The Impacts of Conditional Cash Transfers in Four Presidential Elections (2002-2014)

(Replication Materials)*

Cesar Zucco Jr.
Brazilian School of Public and Business Administration
Getúlio Vargas Foundation
cesar.zucco@fgv.br

This version: September 21, 2015

This document provides a description of the replication materials and instruction to replicate the results in the paper. It is divided into three parts that correspond to the aggregate data analysis (at the municipal level), individual data analysis, and the speculative simulations discussed at the end of the paper. Note that some results in the paper employ techniques that can yield slightly different results each time they are performed. Please contact the author for saved results that allow for exact replication of the estimates and figures reported in the paper.

This Dataverse Study includes the following data files, which are explained, below:

- `data_replicGPS2014.RData`
- `data_replic2014beps.RData`
- `data_replic2014ibope.RData`
- `data_replicsimulation2010.RData`
- `data_replicsimulation2014.RData`

And the following .R script files, also explained below:

- `__replic.gps.2014`
- `__replic_2014_beps.R`
- `__replic_2014_ibope.R`
- `__replic_simmulation.R`

1 Aggregate Level Data

This analysis essentially replicates what I did in the AJPS piece for 2002, 2006, and 2010 (Zucco Jr. 2013). The code, in fact, uses the same functions as those presented in that paper, and for this reason, the replication materials included here only cover the 2014 election. One caveat to keep in mind is that while preparing these materials I realized I was dropping more municipalities than needed from the aggregate analysis (i.e. I was dropping municipalities that had missing values

*This is a preliminary version of the replication materials. Please contact me in case of errors or omissions.

on data that are not actually used), so the result one finds here is slightly smaller than the one reported in the paper (0.32 instead of 0.365).

The file `data_replicGPS2014.RData` is the dataset used in in the aggregate data analysis. It is composed of municipal level observations and contains the following variables:

- codeibge: Code of the municipality
- valid.2014: valid votes in the 2014 election
- dilma.vs.2014: Dilma's vote share in the 2014 election
- alcance.2014: Scope of the bolsa família program in October 2014
- spentH.2014.cct: Amount spent per capita in the municipality with Bolsa Família
- hdi.2010: Human development index of 2010 (the most recent one available)
- lula.vs.1998: Lula's vote share in 1998
- targetalcance.2014: The 2014 target of coverage of bolsa família
- pent.2010: Share of pentecostals in the municipality (2010 census)
- nonwhite.2010: Share of non-white population in the municipality (2010 census)
- distcap: Distance to capital city in tens of km.
- dgdpH.1.2014: GDP growth over the previous year (latest figures were used, and refer to 2013)
- gdpH.2014: GDP per capita (latest figures were used, and refer to 2013)
- mayorpty.2012: Party of mayor elected in 2012
- incgov.2014: Party of the incumbent (sometimes outgoing) governor in 2014
- state

The file `__replic.gps.2014` performs analysis, applying the same script as in the AJPS paper, which is based on Imai & van Dyk's (2004) generalized propensity score approach to continuous treatment.

2 Individual Level Data

Replication materials for the individual level data analysis of the 2002, 2006, and 2010 election are available in the DataVerse for the AJPS 2013 article, which is accessible by this link http://hdl.handle.net/1902.1/20257. The only slight difference is that, for symmetry, in the BPSR piece I reported only linear probability model estimates for each year. These are, however, very similar to what was reported in the AJPS piece.

For the 2014 election I analyzed one IBOPE survey taken close to the election and one pre-electoral and one post electoral wave of the BEPS 2014 survey (available soon, from the IADB). For the BEPS survey I extracted the relevant variables from the original dataset and merged in some municipal level variables of interest to produce the dataset that is contained in the file `data_replic2014beps.RData`, which is used here. This file contains the following variables:

- codeibge: Code of the municipality where individual lives
- bf: logical variable, indicating whether individual is a Bolsa Familia beneficiary
- dilma.pre: Intended to vote for Dilma in wave 1 of BEPS (May-June 2014)
- dilma.post Declared voting for Dilma in post first round wave of BEPS (October 2014)

- sexo: Sex
- idade: Age.
- escola: Schooling, ordered factor
- renda: Income/wealth, an index based on responses to household iitems
- region
- state
- race: race of respondent
- idtelefone: a unique identifier for respondents (some answered both waves)
- hdi.2010: Human development index of 2010 (the most recent one available)
- alcance.2014: Scope of the bolsa família program in October 2014
- distcap: distance to capital city
- metropolitan: whether the muncipality is part of a metropolitan area
- munsize: size of the municipality

The code to analyze the these data is provided in file `__replic_2014_beps`. The code performs the following steps:

1. Reads the data
2. Subsets the post-electoral respondents
    - Performs the matching routine
    - Analyzes balance in the matched dataset
    - Estimates a few different models for the effect of BF on the probability of voting
3. Subsets the pre-electoral respondents
    - Performs the matching routine
    - Analyzes balance in the matched dataset
    - Estimates a few different models for the effect of BF on the probability of voting

Note that genetic matching, which is employed here, can take a few minutes to complete, and will not generate the exact same results each time. The data extracted from the IBOPE survey are in the file `data_replic2014beps.RData`., which contains the following variables:

- bf: whether respondent declared being a beneficiary of bolsa família
- dilma: whether respondent intended to vote for dilma
- sexo: sex
- idade: age
- escola: schooling (ordered factor)
- renda: income (ordered factor)
- renda2: income aggregated to three levels
- metropolitan: whether respondent's municipality was part of a metropolitan area
- munsize: size of respodnent's municipality
- region: region
- state: state
- race: race

The code `__replic_2014_ibope.R` performs essentially the same analysis as described above for the BEPS survey.

3 Simulations

The simulations are non-systematic way to think about what might have generated the observed discrepancy between the aggregate and individual level estimates in 2014. It plugs in quantities that are "known" from the estimations reported earlier int he paper, and searches for values that are unknown and that could have led the aggregate results to looks as they did. The simulations are performed in the R script file `__replic_simmulation.R` and start from actual data on coverage of BF from 2010 and 2014 (files `data_replicsimulation2010.RData` and `data_replicsimulation2014.RData` and make some assumptions about the variability of the behavior of beneficiaries and non-beneficiaries that conform to what we "know" was true about the average difference in behavior between these two groups, based on the estimates reported earlier in the paper.

References

Imai, Kosuke & David van Dyk. 2004. "Causal Inference With General Treatment Regimes: Generalizing the Propensity Score." *Journal of the American Statistical Association* 99(467):854–866.

Zucco Jr., Cesar. 2013. "When Payouts Pay Off: Conditional Cash-Transfers, Clientelism, and Voting Behavior." *American Journal of Political Science* 47(3).