Politicians and Tax Policy: The Role of Preferences and Beliefs

María Inés Badin Diana Moreira Juan Francisco Santini Monica Singhal

This file contains general instructions on how to run the Stata do-files and all data files to reproduce the results of the paper. Data manipulation and analysis are performed using STATA 16.1.

The main data used in this paper was collected by *Hjort et al. (2021)*. In addition, we collected demographics, electoral, and budgetary data from official public sources. **To comply with the informed-consent procedures approved by the IRB of *Hjort et al. (2021)* and to protect the confidentiality of the participants, we have removed all identifying information from the proprietary data (i.e. survey data), and delinked the proprietary data and non-proprietary data (i.e. data from public sources) which could identify participants. Thus, some of the results of the paper are not reproducible using the information provided in this publicly posted package. In the do-files, we turn-off the parts of the code that refer to non-reproducible results (since proprietary data cannot be linked to non-proprietary data), and below we indicate with * the tables that are not reproducible. Please reach out to the authors for questions about the analysis not answered by this replication package.**

There are two main folders that contain several datasets and do-files to replicate the results. One folder contains raw public and proprietary data and the do-file that cleans this data. The second folder contains the main elements to reproduce the results in one "master" do-file that produces the results reported in tables and figures.

In sum, to reproduce the results of the paper you will have to run the following do-files:

- *raw data/public_rawdata_cleaning.do*
    - It takes approximately 1 minute to run completely
- *data analysis/tax_policy.do*
    - It takes approximately 10 seconds to run completely

The rest of the readme file is structured as follows. Section 1 describes the raw public data. Section 2 illustrates the proprietary data: the CNG baseline survey. Section 3 describes the results.

1. Raw Public Data

The folder *raw data* contains the following:

- *public_rawdata_cleaning.do* manipulates and constructs all the variables that are based on non-proprietary public data. To replicate the full cleaning, open Stata, and run *public_rawdata_cleaning*.do. **The code should be run from the start, in order, and without skipping lines**.
    - Municipalities' expenditures and tax revenues were retrieved from (year 2016):

        - https://siconfi.tesouro.gov.br/siconfi/pages/public/consulta_finbra/finbra_list.jsf
            - To download revenue data (click): Exercício – *select year* / Escopo – Municípios / Tabela – Receitas orçamentárias (Anexo I-C) / Consultar
            - To download expendture data (click): Exercício – *select year* / Escopo – Municípios / Tabela –Despesas por Função (Anexo I-E)/ Consultar
        - Raw data can be found in the *raw data/ finances* folder
    - Tax policy instruments variables are obtained from:
        - https://www.ibge.gov.br/en/statistics/social/health/19143-survey-of-basic-municipal-information-editions.html?=&t=o-que-e
            - To download the MUNIC data (click): Downloads – *2015* / Base_de_Dados
        - Raw data can be found in the *raw data/ munic survey* folder
    - Please refer to Hjort et al. (2021)'s documentation for an explanation on all other public data sources.

2. Proprietary data

The folder *raw data* also contains the following:

- *cng_survey_noid.dta* is the main dataset of the CNG survey. It can be found in the *raw data/cng survey* folder. The dataset is in "long" format, with one observation per municipality.

3. Results

The folder *data analysis* contains the following:

- *tax_policy.do* is the "master" do-file. It merges the manipulated public data and proprietary data, constructs the main variables used in the analysis, and produces - in order of appearance - the results reported in tables, of paper and appendix. To replicate the full analysis, open Stata, and run tax_*policy*.do. **The code should be run from the start, in order, and without skipping lines**.

    - Table 1: lines 27-106
    - Table 2: lines 108-173
    - *Table 3: lines 175-217
    - *Table 4: lines 219-234
    - *Table 5: lines 236-261
    - Figure 1: lines 268-283
    - Figure 2: lines 285-309
    - Appendix Table A.1: lines 315-345
    - Appendix Table A.2: lines 347-412

- secondary folder:
    - *samples.dta* is a secondary dataset that identifies whether a municipality (and its mayor) was assigned to the treatment or control group of *Hjort et al. (2021)* (treatment_assignment); attended Hjort's information session (attended); completed the CNG survey before or without attending the session (invited or not, sample_notinfluenced); and completed the CNG survey before the session and was invited to the session (sample_session).

Notes:

- All results are shown in Stata "Results Window".