# Readme for the "Calibration" folder for: "Automation and the Future of Work: Assessing the Role of Labor Flexibility"

Michele Fornino*
Andrea Manera†

This Readme illustrates how we obtained the target statistics of interest described in the Calibration appendix of our paper. Following the steps outlined below will produce the file "Statistics.csv" that is needed as an input for the calibration exercise, specifically for the function "setParametersGE".

## 1 Folder Structure and Master File

The main "Calibration" folder contains three subfolders which have all the datasets and codes needed to replicate the data used for our calibration:

- "dataRaw" contains the data from the sources detailed in the appendix "Data Sources and Raw Files";
- "bin" contains the Stata do-files needed to build our statistics. The **master file** is "runGenerateAllData.do" which executes all the other chunks of code contained in the subfolder;
- "out" collects the output datasets and the file "Statistics.csv".

An important note is in order. To correctly run the codes, users should modify the global variable "$masterpath" in the master file "runGenerateAllData.do". The "masterpath" folder should contained the subfolders listed above in order for the code to run correctly. All files can also be run as stand-alone by uncommenting the preamble and setting the "masterpath" as needed. A more detailed description of the subfolders and files follows.

*MIT. Email: mfornino@mit.edu
†MIT. Email: manera@mit.edu

## 2 “dataRaw”

See the next section for a description of the input files for the Stata do-files. Here we just mention the source of the robot purchase price to mean wages of production-line employees, $p_R/w$, used in our benchmark calibration.

The series on robot prices comes from the first figure in Korus (2019) ("Industrial Robot Cost Decline"). The data on mean wages of production-line employees in the various reference years 2004, 2010, 2014 comes from the OES, and is contained in the respective subfolders of "BLSmeanwages". In each subfolder, there is a "xsls" file named according to the following pattern "national_MYYYY_dl". Upon opening, the file will show a highlighted line containing, in order, the wage in 2015 terms of all production occupations in manufacturing (column U), the price of robots in 2015 dollars from Korus (column V), and their ratio (column W). The wage in 2015 dollars is obtained using the series in "CPIAUCSL.csv".

## 3 “bin”

This folder contains the do-files needed to build our data. They are generously commented. The inputs, outputs and main steps of each files are as follows:

- “runThetaProd” computes the shares of production-line employees on value-added in the various IFR sectors. Inputs:
  - "CPIAUCSL.csv", the CPI series for all urban consumers from FRED;
  - "sic5811.dta", containing SIC-level data from the NBER-CES dataset;
  - "CrossWalkSICIFR.dta" is the crosswalk between SIC codes and IFR codes kindly provided by Daron Acemoglu and Pascual Restrepo, who used it in Acemoglu and Restrepo (2018).

  Outputs:
  - "ThetaProdTsIfr.dta" containing the full time series of labor shares (not used in our Matlab code);
  - "ThetaProdIFR.dta" containing the average share of production-line employees for each IFR sector before 1980, to use as a value for the DRS parameter $\theta$.

  Summary of the procedure: the value added data from the NBER-CES is matched to the IFR code using the crosswalk. The parameter $\theta$ is obtained as the average of the trend component of the share of production-line employees on value added before 1980.

- “runProdLineEmp” computes the number of production-line employees in the various IFR sectors in 1989. This data is currently in our calibration, but it ia needed for the file "Statistics.csv" Inputs:
  - "BLSProdLineEmp/mf89d3.csv", containing the number of employees by OES occupation code in 3-digit SIC sectors in 1989;
  - "CrossWalkSICIFR.dta" is the crosswalk between SIC codes and IFR codes kindly provided by Daron Acemoglu and Pascual Restrepo, who used it in Acemoglu and Restrepo (2018);
  - "dataRaw/BLSProdLineEmp.dta", intermediate output of the code containing the production line in the BLS matched to the IFR sectors.

  Outputs:
  - "ProdLineEmpIFR.dta" containing the unmber of employees in each IFR sector.

