Advanced Development Microeconomics:

Replication of Duflo, E., Dupas, P., & Kremer, M. (2015)

Saisawat Samutpradit
29186018

August 5, 2019

1 Description of the Replicated Analysis

I replicated 5 tables from Duflo et al. [2015]: Table 1 (Baseline characteristics, by treatment group), Table 2A Panel A (Short-run impacts: roll call data), Table 3 Panel A (Long-run impacts: individual long-term follow-up survey data), Table A2 (Attrition in roll call data), and Table A3 (Survey rates during long-run follow-up). The main results are reported in Table 2 and 3; attrition in Table 4 and 5. It is found that the replicated coefficients and standard errors are consistent with the published version.

However, there is one minor remark: the label for each treatment is different from the published version of the paper. In the published version, education subsidy treatment is abbreviated as “S”, while in the program, it is instead denoted as “U”. In this analysis report, I follow the program’s notation.

There are two issues I would like to address by replicating these tables: sample selection bias and unbalanced baseline characteristics. This report is organized as follows: first I address the attrition issue; next, I incorporate religion into the balance check.

It is found that the results are somewhat robust and baseline characteristics are balanced across treatment arms.

2 Attrition

There seems to be some issues with sample selection, especially for girls. This is prevalent in short- and long-run even after the intensive tracking, but not in the medium-run. Table 4 reports the linear probability estimation of having each dependent variable missing. It is found that in the short-run (after 3 years) follow-ups, some of the missing outcomes appear to be unbalanced across treatment arms, as reported in the p-values in the lower panel of Table 4. As a result, whether dependent variable is observed is correlated with the treatment variable. This violates the exogeneity assumption. On the other hand, in the medium-run follow-up, attrition appears to be balanced across treatment arms.

Seeing this results after 5 years, in the long-run follow-up, the authors may have tried to mitigate this issue by incorporating intensive tracking (IT) in addition to the regular tracking (RT). Nevertheless, in the long-run survey rates reported in Table 5, attrition is still an issue even after intensive tracking. According to column 4, those in the education subsidy and joint program are more likely to be found. To mitigate this issue, in the analysis for the long-run impacts, the authors have incorporated sampling weight to retrieve some observations lost for those in the HIV education program. The results are shown in Table 3. However, some other methods such as Heckman’s two-step model, or Lee’s bound estimator have yet to be attempted.

As the results are imprecise, it is questionable whether they would be robust to other estimation methods. The original short- and long-run results are reported in Table 2 and 3. It is found that most of the coefficients are not statistically significant. In this analysis report, I estimated with Heckman's two-step model and Lee's bound estimator and compared the results.

2.1 Short-run

I first estimated Heckman's two-step model. In the short- and medium-run follow-ups, since attrition appears strong for "Dropout" (whether the student drops out of primary school) and "Presence" (whether the student is presence on survey date) during the 3-year follow-up, we only considered these dependent variables.

The results of this estimation are reported in Table 6. When compared with Table 2, it is found that when impose selection equation, the estimates are similar in magnitude. The effects on "Presence" remains insignificant as found in Table 2. However, in the case of "Dropout" as dependent variable, significance is lost.

In the first step of Heckman's model, additional to the treatment and other control variables, I also included "distance to bus stop" as an instrument that possibly satisfies relevancy and exclusion restriction. If the student's house is closer to the bus stop (i.e. community), it may be easier for her to be found. However, when test for the first-stage relevancy, F-statistics is small (1.91) and insignificant. Exclusion restriction also may not be satisfied due to its correlation with the cost of travel to school. Moreover, it should also be noted that this instrument is inappropriate because it is collected at endline survey (after 7 years). Since there seems to be no appropriate instrument collected at baseline, I further considered Lee's bound estimator using a Stata package by Tauchmann [2012].

`leebounds` command only allows for binary treatment variable and discrete covariates. Consequently, controls we used in Table 2, such as size of the school, are not suitable here. To make the results comparable to OLS, I first used OLS to estimate the impact each treatment to the control group without including any covariates. The results are reported in Table 7. Note that the impact of education subsidy on dropout is statistically significant and similar in magnitude to that reported in Table 2. The difference is that the joint program starts to have impact on dropout as well. The significance is at 5 percent level, however.

The results from `leebounds` are reported in Table 8. When compared with OLS estimates without covariates in Table 7, it is found that the effect of education subsidy on drop-out may be slightly stronger if estimated by Lee's bound estimator. Like the simple OLS estimation, the impact of HIV education is negative but statistically insignificant; and joint program is negative and slightly significant. Note that as `leebounds` command only support binary treatment variable, each column represents effects compared to the pure control group.

