Assignment 1:
Microeconometrics

Daniel Redel (210630)
Instructor: Dr. Alexandros Theloudis

Autumn 2022

1 Randomization

a. The paper studied two randomized experiments conducted in schools in urban India. The first intervention was the **Balsakhi Program**, which is a *remedial education program* targeted at children at the bottom of the grade distribution. This program took place in municipal schools in Mumbai and Vadodara and provided them with a woman teacher (a "balsakhi") from the local community to work with the children lagging behind in basic literacy and math skills in the third and fourth grades. These children are taken out of the regular classroom to work with the teacher in a group of other 15-20 children for two hours per day. Instruction focuses on core competencies the children should have learned in the first and second grades.

The second intervention was the **Computed-Assisted Learning (CAL) Program** focusing on math, where all 4-grade children in treated schools in Vadodara were offered two hours of shared computer time per week (two children per computer) during which they solve math problems through games, according to their current level of achievement. One of these hours was during class time and the other was either immediately before or immediately after school. Although the children were occasionally helped by instructors in terms of getting a better understanding of the task required, at no time did any instructors provide general instructions in mathematics.

b. The main outcome the authors were interested in was, broadly, whether the interventions resulted in any improvement in learning levels. In this case, this was measured using normalized annual pretests and post-tests scores. Let's define $y_{igj,POST}$ the post-test score[1] of child i in grade g and school j and the treatment status D_{jg} of the child's school j and grade g, which is a dummy variable equal to 1 if school j is received a balsakhi (or the CAL

[1] The paper actually uses the change in student's test scores ($y_{igj,POST} - y_{igj,PRE}$) as the main outcome. For simplicity, we are only using post scores, since the pre-tests scores were not affected by the interventions.

program), and 0 otherwise. We can now write the **potential outcomes** as:

$$y_{igj,POST} = \begin{cases} y^1_{igj,POST} & \text{if } D_{jg} = 1 \\ y^0_{igj,POST} & \text{if } D_{jg} = 0 \end{cases}$$

Where $y^1_{igj,POST}$ is the test score of child i have him been in a school with a balsakhi teacher (or CAL program), and $y^0_{igj,POST}$ is the test score of child i have him been in a school without a balsakhi teacher (CAL program). Since a particular individual child i cannot simultaneously be part of the treatment and the control group, we only observed one outcome:

$$y_{igj,POST} = y^0_{igj,POST} + D_{jg}(y^1_{igj,POST} - y^0_{igj,POST})$$

The unobserved outcome is called the **counterfactual outcome**.

c. In general terms, Banerjee et al. (2007) studied whether children improved more -relative to what would have been expected- in treatment schools than in comparison schools through **Randomized Control Trials** methodology (social experiments). In both cities, assignment to a treatment group was random and conducted at a **school level** from a sample of government primary schools. To ensure a balanced sample, the assignment was **stratified** by language, pretest score, and gender[2]. For the Balsakhi intervention, half the schools (Group A) received a balsakhi for grade 3 (and grade 4 served as a control group for Group B), and Group B was given balsakhis for grade 4 (grade 3 served as a control group for Group A). This design also allows the authors to estimate long-run effects, since a child entering grade 3 in a school from Group A would remain treated in the second year when in grade 4.

As said previously, the outcome of interest is the difference in the change in student's normalized test scores between treatment and control groups $(y_{igj,POST} - y_{igj,PRE})$[3]. The authors used OLS regressions to identify **Average Treatment Effects** for schools that received the corresponding interventions.

However, to account for the fact that not all schools in Mumbai assigned to a balsakhi actually received them (year 2), Banerjee et al. used the **Intention to Treat** dummy variable D_{jg} as an **instrument** for the actual treatment status P_{jg}. Note that this ended up capturing a **Local Average Treatment Effects** (LATE) of the compliers groups (schools that participated in the program only because they were offered, otherwise they would have not participated).

d. According to the evidence presented in the paper, the randomization appears to have been successful. To support this, the authors compared the pre-test scores of the treatment and control groups. The results in Columns (1)–(3) of Table II didn't show statistically significant differences between the groups, and the point estimates were also small (less than

[2] Schools in Vadodara received both the Balsakhi and CAL intervention. To keep it balanced, the CAL's sample of schools was also stratified according to treatment group status in the Balsakhi Program

[3] The normalization is done by subtracting the mean pretest score of the control group and dividing by the standard deviation of the control group's pretest score.

a 10^{th} of an std. dev.).

One potential challenge was due to the fact that children in grade 3 in schools i that received the program for grade 4 form the control group for children that receive the program for grade 3 in school j, and vice versa. The program effects would be biased if schools reassigned resources, for example, from the treatment grades to the control grades in response to the program. The authors, however, confirmed that there was no reallocation of resources, making sure that the students in the control group were a reasonable counterfactual.

