Instructions for Recreating Tables

The figures and tables in the paper and appendix can all be reproduced using the data set "Public_Data_AEJApp_2010-0132.dta". Along with the data set, we have included a set of fourteen programs all of which run on the included data set. The program "Figure_1-4_Followup.do" recreates the four figures. The remaining programs recreate each of the individual tables and the names of the individual programs indicate the respective table. For example, "Table_02_Registrant_Comparison.do" recreates Table 2 in the paper. And "Table_A1_Survey_Selected_Comparison.do" recreates Table A1 in the appendix. All programs and data sets are in STATA Version 9.2 format.

Data Documentation

The programs use eight different types of data. What follows is a brief overview of the data sources and the different variables. For a complete overview, we refer the reader to Section III.C of the paper.

First, part of our data comes from the original SISBEN surveys of 2003 and 2004, which contain information on all the families that were eligible to register for the lottery. These data are subsets of the actual surveys that were used to create the SISBEN national poverty index. The second source of data comes from the registration forms that each family filled out during application and the outcome of the resulting lottery. Third, we obtained administrative records from the Secretary of Education that includes the enrollment records of every child in a public school and most private schools in the two localities.

After having collected this information on the 251 schools across which students were spread, we chose to collect baseline data and subsequent attendance data in only the 68 schools with the largest number of registered children. For this specific group of students, we collected attendance data during the last quarter of 2005 through direct in-school observations in the form of roll calls. Finally, we conducted both a baseline (school based) and follow-up survey (home based) on these students.

The rest of this Data Documentation provides an overview of the variables that are included from each data source and a brief description of each variable.

1. Dataset ID's

While information from the registration, random assignment, and SISBEN data sets were available for all students, the other data sets only cover a fraction of the students. These variables indicate for which students the various data are available. Since only one variable was taken from the administrative enrollment data set ("m_enrolled"), no indicator is provided for whether or not we were able to match students to that data base. The variable is simply set to missing if a match was not possible.

Variable Name	Description	Values
bl_observed	Indicator for whether or not a student completed the baseline survey.	1= student surveyed in baseline 0=not
fu_attrit	Indicator for whether or not a student failed to complete the follow-up survey.	1=student attrited 0= student did not attrit
fu_observed	Indicator set to one if follow-up survey was provided.	1=students surveyed at follow-up 0=not
survey_selected	Student is enrolled in one of the 68 schools selected for surveying and verified attendance.	1= student was selected for surveying & attendance 0 = not

2. SISBEN data

As noted, a wide range of our demographic controls are taken from the SISBEN dataset which covers information from the original SISBEN surveys. The SISBEN data provides us with several variables at the family level such as schooling level of the household head, physical characteristics of the dwelling, employment status of adults, and family income. It also provides individual level variables such as enrollment status at the time of the survey, age, income, and marriage status. These variables are included as controls in nearly all of the analyses performed in our study.

Variable Name	Description	Values
s_age_sorteo	Age	between 1 and 85
s_age_sorteo2	Age squared	between 1 and 7225
s_durables	Index of durable goods	between 0 and 4
s_edadhead	Age of Head of Household	between 19 and 98
s_estcivil	Marital Status of head of household.	1= free union 2= married 3= widow(er) 4= divorced 5= single
s_estrato	Estrato classification	0, 1, 2, or 3
s_infraest_hh	Physical infrastructure index of house.	between 3 and 21
s_ingtotal	Household Income	between 0 and 4000
s_num18	Number of kids 18 and under	between 0 and 12
s_over_age	Indicator for whether or not a child is older than normal for their grade.	1= 0=

s_puntaje	SISBEN Score	between 1.45 and 21.99
s_sexo	Gender	hombre or 0
s_single	Single Parent Household	1= single parent household 0= not
s_teneviv	House Possession	1= rent 2= own paying it 3= own paid 4= other condition
s_tpersona	Number of People in Household	between 1 and 22
s_utilities	Utilities	between 0 and 6
s_years_back	Number of years older a child is than the normal children in his or her grade.	between -16 and 70
s_yrs	Years of Education	between 0 and 16
s_yrshead	Years of Education Head of Household	between 0 and 22

3. Registration Data

The registration information stems from the eligibility forms that families had to provide during their child's registration for the actual program. The information that was collected, including data on birth date, gender, last grade completed and year in which that grade was completed, was verified with the actual SISBEN database and, when possible, with the SED's official records.

