README

Replication Package for “What Caused U.S. Pandemic-Era Inflation?” by Ben Bernanke and Olivier Blanchard

Overview

This replication package recreates Figures 1-14, Tables 2-5, and the online appendix from “What Caused U.S. Pandemic-Era Inflation?” by Ben Bernanke and Olivier Blanchard. The replication package was prepared for the American Economic Journal: Macroeconomics by Sam Boocker of the Brookings Institution (sboocker@brookings.edu).

The replication package is organized as follows.

- All data can be found in the folder *(1) Data*.
- All regressions can be found in the folder *(2) Regressions*.
- All core simulations and decompositions can be found in *(3) Core Results*.
- Additional results can be found in *(4) Other Results*.
- Results and code for the appendix can be found in *(5) Appendix*. The online appendix details data sources (in depth), presents detailed results for both the full and pre-COVID samples, alternative specifications of certain variables, and additional tests.

Detailed instructions on running the code can be found below under the section of this README entitled *Running the Code*. These steps should be followed in sequence.

Data and Code Availability

To run this replication package, one needs access Stata, MATLAB (including Dynare), R (preferably RStudio), and Python.

The majority of the data, including data for all of the major results from the paper, is freely available from the FRED database, maintained by the Federal Reserve Bank of St. Louis, or Google Trends. To recreate the principal component analysis, data from Bloomberg and Haver is required.

All public data is included under the folder *Public Data*. The file *Regression_Data.xlsx* contains a data dictionary of the public data used in the main specification of the model in the paper. The *figure_6_data.xlsx* file contains used in plotting Figure 6, from FRED and Google Trends.

Private data, used only for the principal component analysis, can be found under the folder *Private Data*.To recreate the Bloomberg datasets, open the *Private Data* folder and open the *Bloomberg_Commodity_data.xlsx* file. Then, log into the Bloomberg terminal and, using Bloomberg's Excel add-in, refresh the sheet. Bloomberg data should now populate the file.

To recreate the Haver datasets, open the *Private Data* folder and open the *SP_GSCI.xlsx* file. Then, using Haver's Excel add-in, press ctrl+d. Haver data should now populate the file. Because the Haver indices are taken from S&P Global, these data are likely available there as an alternative.

Data used for alternative specifications, including expectations data from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters (SPF) and PCE inflation data from FRED, can be found in the public data folder as well, the former under *SPF_data.xlsx*.

Computational Requirements

The approximate time needed to reproduce the analysis on a standard laptop is less than 20 minutes. The code was last run on a Macbook laptop with an Intel 8-core I9 chip, 32GB of RAM, macOS version Sonoma 14.0, and an Intel UHD Graphics 630 1536 graphics card.

In terms of software, we used: Stata versions 17 and 18, MATLAB R2022a, including Dynare 5.4, Python version 3.8.8 (or later), and R version 4.3.1. Additional packages in Python, R, and Stata may need to be installed by the user:

- In Python, install *pandas*, *os*, and *sklearn*.
- In R, install *tidyverse*, *magrittr*, *ggplot2*, *zoo*, *tseries*, *readxl*, *grid*, *lubridate*, and *reshape2*.
- In Stata, install *estout*.

Running the Code

To replicate the results from the model, take the following steps:

1. Go to *(2) Regressions* and run regression_full.do and regression_pre_covid.do. This file

estimates the empirical SVAR over the full and pre-COVID samples, respectively.

   a. Change the working directory and the export directory at the top of the file. Input data can be found in the *(1) Data* folder.
   b. The output is found in the *Output Data* folder.
      i. The estimated coefficients are in *eq_coefficients_.xlsx* type files.
      ii. The summary statistics are in *summary_statistics_.xlsx* type files.
      iii. The data used in the estimation and for future simulations are in *eq_simulations_data_.xls* type files.
   c. To generate plots of actual data compared to out of sample predictions, go to the subfolder *Predicted vs. Actual* and open *plot_pred_v_actual.R*. Change the working directory and file paths of the Excel files. This R file generates 3, 7, 8, and 9. (Note that to correctly replicate these figures, one needs to use the output from the pre-COVID sample.)