Another potential problem that could affect internal validity is attendance and dropout rates, which we shall discuss in the next question.

e. Differential attrition (due to dropouts) between treated and non-treated schools could potentially bias the results if, for example, weak children were less likely to drop out when they benefited from a balsakhi (negative bias). To minimize attrition, the implementation team visited the schools multiple times and tracked down at home those who did not attend school for the tests. To test for this potential bias, the authors compared attrition rates in treatment and control schools. They showed that attrition levels were low and similar in both groups and each city, suggesting the absence of selective attrition.

f. The authors identify both **direct effects** on the children who were assigned to work with the balsakhi and **indirect effects** on children who stayed behind in the classroom due to *class-size effects* and *peer effects.* The direct effect is due to the fact that the program targets children in at the bottom of the grade distribution. Indirect effects, on the other side, may be related to the fact that the children who didn't attend balsakhi classes now benefit from having a small size classroom (Krueger & Whitmore, 2001). Non-treated students may also benefit from peer effects given that now the composition of that class is more homogeneous (Angrist, 2014).

To disentangle these effects, the authors used **Instrumental Variables** methodology (2SLS), using the predicted probability of a child being assigned to the balsakhi in treatment schools P_{igj} as an instrument of actual assignment. In the first stage, the authors predict a child's assignment as a flexible function of his or her pre-test score. The predicted values are then used in the second stage to identify direct effects (τ) and indirect effects (γ)[4].

g. External validity is the extent to which the results of a social experiment can be generalizable to the general population of interest. In the paper, the authors addressed this by implementing the Balsakhi experiment in two different cities: Mumbai and Vadodara, which are important in India. The results showed that the program was effective in both cities. In that sense, there may be external validity regarding the Indian population, but this could not necessarily be the case for other developing countries such as Latin America, for example. Although they may share similar socioeconomic backgrounds, other variables

[4] The assumption here is that the indirect treatment effect γ does not vary with the child's score in the initial test score distribution.

such as culture or political background could influence how generalizable are the results in Banerjee et al. (2007).

2 Linear Regression

The OLS estimate of our parameters of interest $\beta = (\beta_0, \beta_1)$ is given by:

$$\hat{\beta}_1 = \frac{n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

In matrix notation, this can also be expressed as:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

To show why these two expressions are equivalent, we can decompose the matrix equation as follows:

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

Now we plug all the components:

$$\hat{\beta} = (X'X)^{-1}X'Y = \frac{1}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

Note that:

$$\begin{aligned} n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 &= n\sum_{i=1}^{n} x_i - (n\bar{x})^2 \\ &= n\left(\sum_{i=1}^{n} x_i - n\bar{x}^2\right) \\ &= n\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{aligned}$$

We now can solve:

$$
\begin{aligned}
\hat{\beta} &= \frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix}(\sum_{i=1}^{n}x_i^2)(\sum_{i=1}^{n}y_i)-(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}x_iy_i)\\ -(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)+(n)(\sum_{i=1}^{n}x_iy_i)\end{bmatrix}\\
&= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix}(\sum_{i=1}^{n}x_i^2)(1/n)(\sum_{i=1}^{n}y_i)-(1/n)(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}x_iy_i)\\ -(1/n)(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)+(n/n)(\sum_{i=1}^{n}x_iy_i)\end{bmatrix}\\
&= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix}(\sum_{i=1}^{n}x_i^2)(1/n\sum_{i=1}^{n}y_i)-(1/n\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}x_iy_i)\\ -(1/n\sum_{i=1}^{n}x_i)(n\bar{y}_i)+(\sum_{i=1}^{n}x_iy_i)\end{bmatrix}\\
&= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix}\bar{y}(\sum_{i=1}^{n}x_i^2)-\bar{x}(\sum_{i=1}^{n}x_iy_i)\\ -n\bar{x}\bar{y}+\sum_{i=1}^{n}x_iy_i\end{bmatrix}\\
&= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix}\bar{y}(\sum_{i=1}^{n}x_i^2)-\bar{x}(\sum_{i=1}^{n}x_iy_i)-n\bar{y}\bar{x}\bar{x}+n\bar{y}\bar{x}\bar{x}\\ \sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}\end{bmatrix}\\
&= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix}\bar{y}(\sum_{i=1}^{n}x_i^2-n\bar{x}^2)-\bar{x}(\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y})\\ \sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}\end{bmatrix}\\
&= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix}\bar{y}\sum_{i=1}^{n}(x_i-\bar{x})^2-\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\\ \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\end{bmatrix}
\end{aligned}
$$

Finally, we get:

$$
\begin{aligned}
\hat{\beta} &= \begin{bmatrix}\bar{y}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}-\bar{x}\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\\ \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\end{bmatrix}\\
\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} &= \begin{bmatrix}\bar{y}-\bar{x}\hat{\beta}_1\\ \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\end{bmatrix}
\end{aligned}
$$

