Replication

Randomized Field Experiments. The Difference-In-Differences Estimator

Replication: *'Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger'* (Aker, Ksoll, and Lybbert, 2012)

The paper is uploaded on Canvas.

In this project you will replicate the main results in the paper *'Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger'* (Aker, Ksoll, and Lybbert, 2012). Aker et al. examined the effect of mobile phone usage on adult learning by designing a randomized experiment of an adult education program (the so-called Project Alphabetisation de Base par Cellulaire, or ABC).

Working on this project will be helpful in the future as it will show you the main steps a researched goes through when they carry a research on their own.

Data

There are 3 datasets we are going to use:

Dataset 1: The ABCtestscore.dta

This dataset includes the test score data from all five rounds of writing and math tests, including: January 2009; June 2009; January 2010; June 2010; and January 2011. All of these are identified as rounds 1, 2, 3, 4 and 5, respectively. The definitions of the variables are described in the dataset.

Dataset 2: ABCteacher.dta

This dataset includes information on teachers' gender, age, resident village and highest educational attainment in terms of years of education. Teacher data was collected from each village in each year. The definitions of the variables are described in the dataset.

Dataset 3: ABChousehold.dta

This dataset includes data on household and individual (so called respondent-level) characteristics from household surveys conducted in January 2009 (before the program), January 2010 and January 2011. The definitions of the variables are described in the dataset.

(All of the above discussion is based on the paper. Refer pages 1 through 13 in the paper for more detail.)

1. Preliminary questions

(i) Mechanism of the effect and literature review

Read *section 1. Introduction* on the paper (pages 1 through 7)

Briefly (in about one paragraph) discuss through what channels mobile phone use might affect the educational outcomes of adults. Is the effect of mobile phone usage on learning outcomes unambiguous, i.e. can we say before we have studied it whether it will be beneficial or detrimental?

Briefly review the relevant literature

Check online some of the relevant literature, i.e. what other authors have found when studying this topic. To do so follow the steps below:

- Go to: https://scholar.google.com/
- Type some key words characterizing the paper , e.g. "effect of mobile phone on learning"
- You will see a number of papers published on the topic. Click on "All ** versions to open a paper". Read at the minimum the abstract and introduction to see what the authors did. **Include a few sentences on one paper you found of your choice** – who are the authors, what research question they studied, and what results they obtained. You need to *paraphrase* (i.e. do not copy-paste – this is regarded as plagiarism).

(ii) Experimental design

Read *section 2. Project ABC* of the paper (pages 7 through 9)

Briefly (in about one paragraph) describe the experimental design. Make sure to answer the following questions:

- Where was the ABC project implemented (in which country, how many regions and villages took part in the experiment?
- How many villages were assigned to the treatment group (the so-called ABC group) and how many were assigned to the control group (the so-called non-ABC group)?
- What was the treatment, i.e. how the educational courses and tools differed between the two groups – treatment and control?
- What test scores were collected? When were the tests administered?

2. Replication

(i) First consider the following model:

$$\textit{\textbf{Test score}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}\textit{\textbf{Mobile phone}} + \textit{\textbf{controls}} + \boldsymbol{u},$$

where *Mobile phone* is a dummy variable for whether the individual uses a mobile phone.
What are the potential threats to validity of the OLS estimate of β_1 in this model in the absence of a random experiment?

(ii) In a random experiment setting one big threat to validity is the so-called *"contamination of the randomization"*, i.e. randomization may not have been done properly. Let's check if there is such an issue in the ABC experiment.

In each part of this question, your paper must summarize the results, rather than comment on each single test result.

- Use dataset **ABChousehold.dta** to answer this question.

I would advise you to use a DO file in all parts of the question. You can use the sample DO file I have uploaded on CANVAS.

Use STATA command *des* to see the description of variables *age*, *hhhead*, *assets*, *cellphoneowner* and *usecellphone* (these are all household and respondent characteristics).

Check the estimated differences in means for the two groups in 2009 – before the program started, and test the hypothesis that the mean in the ABC group for each characteristic equals the mean in non-ABC group at the conventional 5% level, i.e. test

$$\boldsymbol{H_0}: \boldsymbol{\mu_{\mathrm{ABC}} = \mu_{\mathrm{non\text{-}ABC}}}$$

against the two-sided alternative.

Do your results suggest that the differences in pre-program household characteristics of the two groups are small and insignificant (as should be if randomization was correctly done)?

Hint: Use STATA command *ttest varname if year==2009, by (abc)*

- Use dataset **ABCteacher.dta** to answer this question.

