Readme file: "Tax progressivity and self-employment dynamics"

by Wiji Arulampalam and Andrea Papini.

April 2021

1. DATA

The main data used in the research is the Norwegian Income and Wealth Statistics for Persons and Families. This data contains information drawn from the annual tax returns, and other administrative registers, including the education register. Since these are drawn from confidential personal files, they are not allowed to be shared outside of Norway according to the country's Statistics Act. Access to the data for research purposes for individuals, can however, be obtained through affiliation with a Norwegian research institution. They will have to fulfil formal requirements of the Central Statistical Bureau of Norway (SSB).

Please contact the SSB (https://www.ssb.no/en/) directly to inquire about the original data and how to access them.

2. Files

We used *SAS* for the initial data download and preparation, and *STATA* for further data preparation and all our estimations. The files are provided in 5 **folders** and referenced by names according to the final set of results reported in the paper.

Folder 1_table_2 is the main folder and contains the command files used in the creation of the samples for the main results reported in Table 2. In each of the folders 2_Table_1 to 5_Table_4, the first *STATA* do-file includes an initial comment that states the name of the do-file to use from folder 1_table_2, for the analysis to be performed.

Majority of the work was carried out using *STATA* version 15. However, some of the analysis required random drawings from specific distributions. We fixed the version number to 13 in order to ensure that the draws were the same from estimation to estimation. This information is provided in the relevant command files.

(a) 1_Table_2:

The main folder has the command files used in the creation of the samples for the main results reported in Table 2. The first program file called 0_data_sas_merge_naming_variables.sas is the one used for the initial download of the data from the SSB server using *SAS*. The *SAS* commands are general and should work with any version of *SAS*.

The rest of the command files are *STATA* do-files, and these are numbered consecutively to show the order in which they are executed. Altogether there are nine files in this folder.

(b) 2_Table_1:

This has only one do-file that provides the summary statistics reported in Table 1 and the analysis of covariance results reported in footnote 26.

(c) 3_Table_3: This contains the five do-files used in the experiments reported in Table 3.

(d) 4_Graphs: This has only one do-file that produced the graphs reported.

(e) 5_Table_4: This has subfolders called Panel_A to Panel_H, containing all the do-files used in the sensitivity checks reported in Table 4.

3. Creation of unemployment rate variable

The regional unemployment rate variable (see STATA data file reg_u_1996.dta) used in Table 4 Panel E estimation, was created using different information available at https://www.ssb.no/en/statbank and these are provided in various files as:

The information about the figures in the tables:

(a) Sheet (a): 10539: Unemployed persons registered at the Employment Offices, by contents, region, age and month:

- The tables are updated only once a year as of 2015. Statistics Month in November.
- . = Category not applicable. Figures do not exist at this time, because the category was not in use when the figures were collected.
- Latest update at time of download: 20210127 08:00
- Source: Statistics Norway
- Units: persons
- Data type: Stock;
- Reference period: End of month
- Database: External PROD; Internal reference code: Registrerte1

(b) Sheet (b): Annual average of (a)

(c) Sheet (c) 05613: Persons in the labour force and employed persons, by region, sex, labour force status, contents and year

- Latest update at time of download: 20200123 08:00
- Source: Statistics Norway.
- Units: Persons (1 000):
- Data type: Average
- Reference period: Continuous survey
- Database: External PROD; Internal reference code: Sysselsatte

(d) reg_u_1996.dta:

- unemployment rate calculated as figures in (c) divided by figures in (d), calculated at the region level.
- Regions: region 1: counties 1 to 7; region 2: counties 8 to 10; region 3: counties 11 to 15; region 4: counties 16 and 17; region 5: counties 18 to 20

The authors can be contacted if there are any queries regarding the program files. Their email addresses are: wiji.arulampalam@warwick.ac.uk and andrpapini@gmail.com.