README

September 6, 2021

Entrepreneurial Human Capital and Firm Dynamics

Francisco Queiró

1 Overview

This replication package reproduces the tables and figures in the paper and in the online appendix. The input folder contains the required data that are publicly available. The analysis also uses restricted-use data that cannot be shared with this replication package, and I describe how to obtain access to these data below. The code folder includes all the required stata and python code. When the code is executed, processed data and estimation results are saved in the output folder and tables and figures are saved in the results folder.

2 Data

The analysis in the paper combines publicly available data, which are included in the replication package, and administrative microdata on firms and workers in Portugal, which are confidential and cannot be included in the replication package.

2.1 Publicly available data, included in the input folder:

- Educational attainment across countries from Barro and Lee (2001)
  Downloaded from:
  https://web.archive.org/web/20050902095058/http://www.cid.harvard.edu/ciddata/ciddata.html
  Datafiles:
  "Appendix Data Tables.xls"
  "Appendix Data Tables - in Panel Set format.xls"
- Development accounting data from Caselli (2005)
  Downloaded from:
  https://personal.lse.ac.uk/casellif/data/handbook2005.zip
  Datafile: handbookdata_stata6.dta
- GDP Deflators for Portugal from FFMS (2019)
  Downloaded from:
  https://www.pordata.pt/Portugal/Deflatores+(base+2011)-502
  Datafile: PORDATA_Deflatores-(base-2011).xlsx
- Labor share of income by sector in Portugal from Stehrer et al (2019)
  Downloaded from:

https://euklems.eu/excel/PT_Growth-Accounts_SDB_2019.xlsx
https://euklems.eu/excel/PT_National-Accounts_SDB_2019.xlsx
Datafiles:
PT_Growth-Accounts_SDB_2019.xlsx
PT_National-Accounts_SDB_2019.xlsx

- Sector characteristics from Ciccone and Papaioannou (2009) and Fracassi (2017)
  Data on human capital intensity, external finance dependence, contract intensity and physical capital intensity are from table 1 in Ciccone and Papaioannou (2009)
  Data on social networks intensity are from table OA.II in the online appendix of Fracassi (2017)
  Datafiles:
  ciccone_papaioannou_sectors.xlsx
  fracassi_SNI.xlsx

- Firm employment from Gennaioli et al (2013)
  Downloaded from:
  https://scholar.harvard.edu/files/shleifer/files/regions_data_web_july2012.xls
  Datafile: regions_data_web_july2012.xls

- Correspondence between CAE-Rev.2.1 and CAE-Rev.3 sector codes
  Created by author
  Datafile: cae_rev21_rev3.xlsx

- Correspondence between 1995-1999, 2000-2005 and 2006 onward educational attainment codes in Quadros de Pessoal
  Created by author
  Datafile: schooling_codes.xlsx

2.2 Restricted-use data, not included in this replication package:

- Gabinete de Estratégia e Planeamento (GEP). "Quadros de Pessoal 1995-2017". Accessed on 21-07-2020

- Instituto Nacional de Estatística (INE). "Sistema de Contas Integradas das Empresas (SCIE) 2004-2017"
  - 2004-2009 versão POC. Accessed on 09-06-2016
  - 2010-2017 versão SNC. Accessed on 21-07-2020
- Instituto Nacional de Estatística (INE). "Inquérito às Práticas de Gestão 2016". Accessed on 21-07-2020

Access to these datasets can be obtained by accredited researchers from the National Statistics Office (INE) in Portugal. Accreditation is granted by Direção Geral de Estatísticas da Educação e Ciência at the Ministry of Education, and is restricted to research teams with at least one member affiliated with a Portuguese institution. Details on how to apply can be found here. The application should identify the required datasets using the citations provided above.

3 Computational requirements

3.1 Software

Stata 14:

- usespss
- iscogen
- qreg2

Python 3.7.8:

- jupyterlab 2.2.8
- sympy 1.6.2
- numpy 1.19.1
- scipy 1.5.2
- pandas 1.1.2
- matplotlib 3.3.2
- seaborn 0.11.0
- statsmodels 0.12.0
- patsy 0.5.1
- pyarrow 1.0.1
- rpy2 3.3.4
- xlrd 1.2.0
- openpyxl 3.0.5
- pyreadstat 1.0.2
- bottleneck 1.3.2
- numexpr 2.7.1
- pip 20.2.3
- econtools: pip install git+https://github.com/fqueiro/econtools (author: Daniel Sullivan)
- binscatter: pip install git+https://github.com/fqueiro/binscatterplot (author: Kevin Chen)

The code/environment.yml file can be used to create a conda environment that includes the correct versions of Python and all required packages.

