
Code and data package for Adukia, Asher, and Novosad, "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction"

Code files are included in code-packet.zip. The do file replication.do executes all the other code files to build the DISE dataset from scratch and to generate all results in the paper

Data files included:

- dise_village_key.dta -- links 2001 population census village identifiers (State code and PLCN) to DISE "vilcd" village identifier
- nss_round_55_individual.dta, nss_round_55_district.dta -- data from 55th round of NSS used to calculate wages and education returns in Table A1, and for treatment heterogeneity in Tables 7 and 8.
- village_vars.dta -- village-level control variables, and variables from PMGSY describing road sanction dates, completion dates, and baseline status of roads.

While the data files contain some village-level variables from the Indian population censuses, and data on road completion and sanction dates from PMGSY, we would advise researchers interested in data like these to explore the Socioeconomic High-Resolution Rural-urban Geographic dataset for India (SHRUG), available at <http://www.dartmouth.edu/~novosad/data.html>. These contain village data and matching codes from 1991-2013 which are more up to date and describe additional variable characteristics.

To obtain data from the District Information System for Education (which are not included in the data packet), please contact the National University of Educational Planning and Administration (NUEPA, the parent organization that administers the DISE). The data from NUEPA can be matched to our data through the "vilcd" variable and the data file dise_village_key.dta above.

