

# README for Jardim et al., 2021

## Overview

The code in this replication package constructs the analysis file from the three data sources (BLS, 2017; Census, 2017; ESD, 2016) using Arc GIS, Stata and R. Two master files run the code to generate the data for 22 tables and figures in the paper. Due to data restrictions, the data to generate 17 tables and figures is unable to be disclosed. (See dataset list and list of tables and programs for additional information). The replicator should expect the code to run for 2-3 days.

## Data Availability and Provenance Statements

### 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

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

### Details on each Data Source

#### ESD DAS

Records on individual employment, hours, place(s) of employment, and earnings were provided to the University of Washington by the Washington State Employment Security Department (ESD) pursuant to an interagency data sharing agreement. These records include employee names, social security numbers, hours worked, and earnings, and reported each calendar quarter by each business entity covered by Unemployment Insurance. Business-level records include the name and ESD account number of the business, a mailing address for correspondence, and a modified county code. For businesses operating in a single Washington state location, the modified county code indicates the Washington county in which it is located. For businesses that operate in multiple locations within Washington state, the modified county code does not indicate a county. Data employed in this analysis span the time period from the first calendar quarter of 2005 to the third calendar quarter of 2016.

The Revised Code of Washington establishes that records held by ESD “shall be private and confidential” unless there is a specific exemption codified in law (RCW 50.13.020). One such exemption applies to “Governmental agencies... conducting authorized statistical analysis, research, and evaluation studies” (RCW 50.13.060 (6)). Permission to access data for the purposes of this study was sought and granted on the basis of the University of Washington’s standing as an agency of Washington state government (RCW 28B.20).

Researchers seeking access to the same private and confidential data would need to first become employees of an eligible state, local, or Federally recognized tribal government agency within Washington. The agency would then need to submit an application to ESD

for review by the chief executive (RCW 50.13.060 (1)). In general, agencies submitting applications are required to serve notice to every individual and employer for whom records are being sought. Contacted individuals then, by law, have five days to register objections to the release of their records with ESD. For purposes of conducting the minimum wage study, the University of Washington was exempted from this notification requirement on the grounds that it was conducting a labor market policy evaluation under contract with the City of Seattle (RCW 50.38.060; 50.13.060 (6)). In the event of any ambiguity in the provisions of state law, the ESD chief executive has sole authority of interpretation (RCW 50.13.030).

The ESD has made an alternate version of the UI data available with broader access. This alternate version removes all identifying information for both employees and businesses. As mailing address information is potentially identifying for businesses, it is not included. Without mailing address information, it is generally not possible to verify that a business is located in the City of Seattle or any other municipality in the state. For this reason, the alternate data cannot be used to replicate this study. Researchers interested in accessing this version of the UI data may contact the office of the Labor Market Information Director at ESD.

For this project, the ESD has approved the disclosure of select data which can be used in the replication process. These data can replicate 22 of the 39 tables and figures in this paper.

Code to process the employee-employer confidential data, which includes the incorporation of a wage deflator (see CPI DAS) and merged geographic information (see PUMA and ArcGIS DAS), is included. Due to confidentiality, the data to replicate this code cannot be provided.

No portion of the worker level data can be provided, due to confidentiality restrictions. Code to create the tables and figures using the worker level data are provided in the spirit of transparency.

Datafiles (the accompanying Data\_dictionary.docx file provides details for each dataset):

- aggregate\_establishments\_cumulative\_all.dta
- aggregate\_establishments\_groups.dta
- region\_level\_cumulative.dta
- nworkers\_bins.dta
- entrants\_only\_15\_5y\_PUMA.dta
- hours\_decomposed\_jobslt19.dta

### CPI DAS

Data on Consumer Price Index (CPI) were downloaded from the U.S. Bureau of Labor Statistics (BLS, 2017). Data can be downloaded from <https://www.bls.gov/cpi/data.htm>. Under “Consumer Price Index (CPI) Databases” select “One Screen” database for “Urban Wage Earners and Clerical Workers (Current Series).” Select “U.S. city average”, “All items” and check the “Not Seasonally Adjusted” box to get data and download. Data can also be

directly downloaded using <https://data.bls.gov/pdq/SurveyOutputServlet> (Series Id: CWUR0000SA0). A copy of the data is provided as part of this archive. The data are in the public domain.

Datafile: `cpi_w_00_16.csv`

### ARC GIS DAS

Data used to geocode the precise latitude and longitude of employer addresses from the ESD data came from the 2016 ArcGIS Business Analyst database. We specifically used geocoding data from the USA Local Composite. This data is proprietary. More information can be found here: <https://desktop.arcgis.com/es/arcmap/10.4/extensions/business-analyst/whats-business-analyst.htm>.

Datafile (not included due to confidentiality): `Geocodes_v4.dta`

### PUMA DAS

Data on Public Use Microdata Areas were downloaded from the U.S. Census Bureau (Census, 2016). Data can be downloaded from <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2010.html>. Click the Web Interface to select the year 2010 and Public Use Microdata Area layer type for Washington download the zipped shapefiles. The file used in geocoding analysis is the `tl_2010_53_puma10.shp` file. A copy of the data is provided as part of this archive. The data are in the public domain.

Data folder: `tl_2010_53_puma10`

### Dataset list

See `Programs_data_README.xlsx` for a complete list of programs and data provided in the replication package.

