

Skill-Biased Structural Change

Readme File for the Data Appendix

Francisco J. Buera, Joseph P. Kaboski, Richard Rogerson, and Juan I. Vizcaino

Folder: HS Sector Definition

This folder contains the files needed to compute the shares of high-skill labor by industry under different employment measures.

Files

- HS_Shares_by_Industry.do
This .do File Computes the Share of High-Skill Labor's Compensation, Employment, and Hours Worked by Industry, for the US, between 1977 and 2005. These shares are used to define the High-Skill Intensive Sector in Section 2.2 of the paper. It is also used to compute the supporting numbers provided in Section 2 of the Online Appendix.
 - *Inputs:* the 2013 version of the World KLEMS' labor input file (usa_wk_apr_2013_labour.xlsx).
 - *Output:* a .dta file (HS_labor_shares_1977_2005.dta) containing the time series and the time average of the high-skill labor shares for the 31 Industries in the World KLEMS database. These data are copied to the Excel spreadsheet (HS_comp_emp_hours_share_WK.xlsx) where further analysis is performed, including summary tables and charts, and where Table 2 in Section 2 of the Online Appendix is produced.

Folder: Figures1&2

Subfolder: Figure1

This folder contains the files needed to produce Figure 1 in section 2.3 of the paper.

Files

- Figure1.do
Figure1 plots on the left panel (right panel) the share of Labor Compensation (Value-Added) in the high skill-intensive and low-skill intensive sectors vs. real GDP per-capita at constant 2011 national prices (in US\$). Both series have been demeaned to take out country-specific fixed effects. We consider 16 countries over 1970-2005. Countries are selected using the criterion in

Buera-Kaboski 2012: they should have an income per-capita of at least \$9,200 in Geary-Khamis 1990 international dollars in 1970. The countries are Austria, Australia, Belgium, Denmark, Spain, France, United Kingdom, Germany, Greece, Ireland, Italy, Japan, Luxembourg, The Netherlands, and the U.S.

- *Inputs*: one Excel spreadsheet per country, from EUKlems database (<http://www.euklems.net/>), November 2009 release, revised in June 2010.
- *Output*: a .dta file with the data final data to produce Figure1 (figure1.dta) and two charts (figure1.png; figure1.gph). The charts are stored in the subfolder Figures. The .do file also produces Figure1 for the two broader definitions of the high-skill intensive sector considered in the Online Appendix and computes the R2 of the regressions in the figure with and without controlling for country-specific fixed effects.

Subfolder: Figure2

This folder contains the files needed to produce Figure 2 in section 2.3 of the paper.

Files

- Figure2.do
Figure2 plots the relative price of High-Skill Intensive to Low-Skill Intensive sectors. It is normalized such that, for each country, the index equals 100 in 1995. Both series have been demeaned to take out country-specific fixed effects. We consider 16 countries over 1970-2005. Countries are selected using the criterion in Buera-Kaboski 2012: they should have an income per capita of at least \$9,200 in Geary-Khamis 1990 international dollars in 1970. Countries are Austria, Australia, Belgium, Denmark, Spain, France, United Kingdom, Germany, Greece, Ireland, Italy, Japan, Luxembourg, The Netherlands, and the U.S.
 - *Inputs*: one Excel spreadsheet per country, from EUKlems database (<http://www.euklems.net/>), November 2009 release, revised in June 2010.
 - *Output*: a .dta file with the data final data to produce Figure2 (figure1.dta) and two charts (figure2.png; figure2.gph). The charts are stored in the subfolder Figures. The .do file also produces Figure2 for the two broader definitions of the high-skill intensive sector considered in the Online Appendix and computes the R2 of the regressions in the figure with and without controlling for country-specific fixed effects.

Folder: Table 1

The exercises performed through the files in this folder constitute the analysis described in Table 1 of Section 2.4.