2.2 Long-run

As discussed above, since we lack an appropriate instrument, Heckman's two-step estimator may not be ideal. Therefore, in the analysis for long-run impact, I considered only Lee's bound estimator. Similar to the short-run analysis, I also estimated the impacts on education attainment [1] using Lee's bound estimator (Table 10) and compared the results with simple OLS estimates (Table 9).

Similar to the results in Table 3, simple OLS yields significant impact of education subsidy on "Reach G8"(student reaching grade 8) and "Grades completed". Additionally, joint program

[1]The reason for choosing "Reach G8" and "Grades completed" is that they are the only outcomes with significance level at 10 percent in Table 3.

also becomes statistically significant when regress on "Grades completed". It is merely at negligible 10 percent, however.

Next, I compare the results of simple OLS with bound estimator. According to Table 10, the impact of uniform subsidy on grades completed still falls in the positive range when estimated with Lee's bounded estimator. On the other hand, the effect of education subsidy on "grades completed" may be zero, conflicting with both Table 3 and 9.

These results imply that the original OLS results in Table 2 and 3 are somewhat robust to selection bias in both short run and long run.

3 Alternative Mechanism: Religion

The original balance check shows that the baseline characteristics of schools and study cohorts are well-balanced between treatment groups. However, it has yet taken into account of another important institution that is how religious each school or students in this study cohort are. It may be the case that students in the some school are more religious than others. For instance, some schools may be organized by religious organizations. Therefore, it shall be more convincing if a proxy for religiosity such as church attendance of students is also included in the baseline balance check.

Here I replicated Table 1 in Duflo et al. [2015] and extended the analysis to factors concerning religion. I used anonymous in-class survey self-administered by students in grades 7 and 8 in 2005, `KAPgirls` and `KAPboys`. Note that this survey is neither collected at baseline nor individual is identifiable. Consequently, only the measure we obtained here has some time lags and is only available at school-level. I linked this information to the main data set. The additional balance check is illustrated in Table 1.

First, I consider the composition of self-reported religion[2] at each school. "Protestant", "catholic", "muslim", and "hindu" refer to the proportion of students who reported having the corresponding religion at each school. About 40 percent of the students are protestants, 30 percent catholic, 6–7 percent muslim[3]. None of the p-values are statistically significant. Thus, we can conclude that the composition of students is well-balanced among schools.

Next, I consider church attendance[4] of the students. I defined three measures of religiosity. "Religiosity" takes the value 1 if the student is more than 75 percent certain that he/she will go to church on the following Sunday; "Religiosity2" takes the value 1 if he/she is 100 percent certain; and "Religiosity3" takes the value 1 if he/she will go to church with probability greater than 0. Across the three scores, none of the p-values are statistically significant. Therefore, there is no school in which students are systematically more religious.

Although data used in this additional balance check is limited, we can conclude that there is no systematic difference among schools in the aspect of religion composition and religiosity.

[2]This is taken from question 12 of the KAP survey: "What is your religion/denomination?"

[3]The rest of the students identified themselves as "others" but reported the name of their churches. To determine the denomination of these churches is data-demanding. Therefore, we omit them from the analysis.

[4]This is taken from question 52 of the survey: "What do you think is the percent chance (what are the chances out of 100) that you will go to church this Sunday?"

Table 1: Extension of Table 1 Panel A – Balance check at school level

var	mean_U	sd_U	mean_H	sd_H	mean_UH	sd_UH	mean_control	sd_control	$U = C$	$H = C$	$UH = C$	$U = UH$	$H = UH$	N
kcpe2003	255.2	29.59	249.4	24.83	248.6	32.04	249.3	26.12	0.0970	0.660	0.468	0.146	0.857	319
schsize	464.6	203.1	489.3	208.8	473.8	185.7	498.9	194.3	0.292	0.777	0.587	0.764	0.611	328
ratio02	1.016	0.124	1.024	0.127	1.012	0.105	1.016	0.135	0.945	0.455	0.823	0.857	0.556	328
latrine_2004	11.63	6.324	11.22	6.420	9.949	5.702	11.08	5.538	0.215	0.635	0.0810	0.0680	0.173	317
urban	0.157	0.366	0.120	0.328	0.0870	0.284	0.109	0.313	0.180	0.879	0.364	0.170	0.512	328
total_2km	2.012	1.954	2.157	1.818	2.063	1.803	2.055	1.761	0.845	0.521	0.925	0.860	0.742	328
TOTteachers	14.18	4.246	14.63	5.291	13.82	4.367	14.58	4.667	0.786	0.439	0.282	0.613	0.257	327
meanage	39.96	3.097	39.63	3.780	39.65	3.795	39.62	3.531	0.455	0.796	0.845	0.570	0.975	321
sexratio_teachers	1.216	1	1.177	0.848	1.297	0.987	1.147	0.829	0.856	0.788	0.281	0.575	0.405	327
protestant	0.407	0.205	0.394	0.202	0.446	0.188	0.411	0.211	0.570	0.204	0.168	0.231	0.105	328
catholic	0.296	0.160	0.318	0.183	0.303	0.170	0.312	0.180	0.538	0.443	0.863	0.791	0.569	328
muslim	0.0670	0.102	0.0700	0.109	0.0630	0.0970	0.0690	0.106	0.989	0.761	0.693	0.814	0.669	328
hindu	0.00500	0.00700	0.00600	0.00700	0.00500	0.00600	0.00600	0.00700	0.296	0.348	0.477	0.848	0.313	328
religiosity	0.571	0.103	0.570	0.118	0.580	0.101	0.574	0.111	0.723	0.594	0.656	0.624	0.550	328
religiosity2	0.484	0.115	0.483	0.130	0.491	0.112	0.483	0.127	0.891	0.813	0.614	0.693	0.649	328
religiosity3	0.925	0.0470	0.921	0.0500	0.925	0.0530	0.920	0.0450	0.536	0.715	0.554	0.994	0.557	328