Variable Name	Description	Values
grade	Students grade at time of registration	between 6 and 11
r_be_gene	gender	F or M
school_code	School Level ID	1 to 278 1 designates students who attend a private school
suba	indicates data from Suba	1= Suba 0= San Cristobal

4. Variables from random assignment

The random assignment process assigned students to treatment and control groups. The following variables indicate the treatment or control status of each student in various ways.

Variable Name	Description	Values
control	Student not assigned to receive any of the stipends	1=control 0=treatment
T1	Indicator for whether or not student is part of the basic experiment. Includes students in control and those assigned to basic treatment from San Cristobal.	1=student in Basic Treatment experiment 0= not
T1_treat	Student received the basic treatment.	1=student assigned to basic treatment 0=otherwise
T2	Indicator for whether or not student is part of the savings experiment. Includes students in control and those assigned to savings treatment from San Cristobal.	1=student in Savings Treatment experiment 0= not
T2_treat	Students assigned to the savings treatment	1=student in T2 treatment group 0=otherwise
T3	Indicator for whether or not student is part of the tertiary experiment. Includes students in control and those assigned to tertiary treatment from Suba.	1=student in Tertiary Treatment experiment 0= not
T3_control	Students in Suba assigned to the control group.	1= student in T3 control group 0= otherwise
T3_treat	Students in Suba assigned to the tertiary treatment.	1=student in T3 treatment group 0=otherwise
Treatment	Students assigned to any of the treatments in either Suba or San Cristobal.	1= treatment 0= not

4. Participation Data: Administrative enrollment & Verified Attendance

The enrollment data stems from the SED administrative dataset. This variable indicates whether a student whether or not a student was enrolled assuming that we had sufficient information to match the student to the SED's data base. Attendance data was collected during the last quarter of 2005 through direct observation. A group of assistants randomly visited schools and classes, where they performed roll-calls to mark students' attendance. They visited a total of 1,069 classes in the 68 selected schools for 13 weeks.

Variable Name	Description	Values
m_enrolled	Indicator for whether o r not a student is enrolled based on administrative enrollment data.	1=enrolled 0=not enrolled Missing = insufficient information to match
at_msamean	Percentage of days absent using verified attendance measure.	between 0 and 1

5. Follow-up Data

These data were collected during the follow-up survey round in February and March 2006.

Variable Name	Description	Values
fu_currently_attending	Indicator for whether or not a student is enrolled based on follow-up survey.	1=yes 0=no
fu_earn_last_wwk	Earnings Last Work Week	between 0 and 408
fu_hrs_last_wwk	Hours Worked Last Work Week	between 0 and 91
fu_nim_hogar	Household ID number	between 1 and 6844
fu_primary_home	Primary Activity, Home	1= 0=
fu_primary_study	Primary Activity, Study	1= 0=
fu_primary_work	Primary Activity, Work	1= 0=
fu_self_attendance	Fraction of days students reported to attend on the follow-up survey.	between 0 and 1
graduated	Is the student reported to have graduated based on the follow-up survey?	1= graduated 0=not
tertiary	Indicator for whether or not students report attending post-secondary school in follow-up survey.	1=Student attend post-secondary 0=not

6. Predicted participation data

Predicted attendance and predicted enrolment are used to understand these rates not including the treatment. These counterfactuals are estimated by modeling the behavior of students in our control groups using only the available baseline demographic characteristics. We then use the treated students' baseline characteristics and the coefficients from the regressions on the control group to project what these students would have done had they not been treated.

Variable Name	Description	Values
at_baseline	Predicted baseline attendance	between 0.2637148 and 1.156992
en_baseline	Predicted baseline enrollment	between -43.61449 and 1.2222

7. Variables created from existing data

The following variables are created in the program of Table 9 to assess the sibling externalities.

Variable Name	Description	Values
num_rsib	Number of children in household registered for the lottery	between 1 and 5
tsib	Whether or not the sibling is treated	1= sibling is treated 0= not
num_tsib	Number of treated children in household	between 0 and 2