Files

- CEXdeomographicdata.dta
The demographic data by household which is merged with the value-added in the expenditure bundle data in the Stats .do file 'Table1regressions.do'

- CEXtoVAdocumentation.xlsx
Contains all of the necessary matrices and cross-walks to reproduce the analysis in Section 2.2, including Table 1, as explained in the file 'Mapping Instructions.docx'. The explanation of each worksheet is given in the first sheet entitled 'ReadMe'.
- *CombinedVADataforRegressions.dta
Contains value-added in the CEX consumption bundle of each household in the CEX. Based on the CEX data from 2012.
- Mapping Instruction.docx
Instructions for mapping U.S. industrial skill intensity data to U.S. CEX consumption data through the U.S. input-output structure.
- MatlabManipulations.m
Matlab matrix manipulation codes for constructing the matrix 'dataforCEXmerge' in CEXtoVAdocumentation.xlsx
- Table1regressions.do
Creates the regression results in Table 1, as well as the results without controls. It calls the data files: 'CombinedVADataforRegressions.dta' and 'CEXdemographicdata.dta', and outputs the results to the ASCII log file 'Table1Results.log'

Folder: Data_for_Calibration

This folder contains the files needed to compute the time series used to calibrate the model, as described in Section 4.

Files

- data_for_calibration.do
This .do file produces the time series (1977-2005) for the variables used to calibrate the model. Time series calculations are for value added, high-skill labor income share, value added prices, and labor productivity by sector, and wages by labor type for the U.S.
 - *Inputs:*
 - * The 2013 version of the World KLEMS database for the US (usa_wk_apr_2013.xlsx): used for VA shares by sector, the chained relative price index, and the labor productivity indices.
 - * The 2013 version of the World KLEMS labor input file (usa_wk_apr_2013_labour.xlsx): the data are bridged to produce time-consistent time series, as suggested in the front cover of the file. It is used to produce compensation share of the high- and the low-skill intensive sectors, the skill intensity by sector, and the skill-premium.
 - * Penn World Table 9.0: (pwt90.dta) used to compute GDP growth for the US.
 - *Output:* an Excel spreadsheet containing the series used in the calibration (Data_for_Calibration.xlsx).

Folder: Shift_Share_KM_Methodology

In this folder, we perform a series of shift-share analyses following the methodology proposed in Katz and Murphy 1992. The results are presented in Section 6.1 of the paper, and more extensively in Section 1.4 of the Online Appendix.

To facilitate the comparison between measurement procedures and datasets, the Excel spreadsheet (Dem_shift_calculations_with_data.xlsx) in the folder takes the series of labor inputs measured in KM and BKR efficiency units and computes the between-industry demand shifts and their contribution to the total change in relative demand in favor of high-skill labor. Calculations and data for industries aggregated at the two-sector level are provided using World KLEMS and EU KLEMS data and for KM and BKR efficiency units. Calculations and data using Value Added as a measure of sector size are also provided.

Subfolder: IPUMS

In this subfolder, we compute the between-industry relative demand shift in favor of high-skill labor using IPUMS data for the US, from 1979 through 1989 (years 1980 and 1990 in the Census).

Subfolder Age:

computes the between industry demand shift using Age as grouping criterion for workers.

Files

- Demand_Shift_KM_IPUMS_E_KM_Age.do
This .do file computes the between industry relative demand shift in favor of high-skill workers, used as an input to compute the between industry demand contribution to the total change in relative demand of high-skill workers.
 - *Inputs:* IPUMS USA .dta file for 1980 and 1990. The data file is also available upon request or can be obtained from IPUMS US. See the .do file for variables needed. We do not provide the .dta file here because of its weight. We do provide a 5 percent sample of the data needed (usa_80_90_5pct_sample.dta), but to obtain the exact numbers the full dataset is needed.
 - *Output:* The final result is the number presented in row (i) of Column B in Table 1 of the Online Appendix.

Subfolder Experience:

computes the between industry demand shift using Age as grouping criterion for workers.

Files

- Demand_Shift_KM_IPUMS_E_KM_Experience.do
This .do file computes the between industry relative demand shift in favor of high-skill workers, used as an input to compute the between industry demand contribution to the total change in relative demand of high-skill workers.