Table 2: Replication of Table 2A Panel A – Impacts on girls after 3 years

	(1) Dropout	(2) Presence	(3) married	(4) pregnant	(5) preg unmar.	(6) mar unpreg.
Uonly	-0.031**	-0.002	-0.026**	-0.027**	-0.004	-0.002
	(0.012)	(0.006)	(0.010)	(0.011)	(0.006)	(0.003)
Honly	0.003	-0.008	0.011	-0.007	-0.014**	0.005*
	(0.011)	(0.006)	(0.009)	(0.011)	(0.006)	(0.003)
UH	-0.016	0.000	-0.000	-0.011	-0.013**	-0.001
	(0.012)	(0.006)	(0.009)	(0.010)	(0.006)	(0.003)
Observations	9,116	8,232	9,107	9,072	9,072	9,072
Control mean	0.188	0.939	0.128	0.160	0.046	0.011
U=UH	0.240	0.716	0.011	0.145	0.090	0.809
H=UH	0.102	0.204	0.235	0.742	0.866	0.058
U=H	0.005	0.328	0.000	0.084	0.092	0.021
UH=U+H	0.474	0.237	0.278	0.133	0.539	0.311

Standard errors in parentheses, clustered by school. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 3: Replication of Table 3 – Long-run impacts

	(1) Reach G8	(2) Grades comp	(3) Married	(4) Pregnant	(5) Preg Unmar	(6) Mar unpreg	(7) Child bf 16
Uonly	0.039**	0.209**	-0.039*	-0.032	0.006	-0.001	-0.020*
	(0.017)	(0.092)	(0.021)	(0.021)	(0.012)	(0.004)	(0.012)
Honly	-0.004	-0.022	0.020	0.017	-0.007	-0.004	-0.009
	(0.017)	(0.089)	(0.020)	(0.022)	(0.014)	(0.005)	(0.014)
UH	-0.009	0.006	-0.012	-0.008	0.006	0.001	-0.005
	(0.018)	(0.097)	(0.021)	(0.022)	(0.013)	(0.006)	(0.012)
Sampling weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,687	5,687	5,715	5,719	5,715	5,715	5,719
U=UH	0.008	0.036	0.218	0.275	0.995	0.621	0.176
H=UH	0.794	0.772	0.144	0.253	0.332	0.276	0.777
U=H	0.018	0.016	0.007	0.027	0.328	0.448	0.362
UH=U+H	0.068	0.150	0.821	0.821	0.708	0.313	0.171

Standard errors in parentheses, no clustering. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 4: Replication of Table A2 Panel A – Attrition in Roll Call Data

	After 3 years				After 5 years		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dropout	Presence	Married	Pregnant	Dropout	Married	Pregnant
Uonly	0.001	-0.002	-0.002	-0.005	-0.014	-0.001	-0.002
	(0.005)	(0.009)	(0.005)	(0.005)	(0.010)	(0.014)	(0.014)
Honly	-0.003	0.013	-0.001	-0.002	-0.019*	-0.021	-0.024
	(0.006)	(0.009)	(0.006)	(0.006)	(0.010)	(0.015)	(0.015)
UH	0.010	0.015*	0.010	0.007	-0.005	-0.008	-0.006
	(0.006)	(0.008)	(0.007)	(0.007)	(0.012)	(0.018)	(0.018)
Observations	9,482	9,482	9,482	9,482	9,482	9,482	9,482
Control attrition	0.037	0.131	0.038	0.044	0.076	0.123	0.132
U=UH	0.115	0.037	0.063	0.061	0.331	0.642	0.789
H=UH	0.041	0.816	0.136	0.235	0.167	0.418	0.263
U=H	0.422	0.109	0.860	0.575	0.485	0.082	0.052

Standard errors in parentheses, clustered by school. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 5: Replication of Table A3 Panel A column 1–4 – Attrition for girls during Long-run Follow-up

