year.dta`). This is a panel dataset of household and individuals tracked in years 2000, 2003 and 2007. Year 2000 was actually the final year of a previous version of OPORTUNIDADES called PROGRESA, which we studied in Modules 2.2, 2.3, and 2.4. The PROGRESA treatment was randomized to 320 villages and 186 control villages. By the fall of 2000 all 506 treatment and control villages were included in OPORTUNIDADES. However, it wasn't decided to track the long term impacts of OPORTUNIDADES until 2003, but by that time the original controls had become the "treatment," leaving only one option: to find a new control group. The study team used matching methods to find 150 control villages for the 506 treatment villages in 2003.

For this demonstration, we will apply DID method to compare outcomes between treatment and control villages in 2000 with those in 2003. Interestingly, the "baseline" year of 2000 already has 320 villages, which were exposed to the treatment. This is different from the typical setting in which baseline measurements are collected prior to program activities in the treatment villages. This is a challenging case for DID because of such contamination in the baseline, as well as because control villages in the baseline receiving subsequent treatment (under OPORTUNIDADES) and each control's being matched to multiple treatment villages.

Please implement the following steps in STATA.

- ✓ We will only reproduce a part of the STATA code below; please refer to the DO file for the complete code and accompanied notes
- ✓ Open the dataset and create flags that identify unique villages and households in our sample. The code below cross-tabulates the treatment and control villages by year. Figure 2 shows the distribution of villages by year and comparison groups. We see that all villages in the 2000 sample become the treatment in 2003 and there were 150 additional controls as described previously.

```
egen uniqvill = tag(year villid)
egen uniqhh = tag(year villid hogid)
tab D year if uniqvill == 1, m
```

```
. tab D year if uniqvil, m
```

Village-Level Treatment status	year 2000	2003	2007	Total
Control	186	150	0	336
Treated	319	506	0	825
.	0	0	808	808
Total	505	656	808	1,969

Figure 2. Tabulation of the number of villages by treatment groups and years

- ✓ We need to create a few additional variables. We will create an indicator ("dummy") variable equal to zero in the initial period (2000) and equal to one in the final period (2003) (similar to the variable *t* in models presented in Section 2). You may have noticed in Figure 2 that the treatment assignment for 2007 is missing. Please refer to the DO file to see how we create a 2007 treatment assignment variable. In essence, we use household-level participation in OPORTUNIDADES (`D_HH`) to assess whether the village in which that household resided participated in OPORTUNIDADES. Since OPORTUNIDADES would not have been possible in control villages, we assign villages to the treatment group if they contain at least one household participating in the program, and assign the rest to the control group. The STATA code is given in the DO file. You will notice in Figure 3 that several 2003 control villages become 2007 treatment villages. This is expected in popular programs, where maintaining controls for a long time may not be feasible. However, we will be restricting our analysis below to years 2000 and 2003.

```
. tab D year if uniqvil, m
```

Village-Level Treatment status	year 2000	2003	2007	Total
Control	186	150	41	377
Treated	319	506	767	1,592
Total	505	656	808	1,969

Figure 3. Number of treatment and control villages in year 2007

- ✓ In Figure 4, we compare the distribution of children's education in the comparison groups in year 2000. The distributions overlap quite well for lower education levels. This is expected because changes in the number of years of education can be expected only in the long term. However, we notice that treatment villages may have had somewhat better outcomes in higher grades. This is expected, because we know that 2000 is not a "true" baseline and that several of treatment villages had benefitted from PROGRESA in the past.
- ✓ In Figure 5, we compare the same distributions in the follow-up year 2003. Here, we see that the treatment villages are certainly faring better than the controls. The STATA code is as follows,

```
twoway histogram edu_child if year==2000 & D == 1 ||
histogram edu_child if year==2000 & D == 0, fcolor(blue)
legend(lab(1 "Treatment") lab(2 "Control"))

twoway histogram edu_child if year==2003 & D == 1 ||
histogram edu_child if year==2003 & D == 0, fcolor(blue)
legend(lab(1 "Treatment") lab(2 "Control"))
```

