Data archive manual for: The Empirics of Hidden Labor Force Dynamics in Germany by Sandro Provenzano for the publication in Journal of Economics and Statistics

Overview:

All raw data can be found in plain text ASCII format in the folder "raw_data". In addition, "raw_data" contains the codes for the formatting and analysis
that was used when conducting the study with Stata in text format. The folder "data_formatting" contains the Stata-do-files for preparing the raw data for
the analysis. Furthermore, it contains data in the .dta format that can be read with Stata. Last but not least, the file "Main Code Hidden Labor Force
Dynamics" is the Stata-do-file to undertake the analysis of the paper and directly corresponds to the table outputs in the paper.

The folder "raw_data":

1) The files "data_formatting_Stata", "main_code_Stata" and "quarterly_data_formatting_Stata" contain the codes of the do-files used to prepare and analyze the data
and are written by the author.

2) The files "population_data", "quarterly_population_data", "quarterly_LFPRs" and "yearly_LFPRs" contain the information about population size and shares as well as
the labor force participation rates over time for the subgroups. In total it contains 88 subgroups that divide the total population starting at the age of 15 into 11
age groups, Eastern and Western Germany, Male and Female as well as German citizens and Non-German citizens. The subgroups are named systematically in order to be able
to directly identify each subgroup. For example, in W_D_M_15_19 "W" stands for "Western German" (as opposed to "O" for "Eastern German"), "D" for "German citizens" (as opposed
to "A" for "Non-German citizens"), "M" for "Male" (as opposed to "F" for "Female") and "15_19" which includes all that are between 15 and 19 years old (including those
who are 15 or 19 years old). The designation "ab_65" means people that are at or above 65 years of age. Those groups are not part of the analysis as they are not
considered to be part of the labor force. The variable "date" contains the year or year and quarter (q1 for first quarter and so on). The variables in "population_data"
refer to the total population of the subgroup in the given year ("abs_Bev_"), their share in the population in a given year ("Bev_"), as well as their average share in
the population ("Bev_mean") and are organized in the long format. The file "quarterly_population_data" contains the total population of a particular subgroup in a given
quarter in 1000 in the wide format. The yearly LFPR and population data have in total 2112 unadjusted observations while the quartlery data have in total 3872 unadjusted
observations. The source of all of this data is the Microcensus in addition to own calculations. More information on the Microcensus can be found on the website of the
Federal Statitical Office (https://www.destatis.de/EN/Homepage.html) as well as on
http://www.forschungsdatenzentrum.de/en/database/microcensus/index.asp.

3) The files "unemployment_rates_yearly" and "unemployment_rates_quarterly" contain the official yearly and quarterly unemployment rates of Germany that relate to the
entire civilian labor force. The yearly unemployment rate contains a total of 24 observation while the quarterly unemployment rate contains 44 oberservations. The data
can be found on the website of the Federal Employment Agency
(https://statistik.arbeitsagentur.de/Navigation/Statistik/Statistik-nach-Themen/Zeitreihen/Zeitreihen-Nav.html).

The folder "data_formatting":

This folder contains two do-files (one for yearly and one for quarterly data) and associated .dta-files to process the raw data and prepare it for the final analysis
using Stata. All the data is based on the data given in the folder "raw_data". The finalized datasets that are being using for the analysis are also saved in this folders
and are called "yearly_dataset" and "quarterly_dataset".

The file "Main Code Hidden Labor Force Dynamics":

This is the code with which the results presented in the paper can be obtained. The variable c_ always refers to the labor force participation rates for the subgroups
given by "group". The variable "ABLQ" refers to the unemployment rate. All other variables are either self-explanatory or have been explained above. In order to replicate
the results of the paper the file directory as specified in line 3 "cd ..." has to be adjusted to the respective computer and file location of the user. Then, the results
can be obtained simply by executing the commands step by step using Stata.

--- end of readme ---