	(1)	(2)	(3)	(4)
	Dead	Found RT	Found IT	Found
Uonly	-0.001	0.087***	-0.044	0.060***
	(0.004)	(0.016)	(0.029)	(0.013)
Honly	0.001	0.044**	0.008	0.021
	(0.003)	(0.017)	(0.030)	(0.014)
UH	-0.004	0.090***	0.041	0.073***
	(0.004)	(0.017)	(0.031)	(0.015)
Observations	9,482	9,354	1,291	9,354
Control attrition	0.013	0.444	0.783	0.565
U=UH	0.401	0.824	0.006	0.369
H=UH	0.124	0.006	0.328	0.001
U=H	0.548	0.009	0.094	0.006

Standard errors in parentheses, clustered by school. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 6: Two-step Heckman's model

	Dropout		Presence	
	(1)	(2)	(3)	(4)
main				
Uonly	-0.032	-0.038	-0.002	-0.008
	(0.020)	(0.023)	(0.005)	(0.009)
Honly	-0.002	-0.002	-0.007	-0.005
	(0.020)	(0.023)	(0.005)	(0.009)
UH	-0.005	-0.025	0.001	0.004
	(0.021)	(0.025)	(0.005)	(0.009)
select				
Uonly	0.013	-0.224	0.018	0.147*
	(0.078)	(0.188)	(0.062)	(0.079)
Honly	0.059	-0.045	-0.116*	0.055
	(0.079)	(0.201)	(0.061)	(0.080)
UH	-0.124	-0.384**	-0.137**	-0.016
	(0.077)	(0.181)	(0.062)	(0.078)
distance		0.002		-0.000
		(0.002)		(0.001)
Observations	9,482	5,722	9,482	5,722
Instrument		Yes		Yes

Standard errors in parentheses, clustered by school. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 7: OLS estimates without any covariates

	Dropout			Presence		
	(1)	(2)	(3)	(4)	(5)	(6)
Uonly	-0.032***			-0.002		
	(0.011)			(0.005)		
Honly		-0.009			-0.005	
		(0.011)			(0.005)	
UH			-0.024**			0.003
			(0.011)			(0.005)
Observations	4,694	4,652	4,486	4,247	4,191	4,052

Standard errors in parentheses, no clustering. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 8: Lee's bound estimator

	Dropout			Presence		
	(1)	(2)	(3)	(4)	(5)	(6)
	Uonly	Honly	UH	Uonly	Honly	UH
main						
lower	-0.035***	-0.012	-0.026**	-0.002	-0.005	-0.003
	(0.012)	(0.012)	(0.011)	(0.005)	(0.012)	(0.012)
upper	-0.032***	-0.008	-0.018	0.005	-0.005	0.004
	(0.011)	(0.011)	(0.012)	(0.011)	(0.005)	(0.005)
Observations	4,876	4,829	4,683	4,876	4,829	4,683
Trimming proportion	0.003	0.004	0.007	0.007	0.000	0.006

Standard errors in parentheses, no clustering. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 9: Long-run impacts with simple OLS

	Reach G8			Grades completed		
	(1)	(2)	(3)	(4)	(5)	(6)
Uonly	0.026*			0.162**		
	(0.014)			(0.069)		
Honly		-0.003			-0.018	
		(0.015)			(0.070)	
UH			0.014			0.120*
			(0.015)			(0.071)
Sampling weights	No	No	No	No	No	No
Observations	2,876	2,758	2,783	2,876	2,758	2,783

Standard errors in parentheses, no clustering. $*p < 0.1, **p < 0.05, ***p < 0.01$

Table 10: Bounded long-run impacts

	Reach G8			Grades completed		
	(1)	(2)	(3)	(4)	(5)	(6)
	Uonly	Honly	UH	Uonly	Honly	UH
main						
lower	0.004	-0.015	-0.013	-0.179*	-0.192*	-0.285***
	(0.016)	(0.017)	(0.017)	(0.102)	(0.104)	(0.106)
upper	0.133***	0.047*	0.138***	0.491***	0.153*	0.505***
	(0.027)	(0.026)	(0.028)	(0.097)	(0.092)	(0.100)
Sampling weights	No	No	No	No	No	No
Observations	4,876	4,829	4,683	4,876	4,829	4,683
Trimming proportion	0.114	0.058	0.131	0.114	0.058	0.131

Standard errors in parentheses, no clustering. $*p < 0.1, **p < 0.05, ***p < 0.01$

References

E. Duflo, P. Dupas, and M. Kremer. Education, HIV, and Early Fertility: Experimental Evidence from Kenya. *American Economic Review*, 105(9):2757–97, 2015.

H. Tauchmann. leebounds: Lee's (2009) treatment effects bounds for non-random sample selection for Stata. 2012.