Group 28 - Andrew Ma, Hiro Schmidt

Replication Study: Housing, Health, and Happiness

For our final project, we will be conducting a replication study of Matias D. Cattaneo et.al "Housing, Health, and Happiness" 2009. We will begin by introducing the paper's main research questions and conclusions along with basic data exploration and a summary statistics table of important covariates.

In a sentence, Matias D. Cattaneo and his peers study the health effects of the *Piso Firme* project, a government-subsidized "slum upgrade" involving the installation of cement floors in households near the Torreon region of Mexico. In particular, they compare outcome variables of parasitic infestations, diarrhea, anemia, height, weight, and cognitive development against the untreated Durango region across state lines. Moreover, the regression model includes 41 covariates, including both numerical and dummy variables such as parasite count, total income, e.t.c.

To ensure causality between the Piso Firme program and an increase in household health/happiness, households in the study undergo a pre-intervention check, to ensure that we are using comparable households from both neighborhoods. This is done through comparing variables such as father presence, presence of diseases, respiratory issues, intellectual scores, e.t.c. - there are no significant differences in mean values between control and treatment groups. This implies that improvements can be attributed to the Piso Firme program and the installation of a cement floor.

With respect to data cleaning, there seems to be the presence of missing/null values within all rows of both datasets; researchers seem to impute missing values with 'zeros' as many of them are dummy variables and the missing values seem missing at random; ie, imputing with zeroes is a quick and easy solution to a nuanced and complicated data-completeness issue. Finally, you may find below a table of the covariates and outcome variables of the regression model implemented by Cattaneo et al.; other variables were used in pre-intervention studies for causality purposes and not shown below.

Covariates:

TABLE 1—DESCRIPTION OF OUTCOME VARIABLES AND SAMPLE SIZES IN 2005 SURVEY

Variable	Description	Observations treatment	Observations control
Share of rooms with cement floors	Share of rooms in house with cement floors	1,362	1,393
Cement floor in kitchen	Indicator equal to one if kitchen has cement floor	1,362	1,393
Cement floor in dining room	Indicator equal to one if dining room has cement floor	1,362	1,393
Cement floor in bathroom	Indicator equal to one if bathroom has cement floor	1,362	1,393
Cement floor in bedroom	Indicator equal to one if bedroom has cement floor	1,362	1,393
Parasite count	Parasite count present in child's stool sample	1,528	1,566
Diarrhea	Indicator equal to one if the mother reports that the child had diarrhea in the last four weeks	1,930	2,105
Anemia	Indicator equal to one if the child's hemoglobin level is less than 11g/dL, adjusted for altitude	1,768	1,951
MacArthur Communicative Development Test score	MacArthur Communicative Development Test for children aged 12–30 months	291	302
Picture Peabody Vocabulary Test percentile score	Picture Peabody Vocabulary Test results in terms of the percentile of the distribution of Latin American outcomes for children aged 36–71 months	757	817
Height-for-age z-score	Child's height-for-age z-score	1,865	2,053
Weight-for-height z-score	Child's weight-for-height z-score	1,881	2,058
Satisfaction with floor quality	Indicator equal to one if respondent reports being satisfied or very satisfied with quality of house's floors	1,362	1,393
Satisfaction with house quality	Indicator equal to one if respondent reports being satisfied or very satisfied with overall quality of house	1,362	1,393
Satisfaction with quality of life	Indicator equal to one if the respondent reports being satisfied or very satisfied with their overall quality of life	1,362	1,393
Depression Scale (CES-D Scale)	Radloff (1977) index of self-reported depressive symptomatology. Range: 0 (low) to 60 (high)	1,354	1,388
Perceived Stress Scale (PSS)	Cohen, Kamarck, and Mermelstein (1983) index of self-reported perceived stress symptoms. Range: 0 (low) to 40 (high)	1,359	1,387
Respiratory diseases	Indicator equal to one if the mother reports that the child had a respiratory disease in the last four weeks	1,930	2,107
Skin diseases	Indicator equal to one if the mother reports that the child had a skin disease in the last four weeks	1,926	2,106
Other diseases	Indicator equal to one if the mother reports that the child had another disease in the last four weeks	1,930	2,106
Installation of cement floor	Indicator equal to one if the household reports having installed a cement floor since 2000	1,362	1,392
Construction of sanitation facilities	Indicator equal to one if the household reports having constructed new sanitation facilities since 2000	1,362	1,390
Restoration of sanitation facilities	Indicator equal to one if the household reports having improved sanitation facilities since 2000	1,362	1,391
Construction of ceiling	Indicator equal to one if the household reports having installed new ceilings since 2000	1,361	1,392
Restoration of walls	Indicator equal to one if the household reports having improved house walls since 2000	1,362	1,392
Any improvement	Indicator equal to one if the household reports having any of the house improvements reported above, excluding installation of cement floor, since 2000	1,362	1,393
Log of self-reported rental value of house	Logarithm of self-reported rental value of the house	1,284	1,285
Log of self-reported sale value of house	Logarithm of self-reported sale value of the house	1,239	1,223
Log total income of mothers of children 0–5 yrs	Logarithm of total income of mothers of children from 0 to 5 years of age	247	301
Log total income of fathers of children 0–5 yrs	Logarithm of total income of fathers of children from 0 to 5 years of age	1,026	1,000
Total consumption per capita	Total consumption per capita reported by the household	1,360	1,391

Outcomes:

TABLE 5—REGRESSIONS OF CHILDREN'S HEALTH MEASURES ON PROGRAM DUMMY

Dependent variable	Control group mean (std. dev.)	Model 1	Model 2	Model 3
Parasite count	0.333 (0.673)	−0.065	−0.064	−0.064
		[0.032]**	[0.031]**	[0.032]**
		−19.545	−19.345	−19.198
Diarrhea	0.142 (0.349)	−0.018	−0.020	−0.018
		[0.009]*	[0.009]**	[0.009]*
		−12.819	−13.834	−12.803
Anemia	0.426 (0.495)	−0.085	−0.081	−0.083
		[0.028]***	[0.027]***	[0.027]***
		−20.059	−18.908	−19.388
MacArthur Communicative Development Test score	13.354 (18.952)	4.031	5.652	5.557
		[1.650]**	[1.642]***	[1.641]***
		30.182	42.325	41.609
Picture Peabody Vocabulary Test percentile score	30.656 (24.864)	2.668	3.206	3.083
		[1.689]	[1.430]**	[1.410]**
		8.702	10.460	10.058
Height-for-age *z*-score	−0.605 (1.104)	0.007	−0.002	0.002
		[0.043]	[0.038]	[0.039]
		−1.161	0.279	−0.323
Weight-for-height *z*-score	0.125 (1.133)	0.002	−0.005	−0.011
		[0.034]	[0.036]	[0.037]
		1.790	−4.119	−8.727

Notes: Regressions computed using survey information (sample sizes reported in Table 1). Missing values in covariates were imputed with zero, and a corresponding dummy variable was then added to the regressions. Model 1: no controls; Model 2: age, demographic, and health-habits controls; Model 3: age, demographic, health-habits, and public social programs controls. Reported results: estimated coefficient, clustered standard error at census-block level in brackets (136 clusters), and 100 × coefficient/control mean.

*** Significantly different from 0 at 1 percent level.

** Significantly different from 0 at 5 percent level.

* Significantly different from 0 at 10 percent level.

Source: 2005 Survey.

References

1. Matias D. Cattaneo, Sebastian Galiani, Paul J. Gertler,Sebastian Martinez and RocioTitiunik ,"Housing, Health, and Happiness", *American Economic Journal: Economic Policy*, 2009