Assignment 2

VSLA Program and School Absenteeism

The issue of school absenteeism is today the subject of intense scrutiny, all around the world. School absenteeism has proven to have long-term effects: given that higher quantity of education increases returns to education between 6.8% to 10.6% per year of schooling (Duflo, 2001), the subsequent loss in education may lead to lower wages.

Using data from Mozambique, this paper studies the effect of the Village Saving and Loan program (VSLAs), a government program helping to cover the everyday costs of schooling in order to decrease school absenteeism. The goal of this study is to determine whether the program has a causal effect on school attendance, studying two different methods of selection into the program: via poverty index or voluntary participation.

Potential Outcomes Framework:

The potential outcomes framework is a model that helps quantify causal effects (VanderWeele, 2016), due to the issue of not being able to observe two outcomes at once. The framework enables a "numeric quantity" (VanderWeele, 2016) to define the causal relationship.

Here, Ti = 1 if individual (i) took part in the treatment, i.e., the VSLA program, and is equal to 0 otherwise. Y= 1 if there is a change in the outcome, i.e. if the number of days missed decreased, and equals 0 otherwise. $Y_i(1)$ represents the potential outcome if T=1 and $Y_i(0)$ represents the potential outcome if T=0.

i	T	Y	Y(1)	Y(0)	Y(1) - Y(0)
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	1	?	1	?

Y(1) - Y(0) would represent a causal effect of the treatment on the individual. Here, the causal impact of the program on absenteeism is driven by the reduced economic costs of attending school. Reducing the everyday costs of school, the program alleviates financial incentives to nonattendance. There were two methods of selection: The first selection system, based on the household poverty index; those with a poverty index <58 are automatically selected. The second selection is based on choice and promotion; households choose to participate in the program or not - with randomised promotion in villages. With promotion impacting enrolment thereby affecting days of school missed.

Regression Discontinuity Design:

Households below the cut-off; a poverty index below 58, were automatically inducted into the VSLA program, whereas households above the cut-off were not. We observed 2,689 households in the program and 1455 not in the program. Figure 1 provides graphical evidence of the effects of enrolling in the VSLA program. The large discontinuity around the cut-off provides motivation for the RD design. The mean school days missed for households in the program was: 6.90, whereas for households not in the program: 16.35. The test of no difference in the means is heavily rejected with a t-stat of 52.78.

The results of the regression: $days_i = \alpha + \beta(poverty\_index)_i + \tau(enrolled)_i + \varepsilon_i$ in Table 1 Column 1, suggest that enrolment in the program caused 7.76 school days fewer missed in the last 30 days.

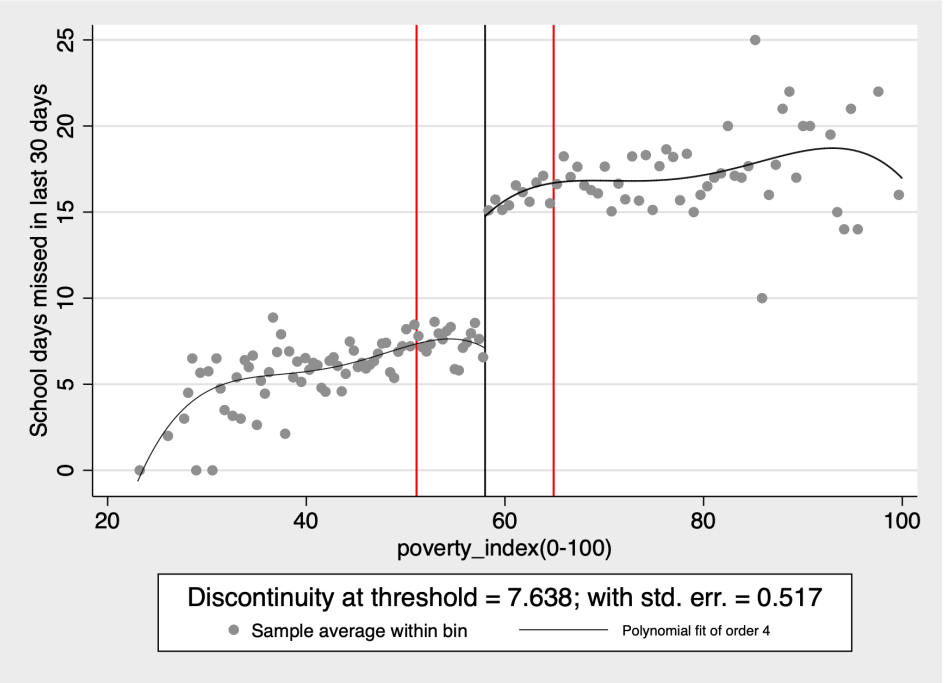

Figure 1: The effect of enrolment in the VSLA program

Notes: vertical lines at cutoff 58 (black), vertical lines at bandwidth cutoffs 51.101 & 64.899 (red lines)