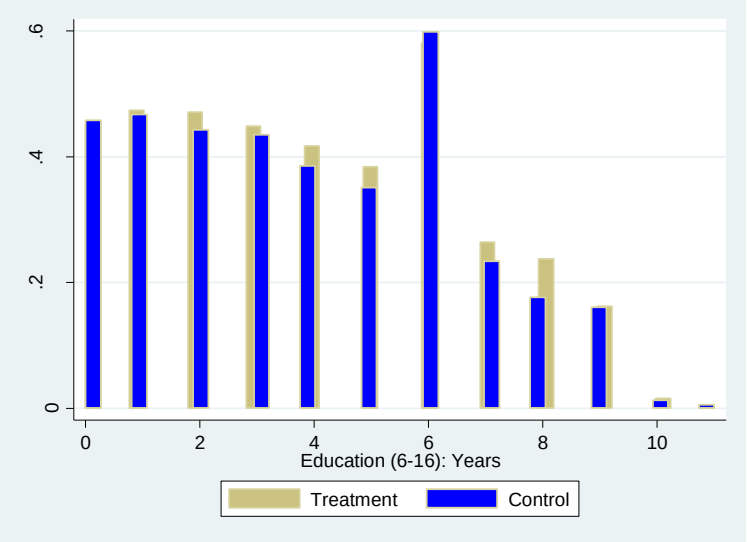

Figure 4. Distribution of number of years of child (6-16 years) education in year 2000

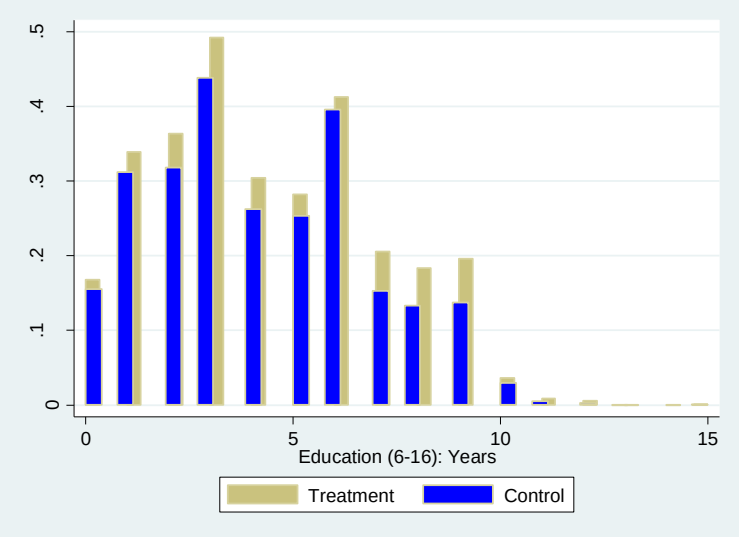

Figure 5. Distribution of number of years of child (6-16 years) education in year 2003

- ✓ Next, we check the baseline balance in covariates. We could use the standard t-test command, but we suggest you install the user-written command "diff" (`ssc install diff`), which is more convenient/informative for DID analysis. Below is the STATA code for the test of baseline balance using the 'diff' command (which estimates multiple t-tests).

```
diff edu_child, t(D) p(period) cov(age sex agehead sexhead) test
```

Figure 6 shows that the key covariates were reasonably balanced at the baseline in an economic sense, even though the differences are significantly different from 0 for the years of child education, age of head of the household, and age and sex of the child. However, we also know that 2000 was the final year of PROGRESA for this sample, for which reason there are certainly differences between treatment and control groups. Further, we tested only a few covariates, and an educational baseline test should be conducted simultaneously with several covariates and the outcomes of interest.

```
. diff edu_child, t(D) p(period) cov(age sex agehead sexhead) test

TWO-SAMPLE T TEST

Number of observations (baseline): 34906
            Baseline        Follow-up
   Control: 13699           -           13699
   Treated: 21207           -           21207
            34906           -
```

t-test at period = 0:

Variable(s)	Mean Control	Mean Treated	Diff.	|t|	Pr(|T|>|t|)
edu_child	3.815	3.896	0.080	2.76	0.0057***
age	25.459	25.226	-0.233	1.84	0.0656*
sex	0.494	0.503	0.008	2.63	0.0086***
agehead	47.334	46.855	-0.479	5.36	0.0000***
sexhead	0.917	0.918	0.001	0.63	0.5266

