Read Me

Overview

The code in this replication package constructs the analysis file from the five data sources (Guner et al., 2014; Acemoglu and Restrepo, 2017; Chetverikov et al, 2016; WID, 2020; BLS, 2017) using Stata, R, and Julia. Three files run all the code necessary to replicate Figure 1, Figure 2, and Table 1. The replicator should expect the code to take less than 10 minutes.

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).

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

- Data on the labor share were downloaded from the Bureau of Labor Statistics (BLS). Data comes from figure 1 in Giandrea and Sprague (2017) available at https://www.bls.gov/opub/mlr/2017/article/estimating-the-us-labor-share.htm. They can be directly read into excel from https://www.bls.gov/opub/mlr/2017/images/data/giandrea-sprague-fig1.stm. A copy of the data is provided as part of this archive. The data are in the public domain.

  Datafile: Table 1 Section 5/RawData/labor_share.xlsx

- Data on income shares of different percentiles were downloaded from the World Income Database (WID). It is downloaded in the code Table 1 Section/Code/Cleandata.do using the special WID stata package. We use the data series spllin and apllin for the US in year 2000. The three data sets derived from this are included as a part of this archive. The data are in the public domain.

Datafile: Table 1 Section 5/RawData/Income_Shares.dta, Table 1 Section 5/RawData/Income_Shares_deciles.dta, Table 1 Section 5/RawData/Income_Shares_percentiles.dta

- Data on marginal income taxes were provided to us from the authors of Guner et al (2014) upon request. It corresponds to data reported in Figures 1, 2, 5, and 6 of that paper. The data are included as a part of this archive and also reported in Appendix E.

  Datafile: Table 1 Section 5/RawData/gkv_Figures_RED_USTaxes_2013.xls

- We get the data on the wage effects of robots upon request from the authors of Acemoglu and Restrepo (2017). It corresponds to data reported in Table 13 of that paper. The data are included as a part of this archive and also reported in Appendix E.

  Datafile: Table 1 Section 5/RawData/results_by_percentile.dta

- We get the data on the wage effects of trade upon request from the authors of Chetverikov et al. (2016). It corresponds to data reported in Figure 1 of that paper. The data are included as a part of this archive and reported in Appendix E.

  Datafile: Table 1 Section 5/RawData/fig1 point estimates for arnaud.xlsx

- Code for data cleaning and analysis is provided as part of the replication package. It is included in this archive.

  File: Table 1 Section 5/Code/CleanData.do

Dataset list

Data file	Source	Notes	Provided
`Table 1 Section 5/RawData/labor_share.xlsx`	BLS		Yes
`Table 1 Section 5/RawData/Income_Shares_percentiles.dta`	WID		Yes
`Table 1 Section 5/RawData/Income_Shares.dta`	WID	Derived from above.	Yes
`Table 1 Section 5/RawData/Income_Shares_deciles.dta`	WID	Derived from above.	Yes

Data file	Source	Notes	Provided
`Table 1 Section 5/RawData/gkv_Figures_RED_USTaxes_2013.xls`	Guner et al. (2014)	Provided upon request.	Yes
`Table 1 Section 5/RawData/results_by_percentile.dta`	Acemoglu and Restrepo (2017)	Provided upon request.	Yes
`Table 1 Section 5/RawData/fig1 point estimates for arnaud.xlsx`	Chetverikov et al. (2017)	Provided upon request.	Yes

Computational requirements

Software Requirements

- Stata (code was last run with version 17)
  - `wid` (as of 2022-05-24)
  - line 12 of the program "`Table 1 Section 5/Code/CleanData.do`" can be uncommented to install all dependencies locally and should be run once.
- Julia 1.6.0
  - `SpecialFunctions` (1.4.0)
  - `Plots` (1.11.2)
  - `PGFPlotsX` (1.4.1)
  - `QuadGK` (2.4.2)
  - `LaTeXStrings` (1.2.1)
  - `Optim` (1.3.0)
  - `XLSX` (0.7.9)
  - `DataFrames` (1.1.0)
  - `CSV` (0.8.4)
- R 3.6.2
  - `haven` (2.3.1)
  - `readxl` (1.3.1)
  - `nleqslv` (3.3.2)
  - `writexl` (1.4.0)

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard (2022) desktop machine:

☑ <10 minutes
☐ 10-60 minutes
☐ 1-8 hours

☐ 8-24 hours
☐ 1-3 days
☐ 3-14 days
☐ > 14 days
☐ Not feasible to run on a desktop machine, as described below.

Details The code was last run on an **Intel-based core i7-8565U laptop with 24.0 GB of RAM and Windows 11 Pro operating system**.

Description of programs/code

- Programs in `Table 1 Section 5/Code` will reformat the datasets referenced above and run the code to replicate Table 1.
  - The file `Table 1 Section 5/Code/CleanData.do` reformats the datasets provided by the authors and downloads the WID.
  - The file `Table 1 Section 5/Code/Table1.R` takes the reformatted data, creates Figure 1, and returns the estimates reported in Table 1. It also calibrates the α and β used in section 6 to produce figure 2.
- Program `Figure 2 Section 6/Figure2a2b.jl` does the calculations to make Figure 2 and creates an excel file and a csv with the numbers.

License for Code

Instructions to Replicators

- For each of the following codes, we use the replication folder as the working directory. `Table 1 Section 5/Code/CleanData.do`, `Table 1 Section 5/Code/Table1.R`, and `Figure 2 Section 6/Figure2a2b.jl` each have code commented out that sets that working directory
- Run `Table 1 Section 5/Code/CleanData.do` with the replication folder as the working directory, uncommenting line 12 to download necessary package.
- Run `Table 1 Section 5/Code/Table1.R` with the replication folder as the working directory, downloading the necessary packages if you do not have them.
- Run `Figure 2 Section 6/Figure2a2b.jl` with the replication folder as the working directory, downloading the necessary packages if you do not have them.