  Summary of the procedure: the OES data is filtered to include only production-line employees ("Production occupations"), matched to IFR sectors and collapsed.
- “runCompustatResidIFR” generates the productivity residuals variables contained in "rawCompustatResidIFR.dta" Inputs:
  - "Compustatquery0726.dta"", containing the WRDS query detailed in the section "Data Sources" of the "Calibration" appendix;
  - "FIXEDINVDEF.csv", the fixed investment deflator from FRED;
  - "CrossWalkSICIFR.dta" is the crosswalk between SIC codes and IFR codes kindly provided by Daron Acemoglu and Pascual Restrepo, who used it in Acemoglu and Restrepo (2018);
  - "apr_measures_ifr19.dta", containing the robot penetration measures from Acemoglu and Restrepo (2018).

  Outputs:
  - "out/dataCleanedCompustat.dta" , intermediate output containing the real net investment in property, plant and equipment;
  - "out/rawCompustatResidIFR.dta", containing the residuals for each firm.

  Summary of the procedure: the series "ppegt" is used to obtain the first value of the capital stock registered for each firm, deflated by the fixed investment deflator. The series "ppent" (net investment in property, plant and equipment" is then interpolated to fill in missing years.

A series for real net investment is then constructed using the changes in the interpolated "ppent" series deflated by the fixed investment deflator from FRED. These are then added to the initial value of capital to obtain the value for the capital stock in each year. We then compute the residuals we need by regressing log-sales on log-employment, log-capital stock (both interacted with the IFR sector), sector by year fixed effects and firm fixed effects.

- "runDetrendResiduals" removes a firm-specific trend from the productivity residual variables contained in "rawCompustatResidIFR.dta" Inputs:

  – "out/rawCompustatResidIFR.dta", containing the idiosyncratic residuals for each firm-year.

  Outputs:

  – "out/detrendedCompustatResidIFR.dta" containing the idiosyncratic residuals for each firm-year purged of firm-specific trends.

  Summary of the procedure: a simple linear trend is estimated for each firm and subtracted from the residual series. Only firms appearing at least three years are included for obvious reasons.

- "runEstimatorsIFR" uses computes maximum-likelihood estimates of the CIR process for each IFR sector, using both raw residuals and detrended residuals. In our Matlab code, we employ the latter series, but also provide the former for robustness. Inputs:

  – "out/rawCompustatResidIFR.dta", "out/detrendedCompustatResidIFR.dta" containing the idiosyncratic residuals for each firm-year.

  Outputs:

  – "rawAlphaBetaIFR.dta", "detrendedAlphaBetaIFR.dta", containing estimates of the parameters $\alpha$ and $\beta$ for a CIR process as described in Wei et al. (2016);

  – "rawCIREstimatorsIFR.dta", "detrendedCIREstimatorsIFR.dta", containing the estimators for the standardized CIR process $\theta_z$, $\sigma_z$ used in the Matlab code.

  Summary of the procedure: we follow the formulas in Wei et al. (2016), applied here at the IFR-sector level. We then standardize the process so that the long-run average of $z$ is 1 in each IFR sector. Note that here we use $p$ instead of $z$ throughout.

- "runMergeAll" merges all the .dta files together with the share of manufacturing value added for each IFR sector. Inputs:

- "VA_DATA14.csv" containing the shares of manufacturing value added of each IFR sector in 2014. The match of GDP-by-sector reported by the BEA can be read from the formulas reported in the Sheet "Manufacturing14" of the Excel file "VALUEADDEDBEA.xls". "VA_DATA14.csv" is a copy of the last Sheet in the file.
- The output files mentioned in the previous steps.

Outputs:

- "Statistics.csv", the input for our Matlab calibration exercise.

# References


Daron Acemoglu and Pascual Restrepo. Robots and Jobs: Evidence from US Labor Markets. Working Paper, MIT, July 2018.

Sam Korus. Industrial Robot Cost Declines Should Trigger Tipping Points in Demand. https://ark-invest.com/research/industrial-robot-cost-declines, April 2019.

Hal Sirkin, Michael Zinser, and Justin Rose. How Robots Will Redefine Competitiveness. Technical report, Boston Consulting Group, September 2015.

Chao Wei, Huisheng Shu, and Yurong Liu. Gaussian Estimation for Discretely Observed Cox–Ingersoll–Ross Model. *International Journal of General Systems*, 45(5):561–574, July 2016.