*** p<0.01; ** p<0.05; * p<0.1

Figure 6. Baseline balance in covariates and outcome of interest at year 2000

- ✓ To estimate the average treatment effect using DID method, we specify the following regression model: `reg edu_child D_period D period, vce(robust)` where `D_period` is an interaction variable created by multiplying the `D` and `period` variables (See Equation 1 in Section 2). Figure 7 is the output of the DID analysis. As discussed in Section 2, the interaction coefficient (β_3 in models presented earlier) provides the average treatment effect. We find an increase of 0.075 years of education for children in treatment villages compared to those in control villages.

```
. reg edu_child D_period D period, vce(robust)

Linear regression                               Number of obs =   82619
                                                F(  3, 82615) =  158.87
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0058
                                                Root MSE      =  2.6468
```

edu_child	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
D_period	.07501	.041586	1.80	0.071	-.0064983	.1565183
D	.0801867	.0289837	2.77	0.006	.0233789	.1369945
period	.3175853	.0348287	9.12	0.000	.2493213	.3858494
_cons	3.815461	.0225669	169.07	0.000	3.77123	3.859692

Figure 7. Regression results for DID analysis

✓ Let's discuss a way of mitigate bias resulting from baseline imbalance in DID analysis. We can do so by including covariates which (we believe) to have been imbalanced, or those which (we believe) could explain the imbalance between the groups well in the regression model specification. For demonstration sake, let's assume that age, sex of child, and the household head age and sex were imbalanced at the baseline. We re-estimate the DID model in Figure 8. Now the coefficient for the interaction term (D_period) is not statistically significant. The estimated magnitude of effect is also very small.

```
. reg edu_child D_period D period age sex agehead sexhead, r

Linear regression                               Number of obs =   77927
                                                F(  7, 77919) =18421.73
                                                Prob > F      =  0.0000
                                                R-squared     =  0.6802
                                                Root MSE      =  1.4966
```

edu_child	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
D_period	.0157462	.0239681	0.66	0.511	-.0312312	.0627235
D	.103376	.0148636	6.95	0.000	.0742434	.1325087
period	.2958656	.020499	14.43	0.000	.2556877	.3360435
age	.7102853	.0020115	353.11	0.000	.7063427	.7142279
sex	-.0966714	.0107214	-9.02	0.000	-.1176854	-.0756574
agehead	-.0013591	.0004906	-2.77	0.006	-.0023207	-.0003976
sexhead	.0940899	.0209877	4.48	0.000	.0529541	.1352256
_cons	-3.906229	.035701	-109.42	0.000	-3.976202	-3.836255

Figure 8. Regression results for DID analysis with covariates

✓ For the demonstration's sake, we present the output of the `diff` command in Figure 9 to show that the estimated impact is the same as that in Figure 8.

```
. diff edu_child, t(D) p(period) cov(age sex agehead sexhead)  robust

DIFFERENCE-IN-DIFFERENCES WITH COVARIATES

Number of observations in the DIFF-IN-DIFF: 77927
            Baseline       Follow-up
   Control: 12337          9418          21755
   Treated: 18615          37557         56172
            30952          46975

R-square:   0.6802
```

DIFFERENCE IN DIFFERENCES ESTIMATION

	BASE LINE			FOLLOW UP			
Outcome Variable	Control	Treated	Diff(BL)	Control	Treated	Diff(FU)	DIFF-IN-DIFF
edu_child	-3.906	-3.803	0.103	-3.610	-3.491	0.119	0.016
Std. Error	0.036	0.035	0.015	0.037	0.034	0.019	0.024
t	-109.42	-108.64	6.95	-98.12	-102.34	6.34	0.66
P>|t|	0.000	0.000	0.000***	0.000	0.000	0.000***	0.511

```
* Means and Standard Errors are estimated by linear regression
**Robust Std. Errors
**Inference: *** p<0.01; ** p<0.05; * p<0.1
```

Figure 9. Results of DID analysis with covariates using the diff command

4. MATCHING AND DID

In the introduction of the previous section, we briefly mentioned that in 2003, the evaluators included additional controls to study longer-term impacts of OPORTUNIDADES using matching methods. Here, we briefly describe how to perform propensity-score-based matching. Our main focus will be demonstrating the basic application; we will not discuss the theory behind propensity-score matching in this course.

