Instructions to construct ATUS datasets, replicate empirical results in Cacciatore, Gnocchi and Hauser (2024)

NOTE: At the top of each Stata .do file, you must add your path to the main CGH_Empirical_Replication folder

To create the ATUS data extract:

For each year=2003, 2004, …, 2019 of data:

1. Download single year datafiles from https://www.bls.gov/tus/data.htm. Specifically, download basic .do and .dat files for
    a. Respondent file
    b. Roster file
    c. Activity file
    d. Who file
    e. CPS file

    For a description of these files that are to be linked together see micro data files. For a description of the variables included in the specific data files see Data Dictionaries.
    The basic .do files that need to be downloaded are: "atusact_year.do", "atuscps_year.do", "atusresp_year.do", "atusrost_year.do", and "atuswho_year.do". The .dat files that need to be downloaded are: "atusact_year.dat", "atuscps_year.dat", "atusresp_year.dat", "atusrost_year.dat", and "atuswho_year.dat".
2. For each year, save the do files in a folder called "Basic Do Files" and save the .dat files in a folder called "Basic Original Data Files". If you use alternative names for these folders, then you will need to make the respective modifications to the directory paths relative to those that are currently in the codes.
3. For each year, create two additional folders. The one folder is called "Data Creation" and contains the do file `format_year.do' that extracts the micro data and codes up the time use categories. The other folder is called "Basic Transformed Data Files". This is the folder where all .dta files will be saved.
4. For each year, run the 5 .do files contained in the "Basic Do Files" folder. To run these, you have to adjust directories in the 5 .do files so that the directories where the do files are located and the directories where the created dta files are placed match the directories in your computer. Running the 5 .do files will generate 5 corresponding .dta files, which will be saved in the "Basic Transformed Data Files" folder.
5. The format.do files for each year will then aggregate the time use activities to the categories used in the paper.
6. See the activity coding lexicon for a list of all activities each year.

To create the AHTUS data extract:

1. The American Heritage Time Use Study data can be accessed from the IPUMS website: IPUMS AHTUS: vars by group (ahtusdata.org)

2. Create separate data extracts for the years 1993, 1995 and 1998 containing the following variables: RECTYPE, SAMPLE, IDENT, HHID, PID, YEAR, RECWEIGHT, AGE, SEX, EDUC, ETHNIC2, INFLTWT, XTIMEWT, STATE, UNDER18, MAIN TIME.
3. Move the three .dat files to the data/ATUS_AHTUS_raw_data folder with the following naming convention: ahtus_YYYY.dat.
4. Although time use in the AHTUS is not captured with the same level of detail as in the ATUS, we match our classification of activities with the classification of AHK as much as possible. The coding of the "main" variable (capturing the main activity done by the respondent) contains the AHTUS coding of activities. Our classification of the more granular activities into the categories used in the paper can be found in lines 429-447 of 06_AHTUS_IPUMS_extracts.do.

To generate the ATUS and AHTUS datasets:

1. **Run 01_ATUS_data_creation_merge.do**

- Runs a set of .do files from the ATUS_AHTUS_raw_data/YYYY/Data-Creation folder for each year which create properly formatted subsets of the ATUS microdata.
- Merges the data for each year into one single file.
- The output is ATUS_ALL_FINAL.dta

2. **Run the codes to aggregate the ATUS microdata**

- Run the following codes:
    - 02_ATUS_generator_1865_AllUS.do
    - 03_ATUS_generator_1865_annual.do
    - 04_ATUS_generator_1865_biannual.do
    - 05_ATUS_generator_1865_quarterly.do
- Input for each of these codes is ATUS_ALL_FINAL.dta
- Each of these programs produce averages of the ATUS data at a specific frequency and geographic level of aggregation.
- The outputted files are called state_dataset_..._temp and contain ATUS data aggregated to the level specified by the code.
- ATUS_generator_1865_AllUS produces data aggregated at the national level while the other codes produce data aggregated at the state level.
- NOTE: These files take time to run.

3. **For each of the state_dataset_ files, manually delete the first row and first column. Rename the "work paper" and "gestfips" columns to "nobs" and "idn", respectively. Save the files under the same name with "_temp" removed so that these files do not get overwritten if the codes are run again.**

4. **Run 06_AHTUS_IPUMS_extracts.do**

- Input is the raw American Historical Time Use Survey (AHTUS) microdata from IPUMS.
- The output is the "AHTUS_all.dta" file, which contains state-level averages for the 1993, 1995 and 1998 versions of the AHTUS.
- Also makes use of 2010 state-level population weights.

To reproduce Table 1:

1. **Run 11_Table1.do**

- Inputs are state_dataset_twoyear.xls and state_dataset_quarterly.xls.
- Uses state_dataset_twoyear to run the state-level regression that generates the results in column 1 and column 5 of Table 1.
- Uses state_dataset_quarterly.xls to run the quarterly state-level regressions and produce the remaining Table 1 results.
- Output in the replication code is a .csv version of Table 1 which contains the same numbers as Table 1 in the paper.
- Also makes use of 2010 population weights and state-level uncertainty dummy variable.

To reproduce Figure 1:

1. **Run 21_regressions_annual_betas.do**

- Inputs are the AHTUS_all.dta and state_dataset_year.xls
- Estimates separate substitution rates for each year in the ATUS and AHTUS sample.
- Also makes use of 2010 population weights (separate files for ATUS and AHTUS due to different coding of states).
- The output is the ATUS_AHTUS_betas.dta file which contains the estimated substitution rates for each year, for each time use category.