Comment on the sample means for both groups – do they seem close to each other? At the 5% level do you see any significant difference between the average age of the teachers (*teacherage)*, the fraction of female teachers (*femaleteacher)*, the fraction of local teachers (*local)* and the average education level of the teachers (*levelno*) in the treatment and control villages in the baseline period (year 2009)?
(Use command *des* to see the variable description and the same t-test command as in the previous part).

- Use dataset **ABCtestscore.dta** to answer this question.

Lastly, compare the average math (*mathzscore*) and writing (*writezscore*) z-test scores in the ABC and non-ABC villages. Note that both test scores are standardized to have a mean zero and a standard deviation of one (such standardized scores are often referred to as z-scores). Test the hypothesis for equality of means at the 5% significance level. What do you conclude?

Summarize the STATA output in the table on pg. 4: report the sample mean for each variable for the control and the treatment villages, report the estimated difference in sample means, and the p-value of the two-sided alternative. This is how the sample statistics are usually reported in publications in the economics and related fields. Refer to Table 1 and 2 in the paper for illustration.

Table 1: Household and teacher characteristics in the treatment and control regions

Characteristic	ABC villages Sample mean	Non-ABC villages Sample mean	Estimated difference in means (p-value of the two-sided alternative)
Panel A: Household			
Age of respondent	37.18	37.86	0.69 (0.37)
Respondent is a household head			
Number of asset categories owned			
Respondent owns a mobile phone			
Respondent has used a mobile phone			
Panel B: Teacher			
Education (number of years)			
Age			
Gender (female = 1)			
Local (teacher from village = 1)			
Panel C: Test score			
Math test score			
Writing test score			

Use the following legend next to each estimated difference in means and hypothesis test, and include the note below:

Notes: * significant at the 10 percent level; ** significant at the 5 percent level; *** significant at the 1 percent level.

```
. ttest age if year==2009, by (abc)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     519    37.86127    .5748581    13.09617    36.73193    38.99061
       1 |     519    37.17534    .5161162    11.75794     36.1614    38.18928
---------+--------------------------------------------------------------------
combined |   1,038     37.5183     .386237    12.44378    36.76041     38.2762
---------+--------------------------------------------------------------------
    diff |            .6859345    .7725528               -.8300121    2.201881
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   0.8879
Ho: diff = 0                                     degrees of freedom =     1036

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.8126         Pr(|T| > |t|) = 0.3748          Pr(T > t) = 0.1874
```

(iii) The econometric strategy the authors of the paper employ is the Difference-In-Differences estimator (DID), which is a commonly used strategy for studying policy interventions.

Recall from class that formally the Difference-In-Differences estimate can be obtained by estimating a regression model with an interaction term as follows:

$$\textit{\textbf{Test score}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}\textit{\textbf{ABC}} + \boldsymbol{\beta_2}\textit{\textbf{Post}} + \boldsymbol{\beta_3}\,\textit{\textbf{(ABC x Post)}} + \textit{\textbf{u}},$$

where $Test\ score$ the learning outcome of interest (math or writing test score); *ABC* = 1 for the treatment villages, and 0 otherwise; *Post* = 1 if after the intervention, and 0 otherwise.

(This is a simplified specification of the model in Section 4 of the pager, on pg. 13.)

- In terms of the parameters of the model, what is the difference-in-difference estimator of the effect of the mobile phone usage? Show formally.

- What is the key assumption for validity of the DID results? Do you this it holds in the context of our model?

- The authors only use the data from rounds 1, 2 and 4 (dropping all data from rounds 3 and 5).

To replicate their results and before doing any estimation let's restrict the sample to the one used in the paper by using the dataset **ABCtestscore.dta**, and typing in STATA:

keep if round==1|round==2|round==4

(This only keeps the observations from rounds 1, 2, 4 dropping all observations from round 3 &5. *Note:* Recall from ECON503 that "|" stands for "OR" in STATA syntax).

***NOTE:* In estimating all specifications use heteroskedasticity-robust standard errors and covariance, i.e. use option "*robust*".**

- ***Model specification 1***

Estimate the model above using dataset **ABCtestscore.dta**, i.e. regress math/write zscore on variables *abc*, *post* and *abcpost* (this is the interaction between the two).

What is the DID estimate? How do you interpret it? Is it statistically significant? Is it economically significant? What do you conclude?

- ***Model specification 2***

Now include controls for the person's **age and gender** (variables *age* and *female*), as well as regional dummies accounting for potential **differences in test scores across region of residence** (the dummy variables for the 4 regions are *hausa zarma kanuri dosso* – if you include all STATA will automatically drop one since they are perfectly collinear).

Is the model robust to including controls, i.e. does the DID estimate change a lot compared to the previous specification?

