Unemployment Variance Decomposition - Code

Cristina Lafuente
University of Edinburgh

March 22, 2019

This document provides technical guidance to the accompanying code to our paper *Unemployment Duration Variance Decomposition à la ABS: Evidence from Spain*.

The code is written in STATA .do files, which calculate and decompose the variance, and Python, that produce the plots and the LaTeX tables for the paper. Other supplementary plots are also produced in Python.

Data

The paper uses the *Muestra Continua de Vidas Laborales* or MCVL. Access to this data is subject to approval by the Spanish Social Security agency. To request access to the data, please go to the official website[1] where you can find the application form. Only the 2013 files with fiscal data are required.

Once you obtain the files, please copy the following 6 files[2] into the "data" folder:

- MCVL2013AFILIAD1_CDF.txt
- MCVL2013AFILIAD2_CDF.txt
- MCVL2013AFILIAD3_CDF.txt
- MCVL2013AFILIAD4_CDF.txt
- MCVL2013PERSONAL_CDF.txt
- MCVL2013PRESTAC_CDF.txt

[1] http://www.seg-social.es/wps/portal/wss/internet/EstadisticasPresupuestosEstudios/Estadisticas/EST211

[2] It is possible that in their original format these files are named *MCVL2013AFILIAD1_CDF.TXT.txt*, with *TXT.txt* at the end. Please correct this typo, so at the end of the file, after *CDF*, there is only *.txt*

Decomposition code

The plots and tables in the paper are obtained in 3 steps:

1. Format the code. The .do file "format_data.do" formats the pension, afiliattion and personal files and saves them in .dta format. Then, it reads the 4 affiliation files and combines them with the personal file. This creates a file called "afilianon13.dta". Then the .do file "baseline.do" reads "afilianon13.dta", incorporates data from the pension file and performs other data cleaning operations (see http://github.com/crisla/MCVL for more details). The "baseline13.dta" file is produced. This is the base for the decompostition.

2. The "main.do" file reads "baseline13.dta" and depending on the type of decomposition, it applies the unemployment expansions as detailed in the paper. These are contained in the "cleaning_new.do" file. There is a separate cleaning file ("cleaning_new_ne.do") for the decomposition of non-employment. Then it carries out the decompositions, for which code is provided into separate files depending on the data expansion applied: LOWER_LOWER.do (raw data), LOWER.do (LTU expansion), LOWER_PLUS.do (STU Expansion) and UPPER.do (Spell Correction Adjustment). After the decomposition is done, it prints in the screen the main figures, and saves the results in .log format in the folder "results". It also saves a comma-separated (CSV) file with all of the spells for each expansion. This forms the basis of the histograms in the paper.

   Similarly, the "BCycles.do" file reads "baseline13.dta" and splits the sample in the 2002-2007 period and the 2008-2013 period. Then it carries out the decomposition as "main.do".

3. The jupyter notebook file "tables_and_plots.ipynb", when opened, reads the output from the results folder and creates the bar charts and tables from the paper. The notebook contains some extra figures and tables and can be used to generate different figures - for example, we only do the decomposition in logs, but the notebook can generate the figures and tables for the level decomposition too. The jupyter notebook file "histograms.ipynb" creates the histogram plots of the paper with the CSV files.

Note that "tables_and_plots.ipynb" is self-contained, in the sense that you do not need access to the data to be able to generate the plots and results from the paper.[3] Access to the data is necessary to generate the decomposition results.

[3] "histograms.ipynb" is not self-contained at the moment, as it needs the CSV files.

Supplementary plots

Other plots of the paper (how duration in the MCVL compares to duration in the LFS) are created in the "supplementary_plots.ipynb" file.