

Replication Package for *Declining Search Frictions, Unemployment and Self-Employment* (by Piotr Denderski and Florian Sniekers)

1 Overview

The codes provided in the package use Stata, Python, and Matlab. We process the data on:

- the composition of the labour force from ILO (n.d.),
- labour market flows from Donovan, Lu, and Schoellman (2023),
- broadband Internet diffusion from ITU (2019),
- GDP per capita, infrastructure and institutional variation (taxes, replacement rates and the size of public sector) from OECD (n.d.),
- infrastructure from OECD (1999),
- prices from the Penn World Tables, (Feenstra, Inklaar, and Timmer, 2021).

The codes are in the main folder of the package. There are four sub-folders. In `/data`, we provide data inputs and outputs (analysis files) in their formats as used by us and in ASCII formats. The folder `/tempFolder` is used by the replication package codes to store and update data files that are created along the way. The codes save the results (tables, figures, and text files) in the `/outputs` folder. We provide information on output/analysis files' contents in the `/codebooks` folder.

2 Data Availability and Provenance Statements

2.1 Statement about Rights

We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

2.2 Summary of Availability

Some data **cannot be made** publicly available. Some data have been updated to newer versions since the submission of the first draft of the paper.

Table 1: Data Sources

Data Name	Files	Provided	Citation
<i>Labour Force Composition</i>	employment_25_more_ilo_estimates.dta	Yes	ILO (n.d.)
	unemployment_25_more_ilo_estimates.dta	Yes	
<i>Self-Employment Composition</i>	EMP_TEMP_SEX_STE_ECO_NB_A-filtered.dta	Yes	
<i>Public Sector</i>	publicSector.dta	Yes	
<i>Broadband Diffusion</i>	broadbandOECDCompleteData.xlsx	No	ITU (2019)
<i>Labour Market Flows</i>	LFSdata_aggregate.xlsx	Yes	Donovan et al. (2023)
<i>GDP</i>	gdpOECDcompleteData.xlsx	Yes	OECD (n.d.)
<i>Replacement Rate</i>	replacementOECDlarge.csv	Yes	
	taxWedgeSingle.xlsx	Yes	
<i>Taxes</i>	taxWedgeCouple.xlsx	Yes	
<i>Infrastructure</i>	initialConditionsOECDCompleteData.xlsx	Yes	OECD (1999)
<i>Prices</i>	PennWorldTables_prices.xlsx	Yes	Feenstra et al. (2021)

All data files, if provided, are located in the /data folder.

2.3 Details on Each Data Source

Note, all provided input files are also made available in non-proprietary ASCII format. These files can be found in /data/ascii_input_files.

Labour Force Composition We accessed the data on ILO estimates total employment, self-employment, and unemployment in March 2020. We make the version of the data we used in the replication package available in files employment_25_more_ilo_estimates.dta and unemployment_25_more_ilo_estimates.dta. The labels in the column `classif2_label` (identical name in both files) are self-explanatory hence we do not provide a dictionary here.

The data on the most current version of the ILO estimates can be accessed at <https://ilostat.ilo.org/data/> - click on *Resources*, then *Data Tools* and *Data Explorer*. In the search bar look for either *Employment by age and status in employment – ILO modelled estimates* (for data on total employment and self-employment) or *Unemployment by sex and age – ILO modelled estimates*. In both cases, choose age 25+, and for unemployment, select *Total* for sex.

Self-Employment Composition We accessed the data on self-employment per sector in June 2022 and make it available in the EMP_TEMP_SEX_STE_ECO_NB_A-filtered.dta file. The labels in the `classif2_label` are self-explanatory hence we do not provide a dictionary here. To download the current version of the data directly from the ILO website, follow the steps as in the case of Labour Force Composition data but choose *Employment by sex, status in employment and economic activity* instead. Select *Total* for sex and narrow it down to self-employed and own-account workers.

Public Sector We accessed the data on the size of the public sector employment in March 2020 and make it available in the publicSector.dta file. The data contents are self-explanatory. To download the current version of the data directly from the ILO website, follow the steps as

in the case of *Labour Force Composition* data but choose *Employment by sex, age, and institutional sector* and select *Total* for sex, *Public* for sector and *25+* for age.

Broadband Diffusion The file `broadbandOECDCompleteData.xlsx` contains data series on the diffusion of broadband Internet from the World Telecommunication/ICT Indicators Database (ITU, 2019). For details on this data, see Section C.1 in the Online Appendix.

Labour Market Flows We provide the data on labour market flows which is used to calibrate the job destruction parameter δ in the `LFSdata_aggregate.xlsx` file. To obtain the documentation for the contents of this file, go to <https://www.lfsdata.com/codes-and-documentation>.

