Replication files for: "Occupational exposure to capital-embodied technical change"

The replication package contains two main blocks. (a) Empirics, which includes Stata do-files that process and analyze the relevant series for this project, including the results in Sections 2 and 3 of the paper. To ease the usage of the dataset we also provide outputs from the data construction exercise that can be directly downloaded by the user in folder "empirics/finaldata". This data is also shared through our dedicated web at https://capitalbyoccupation.weebly.com/downloads.html. (b) Quantitative, which includes Matlab m-files which run the accounting exercises presented in Section 5 and 6 of the paper.

1 Empirics

The raw data subfolder includes public data from the following sources. Additional details on these sources and methodology can be found in the online appendix of the paper.

1. BEANIPA: Nipa Tables from U.S. Bureau of Economic Analysis (Retrieved on June 14, 2021).
2. Crosswalks: Occupation crosswalks from Deming, D. J. (2017). The growing importance of social skills in the labor market.
3. DOT-1977: 4th edition of the Dictionary of Occupational Titles, DOT. https://www.icpsr.umich.edu/web/ICPSR/studies/6100.
4. FED-PCEPI: Federal Reserve Bank of St. Louis (Retrieved on June 14, 2021).
5. FED-working-age-population: FRED, Federal Reserve Bank of St. Louis (Retrieved on January 2019).
6. IPUMS-CPS: Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren and Michael Westberry. Integrated Public Use Microdata Series, Current Population Survey: Version 9.0 [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D030.V9.0 (Retrieved on January 2019), Ruggles *et al.* (2019).
7. Quality-adjusted price of capital: DiCecio (2009) extending the methodology in Cummins and Violante (2002). Most updated data was kindly provided by the author on January 2019.
8. Tool-data-ONET: O*NET Data from: https://www.onetcenter.org/db_releases.html (see version 23.1 of the tools and technology module). (Retrieved on January 2019). The data includes information from the O*NET 23.1 Database by the U.S. Department of Labor, Employment and Training Administration (USDOL/ETA). Used under the CC BY 4.0 license. O*NET® is a trademark of USDOL/ETA.
9. WDI: World Bank. "Exports of goods and services (% of GDP)" World Development Indicators. The World Bank Group. (retrieved on June 2021).

10. Depreciation rates estimates from BEA by equipment type https://apps.bea.gov/national/pdf/BEA_depreciation_rates.pdf, (retrieved on January 2019).

All the data used in the analysis is public and it is contained in the rawdata folder.

Instructions. After downloading the replication folder, researchers should open the do file
.../empirics/dofiles/[0]Master-run.do file.
Researchers must specify the home directory in which the replication folder has been saved. This is the main program for the replication. The program installs the necessary packages and then calls the do files that build the intermediate and final datasets by processing the raw data. This do-file is organized in three blocks, one for the data-construction, one for tables, and one for figures. Each block can be run independently. Results shown in the paper and in the appendix are stored under */empirics/results/*

Packages. All stata packages needed are installed by [0]Master-run.do file (lines 34-35): weibullfit

Hardware and duration: We use a iMac Apple M1 chip with 8-core CPU 16GB memory and the code takes less than 6 minutes to run.

Finally, we provide a dedicated folder "*empirics/toolsDOT*" that contains the python routines used to extract the historical tool requirements of different occupations from the text of the DOT.

Instructions. Can be found in a dedicated *readMe* file, including information on how to create the environment and install libraries.

Libraries. The libraries are listed under *requirements.txt* in the folder above and can be automatically loaded following the instructions in *readMe*.

Hardware and duration: We use SciNet (U of T Supercomputing Centre), 1 nodes with 40 Intel Skylake cores at 2.4GHz, 202 GB per node. The code takes approximately 5 hours to run.

2 Quantitative

The files in the folder replication_quantitative produce the main outcomes of the quantitative analysis in Sections 4.2, 5, and 6 – that is, model calibration, counterfactuals, alternative exercises, and predictions, and produce corresponding tables (in *tables_paper.xslx*, sheet *Tables_paper*) and figures (numbered as in the paper), saved in the folder *quantitative/results*. All files are coded for Matlab and saved in the folder *quantitative/codes*. The file the runs all the above mentioned exercises is *A_main.m*. This file calls sub-routines which can be run separately and are described below.

Data preparation. The main file that loads and prepares the data generated in the *Empirics* is *AA_data.m*; this program calls the following programs:

step 1: *estimates_load.m* – loads the estimates of the elasticity of substitution between

capital and labor, σ_o, the tail of the Frechet distribution, θ, and the elasticity of substitution across capital goods, ϕ.

step 2: *data_load.m* – loads the occupational data on capital, employment, and their rental rates at both annual and decennial frequency and saves these data in the files $data\_info\_yby.mat$, $data\_main\_yby.mat$ (annual), $data\_info.mat$, $data\_main.mat$ (decennial).

step 3: *data_derive.m* – derives statistics on the data loaded in step 2.

step 4: *figures_appendix.m*– generates Figures BI and BII.