2. Go to *(3) Core Results*. This folder contains three subfolders: *Conditional Forecasts*, *Decompositions*, and *IRFs*. (For the remainder of the results, we use the full sample regressions.)
   a. The subfolder *Conditional Forecasts* contains MATLAB code that runs the conditional forecast (Figure 14 in the paper). To run the code, simply open the MATLAB code, change your working directory and the directory of the data output. Note that because the assumed steady state value for V/U (taken to be Q4 2019), we must make an adjustment to the constant term of the wage equation for the simulation; technical details on this adjustment can be found in *wage_constant.pdf*. Output can be found in the *Output Data* folder. To plot the output, run *plot_cf.R*, changing the working directory and file paths of data as appropriate.
   b. The subfolder *Decompositions* contains MATLAB code that runs the price and wage decompositions (Figures 12 and 13 in the paper). To run the code, simply open the MATLAB code, change your working directory and the directory of the data output. Output can be found in the *Output Data* folder. To plot the output, run *plot_decomp.R*, changing the working directory and file paths of data as appropriate.
   c. The subfolder *IRFs* contains MATLAB code that runs the impulse response functions (Figures 10 and 11 in the paper). To run the code, simply open the MATLAB code, change your working directory and the directory of the data output. Output can be found in the *Output Data* folder. To plot the output, run *plot_irfs.R*, changing the working directory and file paths of data as appropriate.

3. In *(4) Other Results*, one can replicate the principal component analysis, and the simple calibrated model.
   a. The file *pca.py* was used to calculate the first principal component of a basket of

19 commodities. These commodities included: crude oil (WTI), heating oil, natural gas, corn, soybeans, live cattle, gold, aluminum, copper, sugar, cotton, cocoa, coffee, nickel, wheat, lean hogs, orange juice, silver, RBOB gasoline, unleaded gasoline, iron, lead, and zinc. This file reads confidential data from Bloomberg and Haver, stored in the confidential data folder discussed above in the files *Bloomberg_Commodity_Data.xlsx* and *SP_GSCI.xlsx*. The best way to run this file is to download Anaconda, a simple interface for Python. Make sure to input the Bloomberg and Haver data into the *PCA* folder before running the *pca.py* file; the output is *Bloomberg_Commodity_Data_Cleaned_Quarterly.xlsx*. Plots are made with *plot_pca.R*.

b. To generate the plots for Figure 6, open the *Shortages* subfolder and run *plot_shortage.R*.
c. To run the calibrated simulations, open *Simple Model* and run *Simple_Model.m*. Be sure to change the working directory. Output is in the *Output* folder in *simple_model_irf_results.xlsx*. Note that to run .mod files separately, simply type into MATLAB's command window "*dynare insert_file_name_here.mod*". The dynare scripts are defined in dynare scripts defined in *Simple_Eq_simulations_weak.mod* and *Simple_Eq_simulations_strong.mod.*

Appendix Results

Variable mnemonics used in the reporting of regression results are given in Table A1 below (see also Table 1 in the text). In Table A2, variables used in Figure 6 are provided. Bloomberg and Haver mnemonics are given in Table A3. Full estimation results in VAR form for both the pre-Covid and full samples are provided in Tables A4 and A5, respectively. (Recall that in the pre-Covid sample, the price equation is estimated over the entire sample period (1989 Q1 to 2023 Q2) and all other equations are estimated over the pre-COVID sample period (1989 Q1 to 2019 Q4). In the full sample, all equations are estimated over the entire sample period.) Estimates and selected statistics for the four equations of the model using alternative measures of the endogenous variables are given below (Tables A6 and A7); to run the regressions to generate these tables, use the code in *PCE Regression* and *SPR Regression*, respectively. Tables A8 and A9 give alternative estimations of the price equation with an alternative measure for shortages and V/U included, respectively; to run these regressions, use the code in *Shortage Alternatives V_U in the price equation*, respectively. Lastly, in Table A10, results on the non-linearity of V/U in the wage equation are given; the code for this table is found in *Nonlinear V/U*.

Where applicable, code and output for each of table can be found in the corresponding subfolder of *(5) Appendix*; input data is included under *(1) Data*.