GDP We have accessed the data on GDP in March 2020, and we provide this data in the file `gdpOECDCompleteData.xlsx`. This file contains a single table with GDP measured using the expenditure approach as GDP per capita in 2010 US Dollars, constant PPPs. The contents of the file are self-explanatory. Our variable precedes the change in the methodology introduced in 2019. To access the current version of the data, go to <https://stats.oecd.org/> and select *National Accounts*, then *Annual National Accounts, Main Aggregates* and *1. Gross domestic product (GDP)*.

Replacement Rates We accessed the data on replacement rates in March 2020 and make it available in the `replacementOECDlarge.csv`. The contents of this file are self-explanatory. To download the most recent version of this data, go to <https://stats.oecd.org/> and select *Social Protection and Well-Being*, then *Benefits, Taxes, Wages* and *Replacement Rates in Unemployment*.

Infrastructure We take the data on *TV penetration* and *Access Lines* in 1997 from the OECD Communications Outlook 1999, (OECD, 1999). We make the data available in the `initialConditionsOECDCompleteData.xlsx` file. The contents of the file are self-explanatory. To download/read the Communications Outlook, go to https://doi.org/10.1787/comms_outlook-1999-en. Registered users can then browse the Outlook online for free. Our data come from Tables 4.2 and 6.13.

Taxes We have accessed the data on taxes in March 2020, and we provide two files that contain the relevant information in the package. The file `taxWedgeSingle.xlsx` contains a table with data on the tax rate for a single person without children earning the average wage. The file `taxWedgeCouple.xlsx` contains a table with data on the tax rate for a one-earner married couple with average earnings with two children. To access the current version of the data, go to <https://stats.oecd.org/> and select *Public Sector, Taxation and Market Regulation*, then *Taxation, Taxing Wages* and *Comparative tables*.

Prices We accessed the PWT data in October 2021. At that time, the most current version was the 10.0 version. We provide the series on prices that we use in the paper in the file `PennWorldTables.prices.xlsx`. The labels in the *Preface* tab are self-explanatory. You

can also download the (full) data in the PWT 10.0 version from <https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt100?lang=en> by clicking on the *Excel* button. Note, the resulting file will have a different outline than the one we provided. The relevant columns in the file downloaded from the website are *AD* and *AT*, which contain the *p1_con* and the *p1_c* variables, respectively.

2.4 Data Output List

Having discussed and listed the input files, we list the output data files below. Note, all of the files which are provided in the package also have their non-proprietary, ASCII formats available. They can be found in `/data/ascii_output_files`. We do not list those separately here.

Table 2: Output Files

Data file	Notes	Provided
<code>/tempFiles/dataForIV.dta</code>	Contains proprietary series on Broadband	No
<code>/data/output_files/predictedBb.dta</code>	Contains series on <i>Predicted Broadband</i> .	Yes
<code>/data/output_files/SEbyEconActivity.dta</code>	Contains series on composition of <i>SE</i> . Needed for additional evidence in Appendix A.	Yes
<code>/data/output_files/dataForPaper.dta</code>	Contains the main data file. Source for results in Section 2.	Yes
<code>/data/output_files/dataForMatlab.xlsx</code>	Contains series on <i>SE</i> , <i>U</i> and <i>Replacement Rate</i> to update <i>b</i> . Used to back out search frictions parameters.	Yes
<code>/data/output_files/dataForMatlab.xlsx</code>	Contains series on λ , <i>E</i> and aggregate price level. Used to show the main mechanism of the model.	Yes
<code>/data/output_files/PWT_prices.dta</code>	Penn World Tables data.	Yes
<code>/data/output_files/lambdaESeriesForPriceReg.xlsx</code>	Series on λ and <i>E</i> .	Yes
<code>/data/output_files/dataForPaper.dta</code>	Combines data on λ and <i>E</i> and that on prices from <code>lambdaESeriesForPriceReg.xlsx</code> and <code>PWT_prices.dta</code>	Yes

The codebooks on relevant variables in the files listed here are provided in `/codebooks/codebook.xlsx`. The sheets' names correspond to those of the files, the contents of which they describe.

3 Computational Requirements

3.1 Software Requirements

- Stata (code was last run with version 18MP)
 - `estout` version 3.31 (as of 26-04-2022)
 - `xttest3` version 1.0.7,
 - the program “`0_installPackages.do`” will install all dependencies locally, and should be run once.

- Python 3.10.9 (packaged by Anaconda, MSC v.1916 64 bit (AMD64)). The codes were run with Jupyter Notebook 6.5.2.
 - pandas 1.5.3
 - numpy 1.23.5
 - matplotlib 3.7.0
 - seaborn 0.12.2
 - tabulate 0.8.10
 - IPython 8.10.0
 - stata_setup 0.1.3
 - re 2.2.1
 - openpyxl 3.0.10
 - ruamel-yaml 0.17.21
 - jinja2 3.1.2
 - pywinpty 2.0.10
 - the program “File_optional_checkAndInstallPackages.ipynb” checks the versions of these packages and, if missing, attempts installing them. We advise running this program only if the execution of the replication codes fails.
- Matlab (code was run with Matlab Release 2023a)
 - Optimization Toolbox

3.2 Controlled Randomness

No random numbers were used.