- *Inputs:* IPUMS USA .dta file for 1980 and 1990. The data file is also available upon request or can be obtained from IPUMS US. See the .do file for variables needed. We do not provide the .dta file here because of its weight. We do provide a 5 percent sample of the data needed (usa_80_90_5pct_sample.dta), but to obtain the exact numbers the full dataset is needed.
- *Output:* The final result is the number presented in row (i) of Column C in Table 1 of the Online Appendix.

Subfolder: World KLEMS

In this subfolder, we compute the between-industry relative demand shift in favor of high-skill labor using World KLEMS data for the US. The between-industry demand shifts are computed measuring human capital in BKRv and KM efficiency units, aggregating industries at the 2 or the 31 sector level, and for different periods.

Files

- Shift_Share_KM_WK2013_Eff_BKRv_2Sectors.do
Computes the between-industry relative demand shift in favor of high-skill labor using World KLEMS data for the US, measuring human capital in BKRv efficiency units and aggregating industries at the two-sector level.
 - *Inputs:* The 2013 version of the World KLEMS labor input file (usa_wk_apr_2013_labour.xlsx). The data are bridged to produce time-consistent time series, as suggested in the front cover of the file.
 - *Output:* The final result is the number presented in row (i) of Columns F (1980-1990) and G (1977-2005) in Table 1 of the Online Appendix.
- Shift_Share_KM_WK2013_Eff_BKRv_31Sectors.do
Computes the between-industry relative demand shift in favor of high-skill labor using World KLEMS data for the US, measuring human capital in BKRv efficiency units and aggregating industries at the 31 sector level.
 - *Inputs:* The 2013 version of the World KLEMS labor input file (usa_wk_apr_2013_labour.xlsx). The data are bridged to produce time-consistent time series, as suggested in the front cover of the file.
 - *Output:* The final result is the number presented in row (i) of Column E in Table 1 of the Online Appendix.
- Shift_Share_KM_WK2013_Eff_KM_2Sectors.do
Computes the between-industry relative demand shift in favor of high-skill labor using World KLEMS data for the US, measuring human capital in KM efficiency units and aggregating industries at the two-sector level.
 - *Inputs:* The 2013 version of the World KLEMS labor input file (usa_wk_apr_2013_labour.xlsx). The data are bridged to produce time-consistent time series, as suggested in the front cover of the file.

- *Output:* The final result is the number presented in row (i) of Table 7 of the paper and in column H (1977-2005) of Table 1 of the Online Appendix. These two are the same number.
- Shift_Share_KM_WK2013_Eff_KM_31Sectors.do
Computes the between-industry relative demand shift in favor of high-skill labor using World KLEMS data for the US, measuring human capital in KM efficiency units and aggregating industries at the 31 sector level.
 - *Inputs:* The 2013 version of the World KLEMS labor input file (usa_wk_apr_2013_labour.xlsx). The data are bridged to produce time-consistent time series, as suggested in the front cover of the file.
 - *Output:* The final result is the numbers presented in Columns D (1979-1989) of the Online Appendix.

Subfolder: EUKLEMS

In this subfolder, we compute the between-industry relative demand shift in favor of high-skill labor using EUKLEMS data for the US. The between-industry demand shifts are computed measuring human capital in BKR and KM efficiency units, aggregating industries at the 2 sector level, and for different periods.

Files

- Shift_Share_KM_EUKLEMS_2Sectors.do
Computes the between-industry relative demand shift in favor of high-skill labor using World KLEMS data for the US, measuring human capital in BKR and KM efficiency units and aggregating industries at the two-sector level.
 - * *Inputs:* The EUKLEMS labor input file for the US version 08I (usa-naics_labour_input_08I.xls) and the EUKLEMS output file for the US version 09I (usa-naics_output_09I.xls).
 - * *Output:* The final result is the numbers presented in Columns I and J (1977-2005) for BKR and KM efficiency units, respectively) of the Online Appendix.

