REPLICATION FILES
"Modeling Time-Varying Uncertainty of Multiple-Horizon Forecast Errors"

This readme file describes the set of replication files ("the replication set") for "Modeling Time-Varying Uncertainty of Multiple-Horizon Forecast Errors" forthcoming in the Review of Economics and Statistics. The replication set contains code as well as all of our input data in raw form as obtained from their original sources described further below.

Authors

- Todd E. Clark (Federal Reserve Bank of Cleveland)
- Michael W. McCracken (Federal Reserve Bank of St. Louis)
- Elmar Mertens, (Deutsche Bundesbank)[1]

Overview

The replication set comes in the form of this *readme* as well as a tar file. The tar file *cmm2018.tar* comprises the contents of several directories with code and data for this project; use of its contents is described further below after a description of our data sources.

Data

As described in Section 2 of our paper, data used for this project comprises SPF survey responses as well as realized values for five different macroeconomic variables. Our data has been obtained from two, publicly available online sources:

1. The real-time data research center (RTDRC)[2] at the Federal Reserve Bank of Philadelphia.
2. The FRED database[3] hosted by the Federal Reserve Bank of St. Louis.

Specifically, from the RTDRC we obtained SPF mean responses for the following variables (with SPF mnemonics in parenthesis as listed at the RTDC website[4]):

- Level of real GDP/GNP (RGDP)
- Level of the price index for GDP/GNP (PGDP)
- CPI inflation rate (CPI)
- Civilian unemployment rate (UNEMP)
- 3-month Treasury bill rate (TBILL)

In addition, we collected first-release data for realized values of RGDP and PGDP from the RTDRC.

- PGDP[5]
- RGDP[6]

From FRED, we collected realized values for CPI, UNRATE and TBILL (mnemonics: CPIAUCSL, UNRATE, TB3MS) using "final" vintage data available per March 21 2018.

The replication set includes copies of the raw input files. Below we also describe code that transforms the input data before further processing by our main estimation routines.

All data has been downloaded on March 31 2018.

Code

All code used for this project has been written in Matlab. The code has been run on various, recent Matlab versions (Versions 2016b through 2018b) as well as different operating systems (Linux, Windows and macOS) without need for any particular adjustments across platforms. The codes uses Matlab's Statistics and Machine Learning Toolbox toolbox as well as (optionally) the Parallel Computing Toolbox. The final results for the published paper were generated using Matlab 2018a on macOS High Sierra.

The Matlab code also creates LaTeX files collecting tables and figures produced by the Matlab code. If a LaTeX installation is present (and if the "pdflatex" command is available on the command line via Matlab's "system" command), the LaTeX files will also be compiled into PDF files.

The replication code is provided as a tar-ball containing four sub-directories:

- **hydeparkDataSPF** contains raw data files obtained from FRB-PHIL as well as matlab files for transforming the raw inputs into a set of mat files (one for each of the five variables). As part of the replication files, copies of these five mat files are also contained in the following two subdirectories.
- **hydeparkMCMC** contains various scripts and functions to perform real-time MCMC estimation of our baseline model as well as the various alternatives described in the paper and the appendix.
- **hydeparkTablesAndFigures** contains various scripts to collect results (as generated by code provided in the **hydeparkMCMC** directory) and produce tables and figures.
- **toolbox** contains various folders providing different auxiliary m-files (Matlab scripts and functions) used throughout. The toolboxes are automatically loaded onto the Matlab path upon invocation of any of the scripts contained in the previously described code directories. Please note that some toolbox files were obtained either from the Matlab file exchange[7] or James P. Le Sage econometrics toolbox[8], which are both freely available; please see the comment headers of the respective toolbox files for further attribution.

When unpacking the tar-ball, these sub-directories should be copied into a common directory.

To prepare input data files for estimation

The directory **hydeparkDataSPF** contains all of the raw data files obtained from the RTDRC and FRED described above. In addition, the directory contains two m-file scripts to transform the raw data input input files for further processing by our main estimation routines contained in **hydeparkMCMC**; both m-file scripts create mat data files in Matlab format.

To process raw data for RGDP and PGDP (which are matched against realized values collected from the RTDRC), please run *hydeparkCollectDataGDP.m*. For the other three variables (CPI, UNRATE, TBILL) please run *hydeparkCollectData.m*. For each variable, a data file is created and stored in Matlab's mat format. (Resulting mat files are also provided as part of the replication set and stored in the **hydeparkMCMC** directory.)

Prior to running these two scripts, some of the Excel xlsx files provided by the RTDRC need to be converted into csv format. Specifically, the DATA sheets contained in the Excel files p_first_second_third.xlsx and routput_first_second_third.xlsx need to be stored as separate csv files. Before storing the data sheets in CSV format, entries of "NA" need to be changed to "-999" and headers should be removed from each DATA spreadsheet. (The resulting csv files are also provided as part of the replication set.) The SPF files "*Mean_XXX.xlsx*" need not be changed prior to processing by Matlab and can be used as downloaded from the RTDRC.