To eliminate misspecification of the functional form, the regression seen in Table 1 Column 2 was run, the coefficient of the enrolment variable decreased by 4.6 percentage points. This is a negligible difference to our initial RD estimate suggesting our model is robust to changes in the functional form of the running variable.

The key assumption when undergoing RDD is that households are similar around the cutoff. We tested this assumption through a balancing table of households within our bandwidth. A bandwidth of 6.89 was selected using the mean square error, this was the optimal bandwidth to minimise the trade-off between bias and variance.(see Table 3). Most of the characteristics were similar, yet we observed noticeable differences in the ages of the head of household and spouses (being on avg. 7 years younger in the treatment group). This may reflect a positive relationship between poverty index and age. Despite this we believe the evidence presented suggests that this assumption holds on average.

Instrumental Variable Approach:

An IV approach is used when the relationship between the treatment, and the outcome cannot be proven to be causal as there may be endogeneity. To establish the causal effect of the program on absenteeism, we thus use an instrument. For the IV to estimate the causal impact three conditions must be fulfilled: relevance condition; exogeneity; and no direct effect on the outcome.

With an IV approach we satisfy the relevance condition with the following regression:

$$HH\ Enrolled\ in\ VSLA = a + \beta(HH\ in\ Promoted\ Community)$$

We find HH located in Promoted Community to be a strong IV for the Enrolment variable. This is because it explains 42.9% of the variation in enrolment (R-squared: 0.2392), and its β coefficient is 0.466*** (s.e.: 0.009, t: 51.070). This means that households in the promoted community are 0.466 times more likely to enrol in the VSLA than villages not promoted in. The villages selected for the promotion campaigns were selected randomly, satisfying exogeneity. Moreover, being part of the promotion campaign for the VSLA does not directly affect school absenteeism but via enrolment.

Using 2SLS we find fitted values for enrolment in the VSLA which are then put into the regression for days of school missed in the past 30 days. Running the column 2 regression in Table 2, we observe R-squared of 0.429, β coefficient for enrolment is -8.422 (s.e.: 0.239) and significant at the 1% level (z: -35.280). Without an IV approach the effect of choosing to enrol is -9.245793 and is significant (t: -79.42) which signifies an overestimation of the effect of enrollment when we do not use the IV approach.

Conclusion:

In summary, the analysis established that the VSLA program had a causal impact on school absenteeism. This was justified using the Regression discontinuity design and the instrumental variable approach under which both models' assumptions were held. The RDD highlighted a significant decline in the number of days missed by 7.764 and the IV approach depicted a 8.421 days decrease, when enrolled in the VSLA. Lastly, the two selection methods used in entering the VSLA alleviated the problem of counterfactuals as addressed in the potential outcome framework analysis.

Table of Regressions Dataset 1: Selection into VSLA if poverty index below 58

Dependent Variable	(1) Days	(2) Days	(3) Days	(4) Days
Poverty Index 1-100	0.103***	0.197***	0.206***	-0.018
	(0.013)	(0.061)	(0.061)	(0.058)
Enrolled in VSLA	-7.764***	-7.810***	-7.785***	-7.474***
	(0.279)	(0.280)	(0.280)	(0.264)
Poverty Index Squared		-0.001	-0.001*	0.000
		(0.001)	(0.001)	(0.001)
HH has Dirt Floor			-0.634***	-1.342***
			(0.183)	(0.175)
Number of HH Members				-0.962***
				(0.040)
Constant	9.604***	7.085***	7.440***	20.623***
	(0.875)	(1.810)	(1.810)	(1.788)
Observations	4324	4324	4324	4324
R-squared	0.400	0.401	0.402	0.472