2. **Manually input the substitution rate (beta) estimates from Column 1 of Table 1 for nonmarket hours and leisure into the SyntheticSeriesForLP_Quarterly code as the values for BETA_Nonmarket and BETA_Leisure (approximately at lines 75 and 76).**

3. **Run SyntheticSeriesForLP_Quarterly.m**

- Inputs are the single beta values from Column 1 of Table 1 (inputted in the previous step), the annual beta values from ATUS_AHTUS_betas.dta and the BLS market hours worked series.
- The outputs are quarterly synthetic series for leisure hours and nonmarket hours constructed using both single and annual substitution rates.

4. **Run 23_Figure1.do**

- Inputs are the quarterly macro dataset and the synthetic leisure and nonmarket hours series.
- Identifies exogenous vix shocks and runs local projections to estimate their impact on market hours, nonmarket hours and leisure hours using a single substitution rates and substitution rates that change in each year.
- The output is a .pdf version of Figure 1.

To reproduce other data files used in our analysis:

1. **Population weights**

- The population weight files for the ATUS and AHTUS are state_pop_shares.dta and state_pop_shares_AHTUS.dta, respectively.
- Data for population weights used for our state-level regressions are from the 2010 state-level population weights in US Census.
- Raw data is available in \uncertainty_popweights\StatePopulation.xlsx and can be downloaded from the Census Website (NST-EST2019-01).
- Since the AHTUS has a different coding for states compared to the ATUS (which uses fips codes), we also generate a file population weight file with population weights matched to the AHTUS state codes.
- The population weights are identical between the two files. Only the coding of the states changes. The coding of the AHTUS state variable can be found on the IPUMS website.

2. **State-level policy uncertainty**

- The state-level policy uncertainty data file is state_uncertainty.dta.
- Data are from the Economic Policy Uncertainty Index and can be downloaded from the Economic Policy Uncertainty Website under "US State-Level EPU".
- To generate the quarterly dummy variable used in the paper, we take the "EPU_State" series for each state from 2003-2019 and create a variable equal to 1 when the index is at least one standard deviation above the mean for that state (means are computed over the same period).
- The monthly dummy variable is aggregated to quarterly frequency by assigning a value of 1 if at least one month in the quarter has a value of 1.
- State_Policy_Uncertainty.xlsx contains the data (Raw Data tab) and all manipulations done to obtain the quarterly dummy variable (Calculations tab).

3. **BLS Market Hours Worked**

- The Bureau of Labor Statistics market hours data file is NONFARM_HOURS_Q.xlsx
- The BLS market hours series is constructed as aggregate hours of wage and salary workers on nonfarm payrolls divided by the working-age civilian noninstitutional population.
- This series is divided by 13 to convert it from a quarterly total to an average weekly total.
- BLS Market Hours and working-age population were obtained from Haver series LHTNAGRA@USECON and LN16N@USECON, respectively.
- The manipulations can be found in the same Excel file.

4. **Quarterly macro data for local projections**

- Manipulations and data sources for our quarterly macro data series are available in the "description" tab of the quarterly_macro_data Excel file.

Instructions to construct ATUS datasets, replicate theoretical results in Cacciatore, Gnocchi and Hauser (2024)

NOTE: The codes call Dynare and Dynare ++.

1. Download and install Dynare 4.4.3.
2. Add the subfolders Dynare/Matlab and Dynare/Dynare ++ to the Matlab path.

The results of the paper were obtained using Matlab 2014b (the Symbolic Math Toolbox and the Optimization Toolbox are needed).

Expected running time (all figures): 1 hour

To reproduce Figures 3, 4, and 5:

Extract the folder *CGH_model_replication* and place it in the root directory (e.g, C:/).

Run the script *run_ge_model.* Choose the figure to reproduce by setting to 1 its corresponding value (only one figure at the time can be reproduced):

- *figure3* = 1 to reproduce Figure 3.
- *figure4* = 1 to reproduce Figure 4.
- *figure5* = 1 to reproduce Figure 5.

The script calls the following auxiliary scripts/files:

- *CGH.mod*: set variable and parameter names and call the file that computes the model steady state (*CGH_steadystate.m*)
- *CGH_steadystate.m*: computes the model steady state.
- *CGH_steadystate_solve.m*: solves for the model steady state.
- *RunDynareppPruning_4th.m* (copyright: Martin M. Andreasen): calls *CGH_dplus.mod which* solves the model with a fourth-order approximation using Dynare++ (the file *image_dyn_plus.m* sets the steady-state values called in the mod file).
- *input_dynare_toolbox.m*: sets parameters needed by *dynare_toolbox.m*
- *dynare_toolbox.m*: obtains the model policy functions and simulates the model. The file uses *input_dynare_toolbox.m* as an input. It also calls *create_g_matrix, Dynarepp_Unfold_Matrices_4th*, (copyright: Martin M. Andreasen), *StateSpace_form_LinearInov_4th* (copyright: Martin M.Andreasen), and *full_block_dr_new* (copyright: Hong Lan and Alexander Meyer-Gohde).
- *save_irfs.m*: simulate the model subject to the chosen shock (figure3, figure4, figure5).
- *plot_irfs_paper.m*: plots figures 3, 4, and 5.