Imputation of missing EE transitions in the monthly CPS
Reference: "Measuring Employer-to-Employer Transitions"
by Shigeru Fujita, Giuseppe Moscarini, and Fabien Postel-Vinay

Download CPS basic monthly micro data files from:
`https://www.census.gov/data/datasets/time-series/demo/cps/cps-basic.html`.

Run Stata scripts in the following order.

1. `1_extract.do`: This code extracts variables of interests from raw CPS monthly files. It produces Stata data files for all months, starting in January 1994. Each monthly file is named `cpsmyyyyymm.raw`, where `yyyy` refers to year and `mm` to month.

2. `2_1_adjust_ind.do`: This code creates industry categories that are consistent over time, and overwrites monthly CPS Stata files.

3. `2_2_adjust_occ.do`: This code creates occupation categories that are consistent over time, and overwrites monthly CPS Stata files.

4. `3_match.do`: This code matches individuals across two adjacent months using the existing files and saves each matched file as `matchedyyyyymm.raw`.

5. `4_genvars_Probit.do`: This code generates additional variables, taking each matched dataset. In particular, it runs Probit regressions, computes predicted probabilities of having an missing answer to the EMPSAME question, and overwrites the data.

6. `5_append_matched.do`: This code appends all monthly matched files into one dataset and saves it as `cps_all_matched.dta`.

7. `6_genvars_matched.do`: This code takes `cps_all_matched.dta`, generates additional variables, and saves it as `cps_matched_all_additional_vars.dta`.

8. `7_UEr_JJrSS_append.do`: This code uses `cps_matched_all_additional_vars.dta`, generates aggregate time series of the UE transition rate and E2E transition rates for selected groups (the overall SS group and the SS group between the first and second interviews), merges these aggregate series back to the original dataset, and overwrites it. The E2E rate for the SS group between the first and second interviews is used in the imputation as the trend and cyclical variables, as described in the paper.

9. `8_1_Imputation_base.do`: This code implements the imputation of the missing observations and generates the aggregate time series including the final E2E rate for the U.S. economy. The final results are stored in `BTwoDummiesJJSS1CycTr.xlsx` and `Imp_2DummiesJJSS1CycTr.xlsx`. The first one includes average bias series for the five respondent groups and the later one includes E2E rates for various groups. As explained below, Matlab codes are written to take these series and draw graphs in the paper. This code also saves the final micro data file as `ee_micro_final.dta`.