Karthik Muralidharan and Nishith Prakash

Below we document the steps required to replicate the tables in the paper, "*Cycling to School: Increasing Secondary School Enrollment for Girls in India*".

We use data from four sources in this paper. They are:

1. District Level Health Survey (DLHS), Round III (2007-08)
    a. Data labeled in STATA 13: dlhs-reg-data.dta
    b. Link: http://rchiips.org/prch-3.html
2. Bihar and Jharkhand Secondary Examination Board Data (2004-2010). This data was collected with due permission from Department of Education, Government of Bihar and Jharkhand.
    a. Data labeled in STATA 13:
        i. exam_data.dta
        ii. exam_data_parallel_trend.dta
3. Enrolment data hand collected from all schools in Bihar and bordering districts in Jharkhand (2003-2006). This data was collected with due permission from Department of Education, Government of Bihar.
    a. Data labeled in STATA 13: bh_enroll_data_reg.dta

Notes on the replication data sets:

1. All data files are in STATA version 13.
2. Do-files run in STATA version 13 unless mentioned otherwise. Some users might need to intall the outreg command: **ssc install outreg2**
3. All key variables have been labeled within each STATA data file.

Table	Data	Do files	Output
1	bh_enroll_data_reg.dta	table-1.do	TABLE-1-PANEL-A.xls; TABLE-1-PANEL-B.xls
2	dlhs-reg-data.dta	table-2.do	TABLE-2.xls
3	dlhs-reg-data.dta	table-3.do	TABLE-3.xls
4	exam_data.dta	table-4.do	TABLE-4.xls
5	dlhs-reg-data.dta	table-5.do	
6	dlhs-reg-data.dta	table-6.do	TABLE-6-Panel-A- Row-1.xls; TABLE-6-Panel-A-Row-2.xls; TABLE-6-PANEL-B-ROW-1.xls; TABLE-6-PANEL-B-ROW-2.xls; TABLE-6-PANEL-

			C-ROW-1.xls; TABLE-6-PANEL-C-ROW-2.xls
7	dlhs-reg-data.dta	table-7.do	TABLE-7-Panel-A.xls; TABLE-7-Panel-B.xls; TABLE-7-Panel-C.xls; TABLE-7-Panel-D.xls
8	dlhs-reg-data.dta	table-8.do	TABLE-8.xls
A.1	dlhs-reg-data.dta	table-A.1.do	
A.2	dlhs-reg-data.dta	table-A.2.do	TABLE-A.2.xls
A.3	dlhs-reg-data.dta	table-A.3.do	TABLE-A.3.xls
A.4	dlhs-reg-data.dta	table-A.4.do	TABLE-A.4.xls
A.5	dlhs-reg-data.dta	table-A.5.do	TABLE-A.5-PCA-ASSET.xls; TABLE-A.5-PCA-SES.xls; TABLE-A.5-SC-GENERAL.xls; TABLE-A.5-ST-GENERAL.xls; TABLE-A.5-OBC-GENERAL.xls; TABLE-A.5-MUSLIM-GENERAL.xls
A.6	exam_data_parallel_trend.dta	table-A.6.do	TABLE-A.6-Panel-A.xls; TABLE-A.6-Panel-B.xls
A.7	dlhs-reg-data.dta	table-A.7.do	TABLE-A.7-PANEL-A.xls; TABLE-A.7-PANEL-B.xls
A.8	dlhs-reg-data.dta	table-A.8.do	TABLE-A.8-PANEL-A.xls; TABLE-A.8-PANEL-B.xls
Figure 1	schooldist.dta	Figure-1.R (in R)	
Figure 2	ddd_long.dta	Figure-2.R (in R)	Figure 2
Figure A.1	dlhs-figure-A.1.dta	Figure-A.1.do	Figure A.1

Further Notes:

1. The do file "dlhs-raw-data-aej-applied.do" uses the raw data from DLHS-III HOUSEHOLD LEVEL (*DLHS3HIND.DTA*) AND VILLAGE LEVEL (*DLHS3VIND.DTA*) and produce the data "dlhs_long_wdist.dta".
2. The do file "dlhs-create-variables-aej-applied.do" uses the data "dlhs_long_wdist.dta" and creates the data "dlhs-reg-data.dta". This data set includes all the variables, including the interaction i.e. the DD, DDD, and DDDD, needed to produce the tables using DLHS-III data in this paper.
3. The do file "data-for-figure-A.1.do" produces "dlhs-figure-A.1.dta" required to produce Figure A.1
4. The do file "Figure-1.do" produces two data sets – "schooldist.dta" and "schoolage.dta" using "dlhs_raw_kids.dta". Run this do file before running the file "Figure-1.R" in R to produce Figure 1.

5. **Steps for Figure 2:**
    a. Use "data-for-fig-2.dta" and run the do file "ddd_nonparametric_point_estimates.do" and to get the data "ddd_long.dta"
    b. Run the do file "ddd_bootstrapping_long_confidence_intervals.do" and get the data "ddd_bootstrap_ci_long.dta"
        i. This do file will also merge the two data sets "ddd_long.dta" and "ddd_bootstrap_ci.dta" and call it "ddd_long.dta"
    c. Run "Figure-2.R" in R to produce Figure 2.