3.3 Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard (2023) desktop machine: 10-20 minutes.

Details The code was last run on a **2-core Intel-based laptop with Windows 11**. The computation took 15 minutes.

3.4 Description of Programs/Code

- The file `File0_installPackages.do` installs the `estout` and `xttest3` packages in Stata if not already installed.
- The file `File1_broadbandIV.ipynb` takes the proprietary raw data on broadband Internet diffusion and creates the variable *Predicted Broadband*. There are two interim data files created in the `/tempFiles` directory at this step: `dataForIV.dta` and `predictedBb.dta`. This file saves the outputs of these analyses in the `/outputs` folder and formats them

- The file `File2.descriptiveStatistics.ipynb` combines the *Predicted Broadband* data with that on *Self-Employment, Unemployment, Own-Account Work, GDP, Tax Burden, Public Sector* and *Replacement Rate* and creates the `dataForPaper.dta` file in the `/tempFiles` folder. This file also creates tables with information on our sample and formats them in the `/outputs` folder.
- The file `File3.runBbEffectEstimations.ipynb` runs the estimation of the effect of broadband on *Self-Employment, Unemployment, Own-Account Work* and deepens this analysis with the sectoral decomposition of self-employment rates. It saves the outputs of these analyses in the `/outputs` folder and formats them. This file creates counter-factual series of interest and appends them to the `dataForPaper.dta` file. Then, it also creates the file `SEbyEconActivity.dta` in the `/tempFiles` folder and merges it with the `dataForPaper.dta` file.
- The file `File4.runNoBbCounterfactual.ipynb` runs the counterfactual analysis of no-broadband diffusion and saves and formats the outputs of these analyses in the `outputs` folder. It also saves the data needed for the exercise of backing out the search frictions parameters in the `dataForMatlab.xlsx` file in `tempFiles` folder.
- The file `File5.calculateSearchFrictions.m` calculates the search frictions parameters as implied by data and in the counterfactual scenario. It then saves and formats the outputs of these analyses in `outputs` folder. It also saves two files in the `tempFiles` folder: `lambdaEseriesForPriceReg.xlsx` which will be needed in the next step of the analysis of the effects of search friction parameters on prices and a file `quantNumbersInText.xlsx` for calculation of the relevant numbers describing the results of this step of the analysis in the main body of the paper.
- The file `File6.priceLevel.ipynb` provides evidence of the effect of model-implied search frictions parameters on the price level. It takes the *Prices* data and the `lambdaEseriesForPriceReg.xlsx` file and saves an analysis file `PWT_prices.dta` in `/tempFiles` folder. Finally, this file saves the outputs of these analyses in the `/outputs` folder and formats them.
- The file `File7.AnalyticalPlots.m` runs the numerical illustrations of selected analytical features of the model. It saves the results in the `/outputs` folder and formats them.
- The file `File_optional_checkAndInstallPackages.ipynb` checks the version of Python packages that we use and if these are missing or are obsolete, attempts installing them in required versions.

3.5 Instructions to Replicators

The only manual interventions required to run the code are concerned with launching Stata from Python. You need to manually update the variables `stataDirectory` with the path of the location of your Stata installation and the `stataVersion` to match the version of Stata you have installed. This has to be done in four files: `File1.broadbandIV.ipynb` (cell 2),

File2_descriptiveStatistics.ipynb (cell 2), File3_runBbEffectEstimations.ipynb (cell 2) and File6_priceLevel.ipynb (cell 2).

Then, the files should be run in a sequence. If you already have `estout` and `xttest3` installed in your Stata, start with `File1_broadbandIV.ipynb`, otherwise first run `File0_installPackages.do`. Having ran `File1_broadbandIV.ipynb`, run `File2_descriptiveStatistics.ipynb`, then `File3_runBbEffectEstimations.ipynb`, then `File4_runNoBbCounterfactual.ipynb`, next `File5_calculateSearchFrictions.m` followed by `File6_priceLevel.ipynb` and finally run `File7_AnalyticalPlots.m`. If you encounter a missing package error at the time of running any of the `.ipynb` files, proceed to running `File_optional_checkAndInstallPackages.ipynb`. However, it is unlikely that this will be necessary as all packages that we used are standard and are included in the Anaconda distribution.