3 Omitted Variables

Let's assume the following linear regression:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

But we also know that x_{2i} is not observed in the data, so that $v_i = \beta_2 x_{2i} + u_i$. In this case, the OLS estimator for the coefficient β_1 is defined by:

$$\hat{\beta}_1 = \frac{Cov(x_{1i}, y_i)}{V(x_{1i})}$$

We can rearrange this equation in order to understand the conditions needed for getting an unbiased estimator of β_1:

$$\begin{aligned}\hat{\beta}_1 &= \frac{Cov(x_{1i}, y_i)}{V(x_{1i})} = \frac{Cov(x_{1i}, \beta_0 + \beta_1 x_{1i} + v_i)}{V(x_{1i})} \\ &= \frac{Cov(x_{1i}, \beta_0)}{V(x_{1i})} + \frac{Cov(x_{1i}, \beta_1 x_{1i})}{V(x_{1i})} + \frac{Cov(x_{1i}, v_i)}{V(x_{1i})}\end{aligned}$$

The first term goes to 0 because β_0 is a constant[5]. We also know that $Cov(X_i, X_i) = V(X_i)$. With that, we end up having:

$$\begin{aligned}\hat{\beta}_1 &= \beta_1 \times \frac{V(x_{1i})}{V(x_{1i})} + \frac{Cov(x_{1i}, v_i)}{V(x_{1i})} \\ &= \beta_1 + \frac{Cov(x_{1i}, v_i)}{V(x_{1i})} \\ &= \beta_1 + \frac{Cov(x_{1i}, \beta_2 x_{2i} + u_i)}{V(x_{1i})} \\ &= \beta_1 + \frac{Cov(x_{1i}, \beta_2 x_{2i})}{V(x_{1i})} + \frac{Cov(x_{1i}, u_i)}{V(x_{1i})}\end{aligned}$$

Because of strict exogeneity $E(u_i|x_{1i}, x_{2i}) = 0$, we can see that:

$$\begin{aligned}Cov(x_{1i}, u_i) &= E(x_{1i}u_i) - E(x_{1i})E(u_i), \text{ by definition of covariance} \\ &= E(x_{1i}u_i), \text{ since } E(u_i) = 0 \text{ by assumption} \\ &= 0, \text{ by the orthogonality condition}\end{aligned}$$

by which we end up with:

$$\begin{aligned}\hat{\beta}_1 &= \beta_1 + \frac{Cov(x_{1i}, \beta_2 x_{2i})}{V(x_{1i})} \\ &= \beta_1 + \beta_2 \times \frac{Cov(x_{1i}, x_{2i})}{V(x_{1i})}\end{aligned}$$

[5] recall property $Cov(X_i, A) = 0$

This implies that in order to get an unbiased OLS estimator of β_1, the right term in the last equation needs to be 0. Intuitively, we need to assume that there is no correlation between x_{1i} and x_{2i} (that is, $Cov(x_{1i}, x_{2i}) = 0$). In practice, however, this assumption is not likely to hold.

To solve this problem, we can rely on Instrumental Variables (IV) methodology. The main idea behind IV regressions is that we can decompose the variation in x_{1i} in two parts: (i) a **non-problematic component** the we can actually explain using exogenous variables such as z_i and (ii) a **problematic component** which is correlated with our unobserved x_{2i}. By isolating the first component, it is possible to recapture an estimate of β_1.

By choosing an instrument z_i, so that:

1. Relevance: $Cov(z_i, x_{1i}) \neq 0$
2. Exclusion: $Cov(z_i, \beta_2 x_{2i} + u_i) = 0 \rightarrow$ The instrument is uncorrelated with both x_{2i} and u_i.

we can solve this in two steps. Using the **first stage** equation we can estimate the variation in x_{1i} that is exogenous (that is, the variation explained by our instrument z_i):

$$\hat{x}_{1i} = \beta_f + \gamma z_i$$

In the second step, we use the predicted $\hat{x}_{1i}$ to show that:

$$\begin{aligned}
\hat{\beta}_1 &= \beta_1 + \beta_2 \times \frac{Cov(\hat{x}_{1i}, x_{2i})}{V(\hat{x}_{1i})} \\
&= \beta_1 + \beta_2 \times \frac{Cov(\beta_f + \gamma z_i, x_{2i})}{V(\beta_f + \gamma z_i)} \\
&= \beta_1 + \beta_2 \times \frac{\gamma Cov(z_i, x_{2i})}{\gamma^2 V(z_i)}
\end{aligned}$$

If the exclusion assumption holds, the right term of the equation is 0 ($Cov(z_i, x_{2i}) = 0$) and we end up with an unbiased estimator of β_1.

However, in order for this to be true, We also need to impose and additional assumption, which is *homogeneous treatment*, otherwise our estimator $\hat{\beta}_1$ would be the Average Treatment Effect of a particular sub-population (LATE on the complier group).