We usually use quasi-experimental and matching-based methods to generate control groups when randomization is not feasible. In non-randomized treatment assignment, our main concern is selection bias. We tested for the presence of selection bias by evaluating the baseline balance in covariates, outcomes, and confounders. If the treatment and control groups are observationally similar at baseline, then we have higher confidence that the two groups preserve exchangeability. Matching is a statistical method of reducing baseline heterogeneity, and Propensity Score Matching (PSM) is one of the more popular matching techniques. Note that matching does not *always* produce matched groups that are more similar than randomly selected group (without matching) would have been.

The propensity matching method identifies treatment and control groups with similar probabilities (or ***propensity scores***) of being selected in a treatment. Therefore, PSM does not match villages (or individuals or households) directly on their observed characteristics, but instead matches them on their likelihood, conditional on observables, of being selected for treatment. PSM is most successful when the propensity scores in the true treatment and control groups are within the same wide range; this is called the **common support condition**. This condition fails when observable characteristics highly correlated with treatment are very different between the treatment and control groups; in these cases, PSM is not an effective quasi-experimental tool.

PSM is mainly used in two circumstances. The first and the most common use of PSM is when treatment villages are pre-selected and we need to find a control group. For example, this occurs whenever evaluation of an intervention commences after the intervention has been completed, without the pre-intervention specification of a control group. This often happens when impact evaluation is not planned along with the intervention and the opportunity for a baseline survey is missed. However, this does not suggest that PSM mitigates the need for baseline survey; on the contrary, the need for a baseline survey is even higher in the PSM framework, which requires that we we evaluate whether the PSM successfully results in balance in the observable coefficients.

Second, sometimes PSM precedes the randomization of treatment. We know that randomization works best when the number of treatment units is high. When we have only a few units available to randomize, chance could result in imbalanced groups at the baseline. PSM allows us to find a group of units which are similar to each other and then randomize the treatment within those groups.

4.1 Implementing PSM in STATA

We continue the demonstration using the dataset for OPORTUNIDADES from Section 2.4 in this module. PSM usually requires a large sample of treatment and control units, with PSM picking the best matched groups. However, for this demonstration we are working with a "smaller" dataset and use PSM to match groups over time (baseline-endline matching, instead of treatment and control groups matching at baseline). We acknowledge that the estimated impacts will not be causal (due to the small sample size), but we only seek to demonstrate PSM. The main sections of the STATA code are below, but we refer you to the DO file for the complete code.

- ✓ Create a variable in year 2000 to reflect the treatment status in 2003. This is done because we assume that we knew the 2003 treatment status at the baseline in order to construct propensity scores. We then keep only the observations from year 2000 in the dataset.

```
gen aux = D_HH == 1 & year == 2003

egen participation_03=max(aux),by(hogid2)
drop aux

keep if period==0
```

- ✓ Estimate propensity score and match households
  - Install a user-written STATA program called psmatch2 (`ssc install psmatch2`). Along with this command, two other programs (`pstest` and `psgraph`) automatically get installed.
  - We estimate the propensity scores and match treatment and control households on basis of household level covariates with: `psmatch2 participation_03 famsize agehead sexhead Income_HH_per, logit`. We could have included village or higher level of aggregate variables as well, but because the match is at household level, we cannot include individual-level covariates. Figure 10 presents the output of a propensity score logit regression. Propensity scores are the "predicted probability" based on the estimated model.

```
. psmatch2 participation_03 famsize agehead sexhead Income_HH_per, logit

Logistic regression                               Number of obs   =      97305
                                                  LR chi2(4)      =    2572.82
                                                  Prob > chi2     =     0.0000
Log likelihood = -58937.123                       Pseudo R2       =     0.0214

------------------------------------------------------------------------------
participation_03 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
        famsize |    .110427   .0029879    36.96   0.000     .1045708    .1162832
        agehead |   -.0103875   .0004896   -21.21   0.000    -.0113472   -.0094279
        sexhead |    .3647205   .0246022    14.82   0.000     .3165011    .4129399
  Income_HH_per |   -4.58e-07   3.33e-07    -1.37   0.169    -1.11e-06    1.95e-07
          _cons |    .2866613   .0386452     7.42   0.000     .2109181    .3624045
------------------------------------------------------------------------------
There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.
```

