Code Documentation for: Paying Outsourced Labor:

Direct Evidence from Linked Temp Agency-Worker-Client Data

Andres Drenik	Simon Jäger	Pascuel Plotkin	Benjamin Schoefer
Columbia	MIT	UBC	UC Berkeley

February 2021

Our study is mainly based on the Argentine Social Security dataset SIPA (Sistema Integrado Previsional Argentino) and Simplificacion Registral (SR). Earnings are deflated using the CPI from Cavallo & Bertolotto (2016) downloadable from (**http://www.thebillionpricesproject.com/our-research/**). The hours analysis relies on the Argentina LFS, Encuesta Permanente de Hogares (EPH) which can be downloaded from (**https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos**) (2012-2018).

Software requirements: Stata 16. There are a number of commands used throughout the code that must be installed prior to running it. We provide a comprehensive installation section in the "Installing Packages" section of do-file **00a_SetPaths.do**. Tables may require additional manual formatting. We ran our code on an 4-core Intel Xeon processor with 128GB of RAM running on Windows 10 pro.

1 General instructions to run the code

1. The two main datasets we use in our analysis are confidential administrative, which we analyzed on a secure server base at Ministerio de Trabajo, Empleo y Seguridad Social. A subset of SIPA can be acquired from **http://www.trabajo.gob.ar/estadisticas/oede/mler.asp**. Please contact the authors for the replication of specific tables and figures in the paper.

2. Place all do-files in the appropriate folders. See the **00a_SetPaths.do** description below.

3. Modify the home global in do-file **00_MasterDoFile.do** to an appropriate folder.

4. Do-file **00a_SetPaths.do** sets path for:

   (a) Input data, containing the raw unfiltered SIPA dataset

   (b) User provided data containing RENAPER, Simplificacion Registral, minimum wage data, the consumer price index (CPI), and 4-Digit Industry codes to filter (and merge with) the SIPA dataset

(c) The filtered data

(d) Encuesta Permanente de Hogares (EPH)

(e) Utilities where a do-file related to specific calculations of the EPH is located

(f) Two intermediate data folders, respectively for data cleaning and empirical analysis

(g) Results data, which we use to generate figures and tables

(h) Do-files, where we place all of the do-files for this paper

(i) Logs, where we store our log files

5. To install packages, set "scalar install_packages == 1" in the do-file **`00a_SetPaths.do`**.

6. To replicate any part of the analysis, we suggest loading do-file **`00_MasterDoFile.do`** and uncommenting the specific section of the analysis you wish to run. We have commented out all the do-files except do-file **`00a_SetPaths.do`** for convenience. Additionally

Output Conventions In **`00_MasterDoFile.do`**, figures and tables are run in the same order as they appear in the text and appendix. Figures are saved in **`.pdf`** format, while tables are saved in .tex format. One do-file is usually provided for each figure or table. Log files match name of do-file. File names for figures and tables begin with **`FigXXX`** or **`TabXXX`**, where **`XXX`** is the reference number in the main or Appendix text.

In-text Calculations There are some computations mentioned in the text (yet not explicitly shown in Tables or Figures) that are carried out at different stages in our code. Three of these are explicitly done in separate do-files which are denote under **`*_Calculations`** where * is the description of the calculation. We provide a list of in-text calculations and their relation to a specific do-file at the end of this document.

Additional Clarification on EPH The calculations that rely on the EPH, use as inputs the raw data that can be downloaded from (**`https://ww w.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos`**). The raw data files are in a **`.txt`** format and are named **`usu_individual_*`** where **`*`** represents the quarter and the year of the data.

2 Code Structure

Building the Data

- **`01_Registro_Merge_Prep.do`** cleans the SR dataset and prepares it in order to be merged with SIPA.
- **`02_Data_Prep.do`** cleans the SIPA dataset, flags public workers, and merges the SIPA dataset with the SR data.
- **`03_Eph_Cont_2012_2018_Creation.do`** generates a dataset of the appended EPH samples from 2012 to 2018.

Preparing the Regression Sample

- **04_Wage_Threshold.do** generates the minimum wage threshold.
- **05_Filter_Full_Data.do** filters the main dataset for the AKM analysis. It drops public workers, generates real wages, applies the minimum wage threshold, and keeps the highest paying job in a month for each worker.
- **06_Workers_Group.do** creates a dataset of a random group of workers for the measurement error correction analysis.
- **07_Multicolinearity_Check.do** creates the dataset that allows us to check for potential multicolinearity in temp agencies when running Equation (2).

Main analysis files

- **08_Reg_Tenure.do** creates the dataset that contains regular worker tenure by firm.
- **09_Akm_Preparation.do** is the main do-file for the analysis exposed in the paper. This do-file prepares and estimates equation (2).
- **10_Mincer_Equations.do** runs equation (1) in order to estimate the results presented in Table 1.

Auxiliary files for Figures

- **11_CBA_Coverage.do** generates a dataset that indicates the share of CBA coverage by 4-digit industry.
- **12_CBA_Collapse.do** reshapes and filters dataset for CBA coverage by industry.