Details

- `Table 1 Section 5/Code/CleanData.do`: will format the data provided by the authors of Guner et al. (2014), Acemoglu and Restrepo (2017), and Chetverikov et al. (2016). It also downloads the data from WID and formats it into how we will use it.

  - This code requires `Table 1 Section 5/RawData/fig1 point estimates for arnaud.xlsx`, `Table 1 Section 5/RawData/gkv_Figures_RED_USTaxes_2013.xls`, and `Table 1 Section 5/RawData/results_by_percentile.dta`
  - Uncommenting line 12 installs the wid stata package to download from WID.
  - Uncommenting line 13 and changing the path automatically sets the working directory
- `Table 1 Section 5/Code/Table1.R`:
  - Uses the data formatted by `Table 1 Section 5/Code/CleanData.do` along with the labor share data `Table 1 Section 5/RawData/labor_share.xlsx`.
  - Creates Figure 1.
  - Calculates the taxes reported in Table 1. The names of the variables map descriptively into those reported in the table with t at the beginning followed by trade or robot, and then a description. Outputs `Table 1 Section 5/Table1.xlsx`
  - Calibrates the α and β used in section 6. These are the variables alphajoint and betajoint.
  - Uncommenting line 4 and changing the path automatically sets the working directory
- `Figure 2 Section 6/Figure2a2b.jl`: Creates figures 2(a) and 2(b).
  - Outputs four files: a tex file for figure 2a, a tex file for figure 2b, an excel file showing the raw data in those figures and the results using different formulas, and a CSV file with the same thing.
  - Uncommenting line 5 and changing the path automatically sets the working directory

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 Number	Output file	Note
Figure 1(a)	Table 1 Section 5/Code/Table1.R	45	Table 1 Section 5/figure1a.png	
Figure 1(b)	Table 1 Section 5/Code/Table1.R	49	Table 1 Section 5/figure1b.png	
Table 1	Table 1 Section 5/Code/Table1.R	272	Table 1 Section 5/Table1.xlsx	

Figure/Table #	Program	Line Number	Output file	Note
Figure 2(a)	Figure 2 Section 6/Figure2a2b.jl	217	Figure 2 Section 6/Figure2a.tex	
Figure 2(b)	Figure 2 Section 6/Figure2a2b.jl	228	Figure 2 Section 6/Figure2b.tex	

The code does not reproduce tables E.1 or E.2 as those simply display variables from Table 1 Section 5/Code/table1.R. The next table reports the mapping between variables in Tables E.1 and E.2 variables in the code.

Table Variable	Code Variable
E.1.A $\{\tau(q)\}$	TempIncome$StatTax
E.1.A $\{\bar{s}(q)\}$	Income_Shares_decile$labor_share
E.1.A $\{\frac{\Delta \log \omega(q)}{\Delta \log y_m^*} \times 100\}$	(AR_estimates$'Long differences'[i] - AR_estimates$'Long differences'[i-1]) / (log(Income_Shares_decile$labor_share[i]) - log(Income_Shares_decile$labor_share[i-1]))
E.1.C $\{\tau(q)\}$	TempIncome$AveTax
E.1.D $\{\frac{\Delta \log \omega(q)}{\Delta \log y_m^*} \times 100\}$	(predict(Robots_eta_1)[i] - predict(Robots_eta_1)[i-1]) / (log(Income_Shares_decile$labor_share[i]) - log(Income_Shares_decile$labor_share[i-1]))
E.1.E $\{\frac{\Delta \log \omega(q)}{\Delta \log y_m^*} \times 100\}$	predict(Robots_domega_1)
E.2.A $\{\tau(q)\}$	TempIncome$StatTax
E.2.A $\{\bar{s}(q)\}$	Income_Shares$labor_share
E.2.A $\{\frac{\Delta \log \omega(q)}{\Delta \log y_m^*} \times 100\}$	(CLP_estimates$estimate[i] - CLP_estimates$estimate[i-1]) / (log(Income_Shares$labor_share[i]) - log(Income_Shares$labor_share[i-1]))
E.2.C $\{\tau(q)\}$	TempIncome$AveTax
E.2.D $\{\frac{\Delta \log \omega(q)}{\Delta \log y_m^*} \times 100\}$	(predict(Trade_eta_1)[i] - predict(Trade_eta_1)[i-1]) / (log(Income_Shares$labor_share[i]) - log(Income_Shares$labor_share[i-1]))
E.2.E $\{\frac{\Delta \log \omega(q)}{\Delta \log y_m^*} \times 100\}$	predict(Trade_domega_1)

References

[1] Acemoglu, Daron and Pascual Restrepo, “Robots and Jobs: Evidence from US Local Labor Markets,” *NBER Work. Pap. 23285, 2017*

[2] Chetverikov, Denis, Bradley Larsen, and Christopher John Palmer, “IV Quantile Regession for Group-Level Treatments, with an Application to the Effects of Trade on the Distribution of Wage,” *Econometrica*, 2016, *84* (2), 809-833.

[3] Giandrea, Michael D. and Shawn Sprague, "Estimating the U.S. labor share," Monthly Labor Review, U.S. Bureau of Labor Statistics, February 2017, https://doi.org/10.21916/mlr.2017.7

[4] Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura, “Income Taxation of U.S. Households: Facts and Parametric Estimates,” *Review of Economic Dynamics*, 2014, *17* (4), 559-581.