3.6 List of Outputs and Programs

Table 3: Outputs and Programs: Paper

Output	Program	Files
Table 1	File2_descriptiveStatistics.ipynb, cell 12	Table1Paper.tex
Table 2	File3_runBbEffectEstimations.ipynb, cells 5 and 19	Table2Paper.tex
Table 3	File3_runBbEffectEstimations.ipynb, cells 6 and 19	Table3Paper.tex
Table 4	File6_priceLevel.ipynb, cells 11 and 18	Table4Paper.tex
Table A1	File3_runBbEffectEstimations.ipynb, cells 7 and 19	Table5AppendixA.tex
Table A2	File3_runBbEffectEstimations.ipynb, cells 15 and 19	Table6AppendixA.tex
Table A3	File3_runBbEffectEstimations.ipynb, cells 16 and 19	Table7AppendixA.tex
Table A4	File3_runBbEffectEstimations.ipynb, cells 17 and 19	Table8AppendixA.tex
Table A5	File3_runBbEffectEstimations.ipynb, cells 18 and 19	Table9AppendixA.tex
Figure 1	File4_runNoBbCounterfactual.ipynb cell 10	Figure1Paper.SE.pdf Figure1Paper.U.pdf
Figure 2	–	–
Figure 3	File7_AnalyticalPlots.m lines 54 - 74	Figure3Paper.pdf
Figure 4	File7_AnalyticalPlots.m lines 106 - 121	Figure4Paper.pdf
Figure 5	File5_calculateSearchFrictions.m lines 107-155	Figure5aPaper.pdf Figure5bPaper.pdf
Figure 6	File5_calculateSearchFrictions.m lines 88-100	Figure6Paper.pdf
Figure A1	File4_runNoBbCounterfactual.ipynb cell 13	Figure7AppendixA.SE.pdf Figure7AppendixA.U.pdf
Numbers in Section 2.2.	File4_runNoBbCounterfactual.ipynb, cell 9	cfEndBeginning.tex
Numbers in Section 5.1	File4_runNoBbCounterfactual.ipynb, cell 17	section5Calibration.tex
Numbers in Sections 5.2 and 5.3	File6_priceLevel.ipynb, cell 17	section5Numbers.tex
Numbers in Appendix A	File4_runNoBbCounterfactual.ipynb, cell 14	correlations.tex

Note: Figure 2 is a diagrammatic representation of the structure of the model. This figure does not require any calculations or data processing. All outputs are stored in the /outputs folder.

References

- DONOVAN, K., W. J. LU, AND T. SCHOELLMAN (2023): “Labor market dynamics and development - Data files,” *The Quarterly Journal of Economics*, data deposited at 1fsdata.com (last accessed: 6 April 2022).
- FEENSTRA, R. C., R. INKLAAR, AND M. P. TIMMER (2021): “Penn World Tables (PWT) version 10.0,” Tech. rep., Groningen Growth and Development Centre, doi:10.15141/S5Q94M (last accessed: 5 October 2021).

Table 4: Outputs and Programs: Online Appendix

Output	Program	Files
Table C1	File1_broadbandIV.ipynb, cells 7 and 8	Table10AppendixC.tex
Table C2	File1_broadbandIV.ipynb, cells 5 and 9	Table11AppendixC.tex
Table C3	File2_descriptiveStatistics.ipynb, cells 13 and 14	Table12AppendixC.tex
Table C4	File3_runBbEffectEstimations.ipynb, cells 8 and 19	Table13AppendixC.tex
Table C5	File3_runBbEffectEstimations.ipynb, cells 9 and 19	Table14AppendixC.tex
Table C6	File3_runBbEffectEstimations.ipynb, cells 10 and 19	Table15AppendixC.tex
Table C7	File6_priceLevel.ipynb, cells 9 and 18	Table16AppendixC.tex
Table C8	File6_priceLevel.ipynb, cells 11 and 18	Table17AppendixC.tex
Figure C1	File1_broadbandIV.ipynb, cell 6	Figure8AppendixC.pdf
Figure C2	File6_priceLevel.ipynb, cell 7	Figure9AppendixC.pdf
Figure C3	File6_priceLevel.ipynb, cell 7	Figure10AppendixC.pdf
Figure C4	File5_calculateSearchFrictions.m lines 107-155	Figure11aAppendixC.pdf Figure11bAppendixC.pdf
Figure C5	File5_calculateSearchFrictions.m lines 107-155	Figure12aAppendixC.pdf Figure12bAppendixC.pdf

Note: All outputs are stored in the /outputs folder.

ILO (n.d.): “Labour Force Statistics (LFS),” Tech. rep., ILOSTAT, <https://ilostat ilo.org/data/> (last accessed: 23 March 2020).

ITU (2019): “World Telecommunication/ICT Indicators Database 2018,” <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx>.

OECD (1999): *OECD Communications Outlook 1999*, doi:10.1787/comms.outlook-1999-en.

——— (n.d.): “Data warehouse,” doi:10.1787/data-00900-en (last accessed: 23 March 2020).