Datasets and Code for "The Dynamic Electoral Returns of a Large Anti-Poverty Program"

Election Data

The election data used in the analysis comes from the Election Commission of India. Data are publicly available at http://eci.nic.in. Polling-station level election data is only available in pdf documents. Those were digitized to create a dataset of election outcomes for the states with NREGS algorithm information. The data contains the names of all candidates, their party affiliation, and the number of votes received per candidate at the polling station as well as some limited candidate background information like age and broad caste category. Unfortunately, information on the number of eligible voters is often missing in the pdf documents, so voter turnout cannot be studied at this disaggregated level. In a small number of pdf documents, low-quality scans make some pieces of election outcome information illegible. Those are set to missing in the dataset.

A common problem with using election data is that election constituency boundaries do not coincide with administrative boundaries. Parliamentary constituencies (PCs) are created to ensure a roughly equal vote-to-seat ratio across the country, and each parliamentary constituency in India elects one politician to the Indian Parliament. NREGS was rolled out at the district level, however, so different parts of the district can be part of different parliamentary constituencies. The use of polling station data allows matching each observation to the correct district regardless of its parliamentary constituency, so there is no concern about measurement error when combining this data with other information.

Obtaining the National Sample Survey (NSS) datasets

The economic outcome data used in the analysis comes from the NSS, which is a large representative household survey collected by the Indian government. The paper uses data from the employment survey of the 61st round (2004-2005) and the 64th round (2007-2008). The NSS data is propriety and needs to be purchased from the Ministry of Statistics and Programme Implementation (MOSPI). Validated unit level data relating to various survey rounds are available on CD-ROMS which can be obtained from the Deputy Director General, Computer Centre, M/O Statistics and PI, East Block No. 10 R.K. Puram, New Delhi-110066 by paying the price along with packaging and postal charges as well as giving an undertaking duly signed in a specified format. The amount is to be paid by way of demand draft drawn in favor of Pay & Accounts Officer, Ministry of Statistics & Programme Implementation, payable at New Delhi.

Further details can be found on their website: http://mospi.nic.in/

Details on how the NREGS algorithm was constructed can be found in the paper and in the Online Appendix section "Additional Details on Algorithm Prediction Success". I summarize some of this information here.

Creating the normalized state-specific rank (running variable in the RD)

The datasets contain the district-specific normalized state rank that determines assignment to the various NREGS phases. The 'prevalence of poverty' measure is the state headcount ratio times the rural state population, which provides an estimate of the number of below-the-poverty-line people per state. A state is assigned the percentage of treatment districts that is equal to the percentage of India's poor in that state. For the calculations, I use headcount ratios calculated from 1993-1994

nationally representative National Sample Survey (NSS) data. I use the state headcount ratios from Planning Commission (2009), since the original headcount ratio calculations do not have estimates for new states that had been created since then. As official Planning Commission estimates, they are likely to be closest to the information the Indian government would have had access to at the time of NREGS implementation.
The development index used to rank districts within states comes from a Planning Commission report from 2003 that created an index of economic underdevelopment. The index was created from three outcomes for 17 major states: agricultural wages, agricultural productivity, and the district proportion of low-caste individuals (Scheduled Castes and Scheduled Tribes) (Planning Commission 2003). Data on the outcome variables was unavailable for the remaining Indian states, and it is unclear whether a comparable algorithm using different outcome variables was used for them. The empirical analysis is therefore restricted to these 17 states. Districts were ranked on their index values. In addition to the algorithm, the government had a separate list of 32 districts heavily affected by Maoist violence. These districts were not subject to the algorithm and all received NREGS in the first implementation phase. In order to closely replicate the algorithm used, these districts are dropped from the sample.
Ranks are made phase- and state-specific and are normalized so that a district with a normalized state-specific rank of zero is the last program-eligible district in a state in a given phase. Rank data in the 17 major states is complete for all rural districts. Rank data is available for 447 of 618 districts in the 2001 Census.

Datasets relevant to the analysis (files were created on Windows 10 operating system):

1. Election_data_2009.dta (Stata 16 format)
   This dataset contains polling-station level election results for the 2009 general election. It is used to generate the main variables of interest used in the analysis. Election data was digitized from pdf documents by the Election Commission of India. This data is merged with the state-ranks that determine NREGS assignment.

2. General_election_const.dta (Stata 16 format)
   This dataset contains constituency-level election results for 2004 and 2009. It is used for some baseline balance checks and to generate summary statistics. Election data comes from the Election Commission of India.

3. Phase_nregs.dta (Stata 16 format)
   This dataset contains just the NREGS information for each phase and is used to generate the figures for the algorithm distribution and first stage.

This folder contains the following do-files

a. Political_analysis_setup.do
   This do-file generates the samples for the political variables used in the analysis.

b. Setup do-files for economic outcomes
   Economic outcomes are analyzed using proprietary data from the 61st and the 64th round of the National Sample Survey (NSS) data. The file "NSS setup readme" in the nss folder describes the steps to follow to prep the raw data for the analysis. A series of .do files will

read in the raw datasets using dictionary files provided in the dict sub-folders and generate the variables for the analysis.

c. Analysis.do
This do-file replicates all the tables and figures in the paper.

Data Dictionary

state_code	State code
dist_code	District code
pc_code	Parliamentary constituency code
ac_code	Assembly constituency code
polling_station_code	Polling station code
state_name	State name
pc_name	Parliamentary constituency name
ac_name	Assembly constituency name
dist_name	District name
polling_station_name	Polling station name
ac_electors	Assembly constituency electors
ps_total_valid_votes	Polling station valid votes
ps_rejected_votes	Polling station rejected votes
ps_total_votes	Polling station total votes
ps_tend_votes	Polling station tendered votes
cand_name	Politician name
cand_sex	Politician sex
cand_category	Politician category (general, SC, ST)
cand_age	Politician age
partyabbre	Political party
totvotpoll	Politician votes received
cand_vote_share	Politician vote share
incumbent_from_2004	1 if incumbent from 2004
ac_name_2004	2004 assembly constituency if incumbent
partyabbre_2004	2004 party abbreviation if incumbent
totvot_2004	2004 votes received if incumbent
vote_share_2004	2004 vote share if incumbent
ac_type_2004	2004 ac type if incumvent
phase	NREGS implementation phase
nrega_phase_1	1 if in NREGS Phase 1
nrega_phase_2	1 if in NREGS Phase 2
nrega_pred_1	1 if predicted to be NREGS Phase 1
nrega_pred_2	1 if predicted to be NREGS Phase 2
comp_index	Composite index
rank	District rank based on composite rank
state_rank_norm_1	Normalized rank NREGS Phase 1
state_rank_norm_2	Normalized rank NREGS Phase 2
same_pc_diff_nregs_p1	Same parliamentary constituency different NREGS phases (Phase 1)
same_pc_diff_nregs_p2	Same parliamentary constituency different NREGS phases (Phase 2)
pop_2001	2001 Census district population
hh_dem	Number of households that demanded NREGS work (district admin data)

person_dem	Number of persons that demanded NREGS work (district admin data)
hh_empl	Number of households employed under NREGS (district admin data)
person_empl	Number of persons employed under NREGS (district admin data)
pd_total	Number of person-days under NREGS (district admin data)
hh_100	Number of households employed for 100 days (district admin data)