## Computational requirements

### Software Requirements

- Stata (code was last run with version 15.1)
  - `synth` (as of 2021-02)
  - `grc1leg` (as of 2021-02)
  - the program “`packages.do`” will install all dependencies locally, and should be run once.
- R (version 4.0.3 (updated 2020-10-10))
  - Platform: `x86_64-w64-mingw32/x64` (64-bit)
  - Running under: Windows Server x64 (build 14393)

- Tidyr\_1.1.2
- maptools\_1.0-2
- rgeos\_0.5-5
- ggmap\_3.0.0
- RColorBrewer\_1.1-2
- tis\_1.38
- systemfit\_1.1-24
- multiwayvcov\_1.2.3
- sandwich\_3.0-0
- zoo\_1.8-8
- data.table\_1.13.2
- rgdal\_1.5-18
- sp\_1.4-4
- psych\_2.0.9
- tidyr\_1.1.2
- scales\_1.1.1
- dplyr\_1.0.2
- stringr\_1.4.0
- lmtest\_0.9-38
- limSolve\_1.5.6
- readstata13\_0.9.2
- ggplot2\_3.3.2
- MatchIt\_4.1.0 (used with confidential data only)

the file “packages\_and\_functions.R” will install all dependencies (latest version), and should be run once prior to running other programs.

## Memory and Runtime Requirements

### Memory

The code was last run on a Windows Server x64 (build 14393) with 256 BG of RAM. It is a Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40 GHz with two processors.

### Runtime

Approximate time needed to reproduce the analyses on a standard desktop machine: 1-2 days. Most scripts take minutes to run. The following are more computationally expensive.

- table\_6.r (~10 hours)
- table\_a5.r (<1 hour)
- figure\_3.r (<1 hour)
- figure\_5\_a12.do (~4 hours)

## Description of programs/code

- The main replication folder contains the folder/code, which contains all of the scripts necessary to replicate the analysis.
- The folder /data is where we store the datasets necessary to run the code.
- The folder/output begins empty and is where the output of the scripts will be stored once you begin.
- Programs in the /code folder generate 22 of the tables and figures in both the main body and the appendix of the paper.
  - The program code/0\_Master\_all\_aggregate\_analysis.r will run all of the parts of the replication which are done in R and a file utilized in code/1\_Master\_entrants.do (line 21).
  - The program code/1\_Master\_entrants.do will run all of the parts of the replication which are done in Stata. Make sure lines 10-21 are run from 0/Master\_all\_aggregate\_analysis.r.
- Each program called from the master files above identifies the table or figure it creates (e.g., code/table\_5.r). Output files are called appropriate names (e.g., /output/table\_5.csv) and should be easy to correlate with the manuscript. Details on each script, the data it requires, and the output associated, can be found in Programs\_data\_README.xlsx.
- The folder code/code\_confidential contains the code used to clean and process the ESD data and code for the non-replicable tables and figures. These programs cannot be run by replicators but are included in the spirit of transparency. The file /code\_confidential/00\_building\_Master.txt lists the order to which the cleaning and processing files are executed. The file /code\_confidential/1\_STATA\_tables\_figures.do runs all the confidential tables and figures which are done in STATA. The file /code\_confidential /1\_R\_tables\_figures.r runs all the confidential tables and figures which are done in R.

## Instructions to Replicators

- Edit code/0\_Master\_all\_aggregate\_analysis.r and code/1\_Master\_entrants.do to adjust the default path
- Run code/0\_Master\_all\_aggregate\_analysis.r all at once to replicate all analysis done in R.
  - These programs were last run top to bottom on February 12, 2021.
  - These programs last took 12.5 hours to run.
- Run code/1\_Master\_entrants.do to replicate all analysis done in Stata.
  - These programs were last run top to bottom on February 26, 2021.

- These programs last took 4.5 hours to run.

### Details

- `code/packages_and_functions.r` has loads all of the required packages, and it calls `code/functions.r` which contains all of the necessary functions.
- To cherry pick tables/figures to run, make sure you first run lines 10-21 of `code/0_Master_all_aggregate_analysis.r` once on each new system to set up the working environment before running any subsequent R script. This chunk of code creates the folder structure and dependencies. Line 21 must also be run prior to running `code/1_Master_entrants.do` as it creates a dataset required in the STATA analysis.
- Many scripts depend on output generated in `code/table_6_a.r` and `code/table_6_b.r`. Make sure to run those two scripts before any other r script that begins with “table” or “figure.” Exact dependencies are detailed in `README_table.xlsx`. `Table_6_a.r` and `Table_6_b.r` can be run in parallel if necessary.
- Table 1 is information from the Seattle Minimum Wage Ordinance. A link to the legal text has been provided in the `Programs_data_README.xlsx` file. Figure A3 is a map of the PUMAs of Washington state. A link to Washington Public Use Microdata Area maps are provided in the `Programs_data_README.xlsx` file.

### List of tables and programs

The provided code reproduces all tables and figures in the paper. Some of these tables and figures can be reproduced with the data we provide. The tables and figures that are reproducible with the data we are providing and those that require the confidential data are clearly labelled in `Programs_data_README.xlsx`

### References

U.S. Bureau of Labor Statistics (BLS). 2017. “CPI for Urban Wage Earners and Clerical Workers (CPI-W).” (accessed February 6, 2017).

U.S. Census Bureau (Census). 2016. “TIGER/Line Shapefiles.” (accessed December 16, 2016).

Employment Security Department, 2005-2016. “Confidential Unemployment Insurance Data [dataset]” *Washington State Employment Security Department (ESD)*. (accessed on varying dates).

ArcGIS software by Esri (GIS). 2017. “2016 ArcGIS Business Analyst database.” [dataset] (accessed December, 2017)

## Acknowledgements

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