Folder: Sensitivity

This folder contains the files for the sensitivity exercises in Section 7 of the paper. In particular, we provide the files for Subsection 7.1 and 7.2 in the paper.

Subfolder: Consumption_Investment_VA

This folder contains the files used to compute the evolution of Value Added in producer's prices that is generated by the final expenditure on consumption and investment in producer's prices for the US. Data availability allows us to do this exercise for five years: 2005, 2002, 1997, 1992, 1987, 1982, 1977 with five years between them.

We download Input-Output data from the BEA and process them in the Matlab .m file for the corresponding year (IOprocessingYYYY.m). To construct the shares of the value added in producer's prices

that are generated by the final expenditure in consumption and investment, we follow the methodology in Herrendorf, Rogerson, and Valentinyi (2013). Specifically, we follow Section B.2 of the paper's online appendix.

- *Inputs*: the I-O tables for the US at the summary level. We use the Industry by Commodity Total Requirements Table, the Input-Output Use Table, and the Input-Output Make Table at the summary level for 1977, 1982, 1987, 1992, 1992, 2002, and 2005.
- *Output*: a series of Value Added shares for the high-skill intensive sector as well as the series of value added shares that are due to consumption and investment for the high-skill intensive sector for 1977, 1982, 1987, 1992, 1992, 2002, and 2005. Data is interpolated when a full-time series is needed.

Subfolder: Trade

This folder contains the Excel file (US Trade Data.xlsx). In this file we provide the calculations and adjustment data series which feeds into our alternative simulations evaluating the impact of trade and its relationship with our decomposition of the growth in the skill premium.

- *Inputs*: Data for the U.S. Trade in Goods and Services, Balance of Payments (BOP) Basis, for 1960 to 2014 from the U.S. Census Bureau. Disaggregated data for the US on Trade in Services by Type of Service, for 1999 to 2014, from the BEA, and US GDP data for 1960 to 2014 from the BEA.
- *Output*: A time series for net exports for the high- and the low-skill intensive sector for 1977 to 2005.

Folder: Cross-Country

This folder contains the files needed to produce the results in Section 8 of the paper and Section 5 of the Online Appendix.

Subfolder: Cross_Country_Calibration_Targets

In this folder, we produce the data series used to calibrate the model and obtain the results presented in Section 8 of the paper.

Files

- CC_cali_targets.do
This .do file produces times series calculations for VA for high-skill intensive and low-skill intensive industries, compensation shares of high-skill labor in high and low-skill industries, wage premium, relative prices of high-skill to low-skill industries, compensation shares of high-skill and low-skill industries, and GDP per capita for 11 countries. Countries are the subset of those in Figure 1 (or Figure 2) that have labor input files in the EUKLEMS database. (Austria, Australia, Belgium, Denmark, Spain, Germany, Italy, Japan, The Netherlands, United Kingdom, and the United States.

- *Inputs*: A series of Excel spreadsheets from the EUKLEMS labor input (XXX_labour_input_08l.xls) and basic files (XXX_output_09l.xls) for the countries above mentioned, contained in the folder LI_files and basic_files_09. Internationally comparable PPP for value-added (double deflated) for each country for the year 1997. This comes from the 'Benchmark_1997.xls' EUKLEMS data file entitled 'Output, input and productivity levels, 1997'.
- *Output*: an Excel spreadsheet (CrossCountry_cal_targets_BKRV.xlsx) containing the time series for the variables and countries above mentioned.

Subfolder: Cross-Country_Model_Fit

In this folder, we produce Figures 7 and 8 in Section 5 of the Online Appendix.

Files

- cross_country_figures.do
Produces Figures 7 and 8 in Section 5 the Online Appendix. These figures measure the ability of the model to fit the actual series for the skill premium and the share of the high-skill intensive sector in the countries mentioned above.
 - *Inputs*: A series of model-simulated and actual data provided in the file (model_fit_crosscountry.xlsx) and exported to the file (cross_country_model_fit.dta).
 - *Output*: Figures 7 and 8 in Section 5 of the Online Appendix.