CPS Labor Market Flows and Wages Replication
README File

Octavio M. Aguilar

February 5, 2025

This README file contains the directory layout and replication steps to calculate labor market flows from the Current Population Survey (CPS) and estimate wages using a mincer earnings function. At the time of this README, data availability spans 1994m1-2024m12. The programs can be easily adjusted to your period of interest. If there are any questions or concerns please email me: octavio.m.aguilar@frb.gov

1 Folders

1. The project "cps" has 2 folders.

    (a) **Data**. In this folder you will find two sub-folders:

        i. **Raw** This folder contains: (1) the raw CPS data from IPUMS; (2) a extended series of the Help-Wanted Index from Barnichon (2010); and (3) RGDP (nonfarm), unemployment in thousands, unemployment to employment flows, and not in the labor force to employment flows.

        A. CPS data: To retrieve the CPS data visit the IPUMS CPS website and download all the necessary variables (the list of variables can be found in cps_variables_list.txt). I downloaded the data in bins due to the large sample size. I downloaded them as: 1994-2001, 2002-2009, 2010-2017, and 2018-2024. After you unzip the data in the raw folder rename it to match the naming convention: cps_startyear_endyear.dta.

B. CompositeHWI.xls: This excel file is taken from Barnichon's website and contains the monthly Help-Wanted Index from 1951 to 2021. I extend this series by pulling JTSJOL and CLF16OV from FRED, which are the job openings (total nonfarm) and civilian labor force level, respectively. His calculation is simply JOLTS/LF or as he says, V/LF.

C. FRED_.xls Files beginning with FRED are pulled directly from the FRED database. This includes RGDP (nonfarm), unemployment in thousands, unemployed to employed flows, and not in the labor force to employed flows.

D. fame_series1_sa.xls This is the output from the FAME program discussed in the programs section below. Briefly, I use the built in FAME function of X-13ARIMA-SEATS to seasonally adjust the main variables used in the analysis.

ii. **Clean**. This folder contains all of the clean data from the raw folder and is defined at the quarterly frequency.

2. **Programs**. In this folder you will find the programs used to clean and analyze the data. In particular, there are three subfolders.

(a) **Clean:** This folder contains the Stata programs to import and clean the data.

i. s1_import_cps.do: This program will import the 1994-2001 raw CPS file and assign it to be the base file. It will then append 2002-2024 data to it. It will save the raw data such that you have a complete dataset from 1994-2024.

ii. s2_import_additional_data.do: This program will import the FRED and Help-Wanted Index and convert it to be compatible with Stata.

iii. clean_cps.do: This program will take the data created from step 1 and clean/generate variables of interest. It will save it as "cps_clean.dta".

(b) **Analysis:** This folder contains the Stata programs to analyze the data and calculate the main statistics of interest: labor market flows, wages, and residuals from the mincer earnings function.

i. The programs are listed from step 1 to 4 and they should be ran in order. The program "s2_create_series1.do" will fail at the last block. This is because you have not yet created the seasonally adjusted series. Please run step 2 up to the "save" block. Then proceed to step 3 followed by executing the FAME program instructions below. After this is done you may return to step 2 and run the remainder of the program.

(c) **Fame:** This folder contains the FAME program to seasonally adjust the cps data.

i. **Instructions for running the program:** To run the FAME program you will need to load the entire folder into a VS-CODE session. Once you do this, take note of the three .PRO files. These are FAME program files. In particular, there will be fame_excel.pro, seasmult.pro, and seasadj.pro. These programs load the excel version of the cps series1 data, created the seasonal adjust function, and executes the function and saves the data as csv, respectively. You will only need to focus on seasadj.pro. Load this program and execute it. Instructions on how to execute the program using your terminal are provided at the top of the program. After running the program, it will export the adjusted data to a csv file in the same fame folder titled "series1_adj.csv". You will need to manually correct the timing convention and rename the variables to your choosing. Once you do this save it as a Stata data file here: "data/clean/fame_series1_sa.dta"

2 Replication Instructions

1. In each program, you will find that I have set the global directory **"home"** to be my local computer. If you keep the folder stucture the same, the only thing you must do is modify this home directory to match your system.

2. After changing the directory, please start by running all the programs in the clean folder. They are label from step 1 to step 4 (see notes in the program section above). After running each program in the clean folder, proceed to run steps 1 to step 4 in the analysis folder.