Standard errors in parentheses
* p<0.1, **p<0.05, ***p<0.01

Notes:
1. HH = HouseHold

Table of Regressions Dataset 2: Household selection into VSLA by choice

Dependent Variable	(1)Days	(2)Days 2SLS	(3)Days
HH Enrolled in VSLA	-9.246***		-9.047***
	(0.116)		(0.135)
HH Enrolled in VSLA Instrumented by HH in Promoted Community IV		-8.421***	
		(0.239)	
Age of HH			0.007
			(0.009)
Age of Spouse			0.019*
			(0.010)
Education of HH			0.018
			(0.031)
Education of Spouse			-0.016
			(0.032)
HH is Female			0.329
			(0.273)
HH Speaks an Indigenous Language			-1.336***
			(0.148)
Number of HH Members			-1.005***
			(0.036)
HH has Dirtfloor			-1.256***
			(0.158)
HH has Bathroom			0.178
			(0.139)
Distance to Closest Hospital			-0.004**
			(0.002)
HH Member Visited Hospital in the Past Year			-0.095
			(0.306)
Constant	16.147***	15.862***	17.724***
	(0.068)	(0.099)	(0.596)
Observations	8298	8298	8298
R-squared	0.432	0.429	0.508

Notes: Standard errors in the parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Days = number of school days missed in the past 30 days. HH is head of household.

Table 3: Balancing of covariates of households within the bandwidth who enrolled in the program (treatment) and those who did not (control)

EXPERIMENTAL CONDITIONS	*(1) CONTROL*	*(2) TREATMENT*	(3) DIFFERENCE
days	15.84798 [4.72257]	7.454406 [6.117932]	-8.393574 [0.234881]
poverty index (0-100)	61.14703 [1.860979]	54.49637 [1.988374]	-6.65066 [0.084515]
age of HH (years)	51.55582 [15.21092]	44.31494 [14.59472]	-7.24088 [0.661824]
age of spouse (years)	44.36698 [12.8584]	39.19004 [12.64865]	-5.17694 [0.564766]
education of HH	2.54064 [2.691265]	2.987018 [2.678485]	0.446378 [0.118742]
education of spouse	2.311758 [2.463023	2.698084 [2.523516]	0.386326 [0.10993]
dirtfloor (binary)	.5866983 [.4927187]	0.6842912 [0.4649757]	0.0975929 [0.021307]
bathroom (binary)	.6615202 [.4734738]	.6022989 [.4896107]	-0.0592213 [0.021212]
hospital distance	105.0696 [40.92888]	111.6142 [40.38037]	6.5446 [1.799722]
hospital visits	0 [0]	0.0490421 [0.2160388]	0.0490421 [0.00598]
female	.1021378 [.3030094]	0.0835249 [0.2767801]	-0.0186129 [0.012952]
indigenous	.3325416 [.4714039]	0.435249 [0.4959797]	0.1027074 [0.02127]
Household size	4.846793 [2.129917]	5.370115 [2.000812]	0.523322 [0.091953]
Observations	842	1305	2147

Notes:

1.In Columns (1) and (2), results are presented as mean [standard deviation]. Column (3) is presented as mean [standard error].

2. No significant differences in observable characteristics, between control and treatment group.

3.HH = Head of Household

4.Column (1) = households with a poverty index > 58 & < 64.899, Column (2) = households with poverty index > 51.101 & < 58

References:

- Brady Neal - Causal Inference (2020) *2.2 - The Fundamental Problem of Causal Inference*. Available at: https://www.youtube.com/watch?v=RGvI0uVMgtw&ab_channel=BradyNeal-CausalInference
- Chang, H. N. & Romero, M. (2008) Present, Engaged, and Accounted for: The Critical Importance of Addressing Chronic Absence in the Early Grades. Report. National Center for Children in Poverty
- Duflo, Esther (2001), 'Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,' American Economic Review, Vol. 91 (4), pp 795-813
- Mod U: Powerful Concepts in Social Science (2016) *The Logic of Instrumental Variables: Causal Inference Bootcamp*. 02/03/2016. Available at: https://www.youtube.com/watch?v=4xF_DMbL14w&ab_channel=Mod%E2%80%A2U%3APowerfulConceptsinSocialScience. (Accessed:28/03/2021)
- Mod U: Powerful Concepts in Social Science (2015) *Instrumental Variables in Action: Education and Wages (graphs): Causal Inference Bootcamp*. 02/09/2015. Available at: https://www.youtube.com/watch?v=vacBsxBgFMY&ab_channel=Mod%E2%80%A2U%3APowerfulConceptsinSocialScience. (Accessed: 28/03/2021).
- VanderWeele, T. J. (2016, Dec). Commentary: On Causes, Causal Inference, and Potential Outcomes. *International Journal of Epidemiology, 45*(6), 1809-1816.