Estimation or the elasticity of substitution across occupational output. The main file for the estimation of the elasticity of substitution across occupational goods, ρ, is *AA_estimation_rho.m*; this program calls the following programs:

step 1: *sel_elast.m* and *data_for_model.m* – computes the statistics needed in the exercise (annual frequency).

step 2: *model_solve.m* – runs the calibration of the model. Note 1: The notation used in the code separates the matrix T_{oh} in two components, T_h and z_{oh}, and calls the occupational price per efficiency unit of labor $\tilde{\varkappa}_o$ instead of λ_o^n, so that, for example, wages can be written as $w_h = (T_h \sum_o (z_{oh}\tilde{\varkappa}_u)^\theta)^{1/\theta}$ instead of $w_h = (\sum_o (T_{oh}\tilde{\varkappa}_u)^\theta)^{1/\theta}$. Note 2: Despite not of relevance for the specifics of the calibration in the paper, the code allows for the case where the occupational production technology is Cobb-Douglas in capital and labor ($\sigma = 1$).

step 3: *est_rho.m* – runs the estimation of ρ and saves the IV estimates in the $rho.mat$.

Baseline exercise. The main file that runs the main counterfactual and prediction exercises is *AA_baseline_model.m*; this program calls the following programs:

step 1: Model fit:

- step 1.1: *sel_elast.m* and *data_for_model.m* – computes the statistics needed in the exercise (decennial frequency).
- step 1.2: *model_solve.m* – runs the calibration of the model.
- step 1.3: *model_fit.m* – runs Table B.IV.

step 2: Counterfactuals:

- step 2.1: *z_decomposition.m* – runs the decomposition of the T_{oht} matrix for the counterfactuals.
- step 2.3: *main_counterfactuals.m* – runs the main counterfactual exercise outlined in Section 5.1.
- step 2.3: *counterfactual_stats.m* and *counterfactual_figures.m* – runs column *CETC/baseline* of Tables 1, 2, columns *CETC/baseline, demand, demographics, composition, CA* of Table B.IV, Table B.V, and Figures 5 and 6.

step 3: Prediction:

step 3.1: *sel_elast.m* and *data_for_model.m* – computes the statistics needed in the exercise (annual frequency).

step 3.2: *model_solve.m* – runs the calibration of the model at yearly frequency.

step 3.2: *prediction.m* – runs the prediction exercise and Figure 7.

Alternative exercises. The main file that runs the alternative exercises is *AA_alternative_exercise.m*; this program calls a subset of the programs in the baseline exercise and outputs columns *CETC/identical elasticities* and *CETC/identical CETC* of Tables 1, 2, and B.VI.

Multiple capital goods. The main file that runs the main counterfactual in the model with multiple capital goods is *AA_model_multiple_equipment.m*; this program calls a subset of the programs in the baseline exercise and outputs columns *CETC/computers*, *CETC/communication equipment* and *CETC/software* of Table 3.

The remaining Matlab files in the folder are sub-routines.

Packages. Users must have installed Matlab's Optimization Toolbox. The code was run on Matlab R2022b.

Hardware and duration: MacBook Pro M1 (2020, 8CPU) 816.04 seconds. Time info is saved in *A-main* file.

References

Cummins, J. G. and Violante, G. L. (2002). Investment-Specific Technical Change in the US (1947-2000): Measurement and Macroeconomic Consequences. *Review of Economic Dynamics*, **5** (2), 243–284.

DiCecio, R. (2009). Sticky wages and sectoral labor comovement. *Journal of Economic Dynamics and Control*, **33** (3), 538 – 553.

DOT (). *Dictionary of Occupational Titles (DOT): Revised Fourth Edition, 1991 [dataset]*. Tech. Rep. 10.3886/ICPSR06100.v1, United States Department of Labor. United States Employment Service, and the North Carolina Occupational Analysis Field Center, https://www.icpsr.umich.edu/web/ICPSR/studies/6100.

O*NET (). *ONET Version 23.1 [dataset]*. Used under the CC BY 4.0 license Version 23.1, U.S. Department of Labor, Employment and Training Administration, https://www.onetcenter.org/db_releases.html.

Ruggles, S., Flood, S., Goeken, R., Schouweiler, M. and Sobek, M. (2019). *IPUMS USA [dataset]*. Tech. Rep. Version V9, Minneapolis, MN: IPUMS, https://doi.org/10.18128/D030.V9.0.