Figure 10. Logit regression to estimate the propensity scores

- ✓ To evaluate the success of our matches, we use pstest to assess the improvement in baseline balance after matching as: `pstest age sex famsize agehead sexhead Income_HH_per, t(participation_03) graph both`. Figure 11 presents the partial output from the command. We find that the matched sample has smaller difference between treatment and control group covariates. Interestingly, using household-level propensity scores even balanced the individual-level age variable, though it led to an unbalancing of individual-level sex.

```
. pstest age sex famsize agehead sexhead Income_HH_per, t(participation_03) graph both
```

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
age	U	24.174	27.654	-16.6		-26.52	0.000
	M	24.297	24.533	-1.1	93.2	-2.18	0.029
sex	U	.50013	.49798	0.4		0.63	0.526
	M	.50343	.51286	-1.9	-338.2	-3.40	0.001
famsize	U	6.4333	5.6675	30.9		48.83	0.000
	M	6.4856	6.3932	3.7	87.9	6.92	0.000
agehead	U	46.246	48.717	-17.0		-26.54	0.000
	M	46.316	45.755	3.9	77.3	7.72	0.000
sexhead	U	.93251	.88506	16.5		25.25	0.000
	M	.9331	.94408	-3.8	76.9	-8.38	0.000
Income_HH_per	U	698.5	938.16	-1.2		-2.01	0.045
	M	704.37	406.24	1.5	-24.4	4.78	0.000

Figure 11. Output of pstest command to assess the improved balance after PSM

Figure 12 is the graphical output from the `pstest` command, which demonstrates that selection bias (in terms of measured and tested covariates) is reduced by matching.

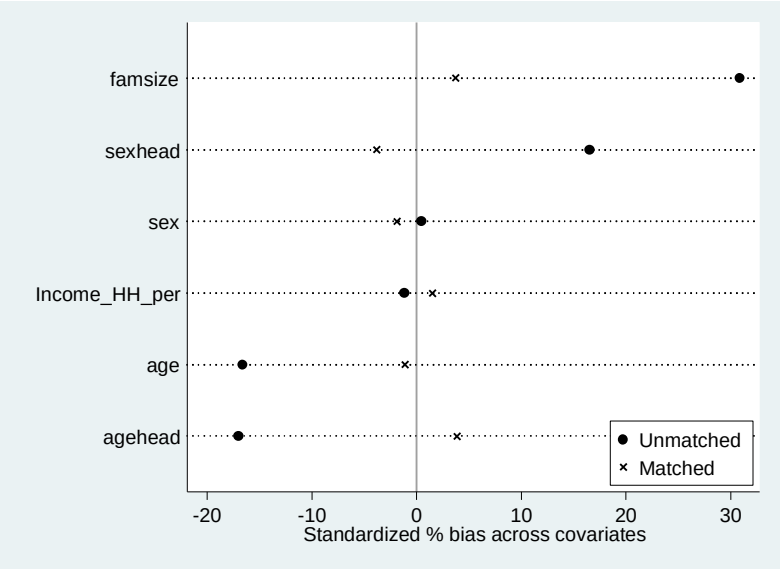

Figure 12. Graph of reduced bias in covariates after matching

- ✓ We can graphically demonstrate the "common support" condition by using command `psgraph` as shown in Figure 13. We can also compare the density distributions using command: `twoway (kdensity _pscore if participation_03==1, clwid(medium)) (kdensity _pscore if participation_03==0, clwid(thin) clcolor(black)), xti("") yti("") title("") legend(order(1 "p-score treatment" 2 "p-score control")) xlabel(0.3(.2)1) graphregion(color(white)).` The graphical output is shown in Figure 14.

