README for: Coarse Wage-Setting and Behavioral Firms

Overview

The code in this replication package constructs the analysis files from six data sources described below using Stata. One main code runs all of the codes to clean and prepare the data for the analysis, and one code generates the figures and tables in the paper. The analysis relies extensively on a confidential microdata that the author is not permitted to distribute due to Data Use Agreements. Details on data sources and how to access them are provided below.

Data Availability and Provenance Statements

☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

If box above is checked and if no simulated/synthetic data files are provided by the authors, please skip directly to the section on Computational Requirements. Otherwise, continue.

Statement about Rights

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Summary of Availability

☐ All data **are** publicly available.

☒ Some data **cannot be made** publicly available.

☐ **No data can be made** publicly available.

Details on each Data Source

The analysis in this paper relies on confidential as well as publicly available datasets. The use of the confidential dataset is covered by Data Use Agreements that prevent the author from distributing the data to anyone else. Below are details on each of these datasets as well as contact info for each data-providing agency. The author will assist with any reasonable attempts to gain data access and replicate the analysis for two years following publication.

The paper uses the following datasets:

1. The main dataset contains the 1999–2017 years of Brazil's employer-employee dataset, the *Relação Anual de Informações Sociais*, or RAIS. This dataset covers the universe of formal-sector jobs in Brazil and is managed by the Ministry of Labor. The following link provides information on how researchers can apply for access to the identified version of the RAIS: http://www.rais.gov.br/sitio/index.jsf. The following link provides access to a de-identified version of the RAIS: https://basedosdados.org/dataset/3e7c4d58-96ba-448e-b053-d385a829ef00?table=c3a5121e-f00d-41ff-b46f-bd26be8d4af3. The de-identified version can be used to replicate all the results from Section 2 of the paper.

2. The second dataset is the 2013 Brazilian Household Survey (*Pesquisa Nacional por Amostra de Domicílios*, PNAD). The PNAD is a nationally-representative survey conducted annually by the National Statistics Office to measure several characteristics of the population, such as household composition, education, and income. The replication package includes a Stata dataset called `pnad_2013.dta`, which is the cleaned version of the microdata used in the paper.

3. The third dataset is the 2013 Brazilian Labor Force Survey (*Pesquisa Mensal de Emprego*, PME). The PME is a monthly survey conducted in six large metropolitan areas to provide frequent updates on the unemployment rate and other labor-market variables. The replication package includes a Stata dataset called `pme_2013.dta`, which is the cleaned version of the microdata used in the paper.

4. The third dataset is the 2010 Brazilian Population Census (*Censo Demográfico*). The census is conducted approximately every ten years to count the population of the country, but it also includes earnings information. The replication package includes a Stata dataset called `census2010.dta`, which is the cleaned version of the microdata used in the paper.

5. The fourth dataset is the 2013 Social Programs Registry of Individuals (*Cadastro Único*). The Social Programs Registry contains information on all beneficiaries of government programs, including their earnings. A de-identified version of this dataset can be accessed here: https://www.gov.br/mds/pt-br/pt-br/servicos/sagi/microdados. The replication package includes a CSV file with the raw data (`base_amostra_pessoa_201312_20190131.csv`) and the code to clean the raw data.

6. The sixth dataset is the universe of collective bargaining agreements (CBAs) signed during 2008–2017. This data was collected by Lagos (2023). The replication pack-

age includes a Stata dataset called `cba.dta`, which is a cleaned version of the microdata collected by Lagos (2023).

7. The paper also uses several other auxiliary datasets that are provided in the replication package: Monthly Consumer Price Index (CPI) information (`inpc_202003SerieHist.xls` and `cpi_mthly.dta`), Yearly CPI (`cpi_yrly.dta`), region-level CPI (`ipca_15.dta`), mapping between municipalities and microregions (`muni_identifiers.csv`, `microregion.dta`, and `rais_codemun_to_mmc_1991_2010.dta`), mapping between municipalities and metropolitan regions (`muni_metreg.dta`), state-level minimum wages (`state_mw.dta` and `state_mw.xlsx`).

Computational requirements

Software Requirements

- Stata (code was last run with version 15)
  - `estout` (as of 2023-12-01)
  - `reghdfe` (as of 2023-12-01)
  - `ftools` (as of 2023-12-01)
  - `gtools` (as of 2023-12-01)
  - `bunching` (provided by the author)

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses:

☐ <10 minutes
☐ 10-60 minutes
☐ 1-2 hours
☐ 2-8 hours
☐ 8-24 hours
☐ 1-3 days
☒ 3-14 days
☐ > 14 days
☐ Not feasible to run on a desktop machine, as described below.

Description of programs/code

- `code/0-master.do`: This code executes all the files that clean the raw data, merge the cleaned datasets, and prepare the final analysis datasets. The codes that this file executes are grouped into three subfolders:

- `code/clean`: There are 4 codes in this folder
- `code/results`: The code in this folder produces the figures and tables in the paper. Below I detail the location of each figure/table in the code.
- The rest of the codes in the `/code` folder are auxiliary files used as inputs in other codes.

List of tables and programs

The provided code reproduces:

☐ All numbers provided in text in the paper
☒ All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified below.

Figure/Table	Program	Line(s)	Note
Figure 1	code/results/figures.do	30 & 56	
Figure 2	code/results/figures.do	96, 150, 195, & 242	
Figure 3	code/results/figures.do	153	Requires confidential data
Table 1	code/results/tables.do	25	
Table 2	code/results/tables.do	181	Requires confidential data
Table 3	code/results/tables.do	181	Requires confidential data
Table 4	code/results/tables.do	459	Requires confidential data

Acknowledgements

Some content on this page was copied from Hindawi. Other content was adapted from Fort (2016), Supplementary data, with the author's permission.