- ***Model specification 3***

The authors include age in the model specified linearly. However, it is likely the relationship between age and test scores is non-linear. Generate a quadratic term of age (call it *agesq*) and include it in the model for *mathzscores* and *wrtiezscores*.

Are *age* and *agesq* jointly significant at the 1% level in each model (i.e. perform an F-test)? Based on this, would you suggest the authors keep *agesq* in the model?

- ***Model specification 4***

Finally execute STATA command:

qui tab codevillage, gen(village_dum)

This creates 113 village dummies – one for each of the values of variable *codevillage* (village code), i.e. it creates a dummy variable for each village in the sample. The dummies will be named *village_dum1* through *village_dum113*.

Include *all* village dummies in the model by executing the command:

reg mathzscore abc post abcpost age agesq female village_dum, robust*
reg writezscore abc post abcpost age agesq female village_dum, robust*

(STATA will drop one of the dummies automatically).

This controls for **differences across villages** (the so-called *village heterogeneity* or commonly referred to as *village fixed effects*).

Is the model robust to including controls for differences across villages, i.e. do your conclusions change regarding the effect of mobile phone learning on math and writing test scores?

Summarize the estimation results by filling in the table on the next page (if you prefer you can create 2 tables – one for write test scores, and one for math test scores).

<u>You need to report every parameter estimate, together with its (robust) standard error in parentheses as shown in the example.</u> Report the number of observations and the R-squared of each model.

Also, <u>use the following notation for significance level:</u> ***Significant at the 1 percent level, ** significant at the 5% level, * significant at the 1% level (as shown in the example). This is how the estimation results are usually reported in publications in the economics and related fields. Refer to Table 3 in the paper for illustration.

Table 2—Impact of the ABC program on average test scores: Difference in Difference

	(1)	**(2)**	**(3)**	**(4)**
Panel A: Math z-scores				
ABC	-0.071*** (0.025)			
Post	0.000 (0.024)			
ABC × Post	0.246*** (0.034)			
Age	---			
Age^2	---	**---**		
Region dummies	No	Yes	Yes	No
Village dummies	No	No	No	Yes
Number of observations	13,420			
R^2	0.009			
Panel B: Writing z-scores				
ABC				
Post				
ABC × Post				
Age	---			
Age^2	---	**---**		
Region dummies	No	Yes	Yes	No
Village dummies	No	No	No	Yes
Number of observations				
R^2				

Notes: Panel A represents the results with math z-scores as the dependent variable. Panel B represents the results with writing z-scores as the dependent variable.
Heteroskedasticity robust standard errors reported in parentheses.
***Significant at the 1 percent level, ** significant at the 5% level, * significant at the 1% level.

```
. reg mathzscore abc post abcpost, robust

Linear regression                               Number of obs     =     13,420
                                                F(3, 13416)       =      43.33
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0093
                                                Root MSE          =      .9653

------------------------------------------------------------------------------
             |               Robust
  mathzscore |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         abc |  -.0712108   .0236441    -3.01   0.003    -.1175566    -.024865
        post |  -2.63e-09   .0248476    -0.00   1.000    -.0487049    .0487049
     abcpost |   .2460714   .0332319     7.40   0.000     .1809322    .3112107
       _cons |  -6.69e-09   .0184518    -0.00   1.000    -.0361681    .0361681
------------------------------------------------------------------------------
```

IMPORTANT NOTE:

You can create the complete table very quickly with STATA command *outreg2*. However, you fist need to install this command. To do so, type in STATA's command window:

findit outreg2

You will see several links. Scroll down to

outreg2 from http://fmwww.bc.edu/RePEc/bocode/o

then click on the link. This will install outreg2. I am not certain this would work on a UM computer (as you don't have admin rights to install on those), so my guess is you need to work on a personal computer.

Once you have outreg2 installed, the command to created a Word file with a table of results is:

reg y x1 x2 x3
outreg2 using FILE_NAME.doc, replace ctitle(TITLE HERE)
reg y x1 x2 x3 x4 x5
outreg2 using FILE_NAME.doc, append ctitle(TITLE HERE) addtext(COMMENT)

This will create a table with all 2 columns – one for each specification, including the parameter estimates and their standard errors. as well as the R-squared and number of observations in each specification. Note that the table will not get over-written, i.e. if you have created the table, each time you want to re-run this code you need to delete the table first.

For instance, to obtain the first and second specification for write score, you would type:

reg writezscore abc post abcpost
outreg2 using WRITE.doc, replace ctitle()
reg writezscore abc post abcpost age female hausa zarma kanuri dosso
outreg2 using WRITE.doc, append ctitle() addtext(contorls, YES)