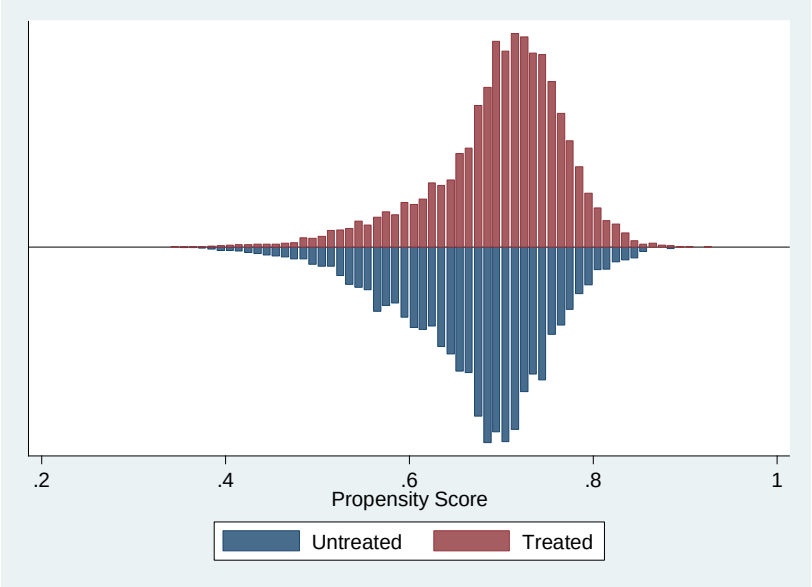

Figure 13. Histogram of propensity score in treatment and control groups

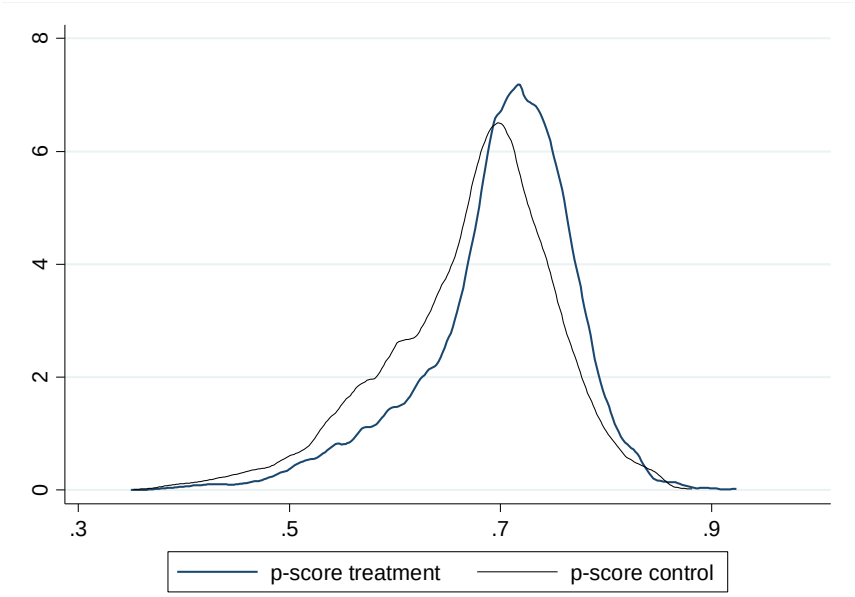

Figure 14. Kernel distribution of propensity scores demonstrating common support

- ✓ In summary, we are able to reduce baseline bias using the PSM technique. We also find common support for the distribution of propensity scores. In reality, however, propensity-score matching does not always reduce bias. When sample size is large, for instance, matching by using random selection for treatments and controls from two separate groups is as good as matching. Also, you may find that several observations can be "off-support", with propensity scores far higher or lower than any scores held by members of the other group; these individuals may (in some cases) be dropped from further analysis.