In-Text Calculations

- **13_SD_Calculations.do** estimates the standard deviation of temporary and regular work arrangement fixed effects that are presented in the text.
- **14_Temp_Calculation_2005.do** calculates the amount of temp workers in 2005.
- **15_Connected_Set_Calculations.do** calculates the size of the connected set.
- **16_User_Firms_Connected_Calculations.do** calculates the magnitude of user firms in the connected set.

Figures

- **Fig2.do**
- **Fig3.do**
- **Fig4.do**

Appendix Tables

- **TabA1.do**
- **TabA2.do**

Appendix Figures

- **FigA1a.do**
- **FigA1b.do**
- **FigA2.do**
- **FigA3.do**
- **FigA4.do**

3 Additional in-text Calculations

Calculation	Source/log file
Specifically, user firm pay premia for temp workers have a standard deviation of 17.2 log points, but this rises to 20.7 log points in regular work arrangements for the sample of user firms. These dispersion measures are robust to a split-sample measurement error correction, which shrinks the standard deviations to 15.2 log points for pay policies for temp work arrangements, whereas it leaves the regular work arrangements largely unaffected, at 20.5 log points.	**13_SD_Calculations.do**
Finally, about 1.7% of employees were employed through a temp agency in 2005	**14_Temp_Calculations.do**
if anything, temporary workers appear to work slightly more hours (36.18 hrs/week, SD 12.15, vs 34.61 hrs/week, SD 13.16)	**FigA1b.do**
We estimate (2) in the largest connected set, which captures 60.8% of firms and 95.9% of worker-year-spell observations	**15__Connected_Set_Calculations.do**
The firms relying on temp labor are larger, as they make up 32.2% (1.6%) of our connected set sample of worker-month (total firms) observations	**16_User_Firms_Connected_Calculations.do**
Instead weighting firm observations by the number of temporary (rather than all) workers yields a weighted-mean difference of 0.13	**Fig2.do**
Instead weighting firm observations by the number of temporary workers (rather than all workers) would yield a slightly higher slope of 0.61 (SE 0.0055)	**Fig3.do**
Figure 4 footnotes	**Fig4.do**
Appendix C results	**FigA4.do**

4 Variables Description

Variable	Description
cuit_empl	Unique firm ID
cuil_trab	Unique worker ID
remuner_total	Total earnings in a month
modalidad	Contract modality
ciiu_4	4-digit industry code
ciiu_4_user	4-digit industry code, on which the temp workers are assigned the code of the user firm
industry_code_2digit	2-digit industry code
real_firm	Firm where the worker is actually working (equal to cuit_empl for regular workers, equal to cuit_empl of the user firm for temp workers)
temp_agency_id	Unique firm ID for temp agencies
temp_worker	Dummy indicating if the worker is a temp worker in that period
index	CPI index used to generate real wages
mw_threshold	Minimum wage threshold used
firm_fe3	Regular firm fixed effects estimated through an AKM model without temp agency fixed effects
firm_fe3_g*	Regular firm fixed effects estimated through an AKM model without temp agency fixed effects, estimated using a random sample of workers
firm_fe3_gd*	Regular firm fixed effects estimated through an AKM model without temp agency fixed effects, estimated using a certain period of time
firm_fe4	Regular firm fixed effects estimated through an AKM model with temp agency fixed effects
firm_fe4_g*	Regular firm fixed effects estimated through an AKM model with temp agency fixed effects, estimated using a random sample of workers
firm_fe4_gd*	Regular firm fixed effects estimated through an AKM model with temp agency fixed effects, estimated using a certain period of time
uf_fe3	User firm fixed effects estimated through an AKM model without temp agency fixed effects
uf_fe3_g*	User firm fixed effects estimated through an AKM model without temp agency fixed effects, estimated using a random sample of workers
uf_fe3_gd*	User firm fixed effects estimated through an AKM model without temp agency fixed effects, estimated using a certain period of time
uf_fe4	User firm fixed effects estimated through an AKM model with temp agency fixed effects

uf_fe4_g*	User firm fixed effects estimated through an AKM model with temp agency fixed effects, estimated using a random sample of workers
uf_fe4_gd*	User firm fixed effects estimated through an AKM model with temp agency fixed effects, estimated using a certain period of time
worker_fe3	Worker fixed effects estimate through an AKM model without temp agency fixed effects
worker_fe4	Worker fixed effects estimate through an AKM model with temp agency fixed effects
connected_set	Dummy indicating if the firm is in the largest connected set
obs	Amount of worker-month observations for each firm
obs_temp	Amount of temp worker-month observations for each firm
tenure	Average tenure of regular workers in firm
lom_tenure	Leave-out tenure used in Figure 4.a
lom_sd_firm_fe4	Leave-out SD used in figure 4.b
lom_firm_fe4	Leave-out regular firm FE used in figure 4.c
share_cov	Share of CBA coverage used in figure 4.d
perc_temp_workers	Percentage of temp workers per industry
perc_non_temp_workers	Percentage of non-temp workers per industry
pp3e_tot	EPH variable: amount of hours worked in a week
pondera	EPH variable: survey-sample weight
industria_ocup_prin_caes	EPH variable: industry of main job
pp07c	EPH variable: If the contract has a fixed term ending
cat_ocup	EPH variable: Occupational category
estado	EPH variable: If employed or unemployed
ano	EPH variable: year