R 4.02:

- binsreg 0.2.0

Windows (the Stata package usespss is not available for Linux/MAC operation systems)

3.2 Hardware

The code was run on an 8-core Intel i7-9800X processor with 64GB of RAM on Windows 10, and took about 4 hours.

4 Code

All required code is included in the code folder:

- The subfolder data_processing includes Stata do files that process the raw data for analysis and produce some of the statistics reported in the paper. The file do_all.do executes these

in the proper order.

- The file main.ipynb produces the main results, including all tables and figures except for those in section 3.4 of the paper and section B.5 of the online appendix.
- The file mechanisms.ipynb produces the tables and figures in section 3.4 of the paper and section B.5 of the online appendix, along with the corresponding numbers reported in the text.
- The file qregs.do runs the quantile regressions used to construct figure 6b
- The files util.py and settings.py include utilities and settings used in the analysis
- The file run_all.ipynb executes all code from start to finish, including stata and python programs

5 List of tables, figures and numbers

All tables and figures are included with this replication package, and can be found in the results folder.

Created by code/main.ipynb (links to specific cells are provided at the top of the notebook):

- Tables 1, 2, 4 and 5 in the paper, and tables E.1 and E.2 in the online appendix
- Figures 1 to 6 and 10 to 13 in the paper, and figures F.1 to F.5 and F.9 in the online appendix
- Relative firm sizes reported in the introduction, in section 3.1 and in online appendix B.1
- Robustness to $\beta^w = 6\%$ and $\beta^w = 10\%$ in section 3.2 of the paper
- Quantile regression coefficients reported in section 3.3 of the paper
- Model-implied aggregate returns to schooling reported in section 5.1 of the paper
- Caselli (2005) development accounting success measures with r=8% in section 5.2 of the paper
- Regression coefficients reported in online appendix B.2
- Correlation between baseline and unique definitions of entrepreneur schooling, reported in online appendix B.3
- Average firm employment in Gennaioli et al (2013) and the employment percentile it corresponds to in this paper's sample, reported in footnote 40 of the paper

Created by code/mechanisms.ipynb (links to specific cells are provided at the top of the notebook):

- Table 3 in the paper
- Figures 7 to 9 in the paper, and figures F.6 to F.8 in the online appendix
- Coefficients from figure 7 reported in section 3.4 of the paper
- Fraction of entrepreneurs and relative sizes by college field of study reported in section 3.4 of the paper
- Coefficients and confidence intervals from figure 9 reported in section 3.4 of the paper

Created by code/data_processing/entrep_firms1995_2017.do:

- Fraction of entrepreneurs identified by ISCO codes reported in section 2 of the paper (line 76)
- Firm exit rate reported in section 4.5 of the paper (line 99)
- 1 and 10 year manager schooling auto-correlations reported in online appendix B.1 (line 113)
- Fraction of college-educated entrepreneurs in 2016 reported in section 4.5 of the paper (line 118)
- Average years of schooling in the workfoce in 2004 and 2017, reported in online appendix B.2 (line 121)

6 Instructions for Replication

- Add restricted-use datasets to the input folder. Files should be unzipped and in the raw .sav format provided by the National Statistics Office. Quadros de Pessoal files should be placed in the input/QP_data folder, SCIE files should be placed in the input/SCIE_data folder and Inquérito às Práticas de Gestão files should be placed in the input/PraticasGestao_data folder.
- Set path where Stata is installed in line 17 of code/util.py
- Execute run_all.ipynb in the code folder

7 References

- Barro, R.J., Lee, J.-W., 2001. International Data on Educational Attainment: Updates and Implications. Oxford Economic Papers 53, 541–563.
- Caselli, F., 2005. Accounting for Cross-Country Income Differences. Handbook of economic growth 1, 679–741.
- Ciccone, A., Papaioannou, E., 2009. Human capital, the structure of production, and growth. The Review of Economics and Statistics 91, 66–82.
- Fracassi, C., 2017. Corporate Finance Policies and Social Networks. Management Science, 63 (8), 2420-2438.
- FFMS, 2019. Indicadores de envelhecimento em Portugal. PORDATA – Estatísticas, gráficos e indicadores de Municípios, Portugal e Europa. Retrieved July 7, 2019.
- Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2013. Human Capital and Regional Development. The Quarterly Journal of Economics 128, 105–164.
- Stehrer, R., Bykova, A., Jäger, K., Reiter, O., Schwarzhappel, M., 2019. Industry Level Growth and Productivity Data with Special Focus on Intangible Assets. Vienna Institute for International Economic Studies Statistical Report.