**“Corruption and Firms” (Colonnelli and Prem)**

**Overview**

This replication package contains a description of the confidential data sources used in the paper Corruption and Firms, as well as the Stata code used to generate the tables and figures of the main paper.

**Data Availability Statement**

The two main sources of data used in the paper are the RAIS and PAC/PAS databases.

The RAIS (*Relação Anual de Informações Sociais*) is a matched employer-employee dataset managed by the Brazilian Ministry of Labor and Employment. The RAIS data are described at this link: <http://portalfat.mte.gov.br/relacao-anual-de-informacoes-sociais-rais/> .

We obtained access to RAIS data under an agreement signed on October 15, 2015, together with Ministry of Labor and Employment-MTE, which nowadays is part of Ministry of Economy. We were given access to these data under the agreement that they could not be made publicly available.

Researchers interested in working with RAIS can apply through their institutions and get access to RAIS after approval and after relevant Non-Disclosure-Agreements are co-signed. Currently, the Secretary of Labor within the Ministry of Economy manages the application process for RAIS data, which is described at the link <https://www.gov.br/pt-br/servicos/solicitar-acesso-aos-dados-identificados-rais-e-caged>.

Notice that the RAIS data undergoes quality checks and updates every year. As a result, there might be minor differences depending on the time in which the data was obtained from the MTE.

These are two examples of papers that have used the RAIS data and that have recently been published in REStud (these two papers also rely on additional data sources):

* Theory and Inequality: From Theory to Estimation (2016) https://academic.oup.com/restud/article/84/1/357/2669956
* Regional Transfer Multipliers (2018) <https://academic.oup.com/restud/article/86/5/1901/5210878>

The PAC/PAS (*Pesquisa Annual de Comercio* (PAC), *Pesquisa Annual de Servicos* (PAS)) data are firm-level datasets managed by the Brazilian Institute of Geography and Statistics (IBGE). The data are described at these links: <https://www.ibge.gov.br/en/statistics/economic/trade/17210-annual-survey-of-trade.html> , <https://www.ibge.gov.br/en/statistics/economic/services/17345-annual-survey-of-services-pas1.html?=&t=o-que-e> . We are not aware of academic papers that have used the PAC and PAS datasets.

We obtained access to the PAC and PAS datasets through an agreement with the Brazilian Institute of Geography and Statistics (IBGE). We were allowed to use the data through physical visits to the Rio de Janeiro office of IBGE. The visits are necessary to obtain access to the confidential room from which we could access the computers linked to the server that contains the datasets.

Researchers interested in working with PAC and PAS can request confidential access by reaching out to [ibge@ibge.gov.br](mailto:ibge@ibge.gov.br).

The PAC/PAS data undergo quality checks and updates every year. As a result, there might be minor differences depending on the time in which the data was obtained from IBGE.

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.

Summary of availability: No data can be made publicly available.

**Dataset list**

Because of the confidentiality issues discussed in the DAS, this replication package does not contain the data. However, this is a summary of the data structure needed to run the replication code we provide, conditional on obtaining access to the RAIS and PAC/PAS data as discussed above:

* Data\_Yearly: RAIS dataset collapsed at the municipality-year level. *– Confidential / Not provided*
* Data\_Quarterly: RAIS dataset collapsed at the municipality-year-quarter level. *- Confidential / Not provided*
* Data\_Firms: PAC/PAS dataset collapsed at the firm-year level. *- Confidential / Not provided*

**Description of programs/code**

There are two do files only.

The *Figures.do* file generates the figures in the paper using the RAIS data, i.e., Figures 1, 2, 3, and 4.

The Tables.do file generates the tables in the paper using the RAIS and PAC/PAS data, i.e., Tables 1, 2, 3, 4, 5, 6, 7, and 8.

**Computational requirements**

The full code can be run using Stata/MP 15.1. The code was developed and tested on a Mac.

The following are the non-default Stata packages that are needed for the programs to run, together with instructions on how to install them.

* reghdfe: to install the latest version write the following in the command window in stata
  + Install ftools (remove program if it existed previously)
    - cap ado uninstall ftools
    - net install ftools, from("https://raw.githubusercontent.com/sergiocorreia/ftools/master/src/")
  + Install reghdfe 6.x
    - cap ado uninstall reghdfe
    - net install reghdfe, from("https://raw.githubusercontent.com/sergiocorreia/reghdfe/master/src/")
  + Install parallel, if using the parallel() option; don't install from SSC
    - cap ado uninstall parallel
    - net install parallel, from(https://raw.github.com/gvegayon/parallel/stable/) replace
    - mata mata mlib index
* distinct: to install the latest version write "ssc install distinct" in the command window in stata
* outreg2: to install the latest version write "ssc install outreg2" in the command window in stata
* parmest: to install the latest version write "ssc install parmest " in the command window in stata

**Memory and runtime requirements**

Once the data is aggregated at the yearly and quarterly level, as discussed above, the full code can be run in less than 2 hours on most desktop machines.

**References to the confidential datasets**

Brazilian Institute of Geography and Statistics (2002-2014). Pesquisa

Annual de Comercio - PAC, Pesquisa Annual de Serviços - PAS.

https://www.ibge.gov.br/en/statistics/economic/trade/17210-annual-survey-of-trade.html.

(accessed April 19, 2021).

Brazilian Ministry of Labor and Employment (2002-2017). Relação Anual de Informações Sociais - RAIS. http://portalfat.mte.gov.br/relacao-anual-de-informacoes-sociais-rais/. (accessed April 19, 2021).