4.2 Evaluating the Impact of the Intervention

For further analysis, we could restrict the sample only to those households which have common support. The command pstest creates a variable called 'support' which can be used to drop observations we don't need: `keep if support == 1`. Additional STATA code is given below:

- ✓ We need to keep from year 2000 only those observations which have common support. We will also save the variables for predicted propensity score (`_pscore`) and the created variable `participation_03`. Then we merge this "matching" related information with original dataset:

```
keep if _support==1 & uniqhh == 1
keep hogid2 _pscore participation_03
sort hogid2
merge hogid2 using "$path/Panel_OPORTUNIDADES_00_07_year.dta"
```

- ✓ We again create the variables `period` (1 if 2003 and 0 if 2000) and `DHH_period` (`period * D_HH`) for use in DID analysis.
- ✓ Estimate the ATE using DID analysis as before, and store the results so that we can later compare them with the results using only the matched sample:

```
reg edu_child DHH_period D_HH period, vce(robust)
estimates store r1
```

- ✓ Estimate the DID effect, but this time restricting the analysis only to the matched sample and store the results.
  ```
  reg edu_child DHH_period D_HH period if matched==1, vce(robust)
  estimates store r2
  ```
- ✓ Using the command xml_tab, export the results of the two regression above to an Excel file as: `xml_tab r1 r2, below stats(N r2) replace save("$path\DID_Results.xml")`. Figure 15 presents a comparison of DID results from these two models. The first column reports the result for the DID using the full sample; the coefficient associated to the interaction is 0.64, significant at 1% significance level. When we use matched sample, this coefficient is 0.43, also significant at 1% significance level (indeed, the 95 percent confidence intervals have no overlap). Our results are therefore sensitive to probability score matching at the baseline, perhaps providing evidence of baseline selection bias which was reduced by matching.

		DID with Original Sample	DID with PSM Sample
DHH_period	Coef	0.642***	0.431***
	SE of Coef	(0.042)	(0.060)
D_HH	*Coef*	-0.276***	-0.269***
	SE of Coef	(0.033)	(0.035)
period	*Coef*	-0.042	0.225***
	SE of Coef	(0.035)	(0.054)
_cons	*Coef*	4.067***	4.102***
	SE of Coef	(0.029)	(0.031)
N		82,321	57,788
R2		0.009	0.012
note: *** p<0.01, ** p<0.05, * p<0.1			

Figure 15. Comparing DID with and without PSM

5. TRIPLE DIFFERENCE IN DIFFERENCES

Triple-Differences Design (DDD) is an extension to the basic DID analysis covered above to multiple groups and for multiple time periods. For example, let's compare three groups. We know that in OPORTUNIDADES only a fraction of the households in the treatment villages are eligible for treatment, so we can compare these three groups: (1) control villages and households; (2) participating households from treatment villages; and (3) non-participating households from treatment villages. While we agree that there are concerns about selection bias across these groups, here we only wish to demonstrate an application of DDD.

Consider the following conceptual model.

(10) The outcome Y_{igst} for an individual i at time t in group g (treatment or control) and subgroup (s) denoting participating in program itself:

$$E[Y_{igt}|G,S,t] = \alpha_g + \pi_s + \theta_t + \beta_1 G + \beta_2 S + \beta_3 t + \beta_4 GS + \beta_5 Gt + \beta_6 St + \beta_7 GSt$$

Where α_g represents group-level time-invariant "fixed" effects; π_s represents time-invariant fixed effects between participating and non-participating households; θ_t represents period fixed effects; G is indicator variable for the treatment (=1) or control (=0) groups; S is indicator variable for the participation (=1) group; t is indicator variable for baseline (=0) or endline/follow-up (=1) measurements, and the βs are the regression coefficients to be estimated. We ignore time-variant effects and assume that the expected mean error in above model is 0. We'll spare you the math, but it can be shown that the average treatment effect of the intervention on the participating households from the treatment villages compared to the control households from control villages is given by coefficient β_7.

6. BIBLIOGRAPHY/FURTHER READING

1. Bertrand, Marianne; Esther Duflo and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?" Quarterly Journal of Economics, 119, 249-275.
2. Donald, S. G. and K. Lang (2007). "Inference with Difference-in-Differences and Other Panel Data", Review of Economics and Statistics, 89, 221-233.
3. Gerber, Alan S., and Donald P. Green. Field experiments: Design, analysis, and interpretation. WW Norton, 2012.
4. McKinnish, T. (2000). "Model Sensitivity in Panel Data Analysis: Some Caveats About the Interpretation of Fixed Effects and Differences Estimators", Mimeo, University of Colorado.