In case of updating the data, please update the definition of data vectors in the scripts *hydeparkCollectData.m* and *hydeparkCollectDataGDP.m* as indicated by comments therein (see lines 33 and 35/50, respectively).

To estimate the various models

Code for estimating the various model variants considered in paper and appendix is provided in **hydeparkMCMC**. In addition, as part of the replication set, **hydeparkMCMC** contains copies of the input data's mat files as created in **hydeparkData**. When updating the data, please copy updated mat files into the MCMC directory.

Run the following Matlab scripts:

- *hydeparkETAsv.m* estimates the baseline SV model described in the main paper. The script loops over all five SPF variables and multiple estimation windows as required for the real time estimation of the model. (This script creates result files both for our baseline choice of an evaluation window starting after 60 quarters as well as the alternative choice of an evaluation window starting only after 80 quarters.)
- *hydeparkFEconst.m* computes the alternative FE-SIMPLE model for various

estimation and evaluation windows.

- *hydeparkETAsvSinglefactor.m* computes a variant of the baseline model that uses a single-factor model for the SV processes.
- *hydeparkETAsvar1.m* computes a variant of the baseline model that estimates an AR(1) model for the log-variances of each SV process.
- *hydeparkETAconst.m* computes a variant of the baseline model that assumes constant variances and rolling estimation windows ("ETA-SIMPLE").
- *hydeparkETAvarsv.m* computes the ETA-VAR-SV model, which models the vector of forecast updates with a VAR-SV model.
- *hydeparkFEvarsv.m* computes the FE-VAR-SV model, which models the vector of forecast errors with a VAR-SV model. The script *hydeparkFEvarIC.m* computes various lag-length selection criteria for this purpose.
- *hydeparkETAJOINTsv.m* and *hydeparkETAJOINTsvSinglefactor.m* estimate a joint model of UNRATE, RGDP and PGDP using the baseline SV specification and the single-factor SV specification, respectively.

General notes:

- Each estimation script generates various figures as well as screen output of results. Tables and figures as shown in the published paper are also compiled via the scripts contained in the **hydeparkTablesAndFigures** directory.
- Computation of the real-time estimates is a massively parallel problem, since each real-time jump-off requires a separate MCMC estimation. To speed up the computation, the code loops over the real-time runs using Matlab *parfor* loops, which are executed in parallel if the Parallel Computing Toolbox is available and a set of parallel workers is available.
  - Whether a pool of parallel workers in used depends in part also on user settings specified in Matlab's preferences. Ideally, a user wanting to use a parallel pool should initialize the pool with Matlab's *parpool* command *prior* to executing our code.
  - When the Parallel Computing Toolbox is available on a user's machine, a corresponding section in Matlab's preference menu allows the user to enable automatic creation of a parallel when needed. If that option is enabled, our code will try to create a parallel pool, but not if otherwise. Please see the function *getparpoolsize* that is contained in **toolbox/emtools** for further details.
- In case of parallelization, separate random number streams will be used for each parallel worker. As a consequence, replication of the MCMC computations will invariably result in marginally (though not significantly) different

results when done using different computational setups.

- Most scripts contain a boolean variable *quicky*, which should be set to false for production quality results. If *quicky* is set to true, the code generates results only for very short MCMC chains and typically only for one variable (instead of looping over all five variables).
- Our code relies on a number of additional toolbox files — mostly developed by the authors — that are provided in a separate directory **toolbox**. As part of running any of the MCMC routines listed above, the Matlab path is reset and the **toolbox** directory and its subdirectories are automatically added to the path.
- Estimation output — figures as well as data files — are stored in a separate directory. By default, figures are stored in a subdirectory of **hydeparkMCMC** called **tmp** (and newly created if necessary); this can be changed by editing the script *localtemp.m* (provided as part of **toolbox/emtools**). Matlab data files containing MCMC results are stored in a directory defined by *localstore.m* (provided as part of **toolbox/emtools**), whose default choice is **tmp/resultfiles**.

To generate tables and figures

The directory **hydeparkTablesAndFigures** provides scripts to generate LaTeX tables and figures (as shown in our paper and the appendix). These scripts assume that MCMC results generated by the code in **hydeparkMCMC** and stored in mat-file format are contained in a directory called **resultfiles** and located one level above **hydeparkTablesAndFigures**. This setting can be changed by editing the *datadir* variable (as well as variants like *datadirFECONST*) in the various scripts contained in **hydeparkTablesAndFigures**. Assuming all of the above-mentioned model variants have been estimated, a full set of tables can be created by invoking *generateAllTables.m*.

Figures of the data can be generated by *figuresETA.m*, figures comparing forecast error and expectational updates against one-standard-deviation bands by *figuresSVbands.m*, and fan charts by *figuresFanCharts.m*.

---

1. Corresponding author: Elmar Mertens, em@elmarmertens.com

2. https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/
3. https://fred.stlouisfed.org
4. https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files
5. https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/real-time-data/data-files/files/xlsx/p_first_second_third.xlsx
6. https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/real-time-data/data-files/files/xlsx/routput_first_second_third.xlsx
7. https://www.mathworks.com/matlabcentral/fileexchange/
8. https://www.spatial-econometrics.com