README - MATLAB Code[*]

Victor Sellemi[†]

May 26, 2021

Overview

This file contains details on implementing sample MATLAB codes for macroeconomic forecasting. The main files are contained in the `matlab` directory and are named main_<model name>.m. These files require necessary helper functions contained in the `util` subdirectory and data files contained in the `data` subdirectory. Users should also ensure that they have installed the `Econometrics` and `Statistics and Machine Learning` Toolboxes with MATLAB version `R2021a+`.

Data

For these examples, we use the FRED-MD monthly panel of US macroeconomic and financial variables from McCracken and Ng (2016). Our version of these data span between 1959:01 - 2021:02 and contain 134 predictors. The folder `fred_md_preprocessing` contains code from the paper to (i) implement data transformations to achieve stationarity of all series and (ii) estimate factors using Principal Components Analysis (PCA) on the panel.

For ease of implementation, the default setting in our forecasting models is to form predictors as a subset of the first 8 PCA factors and their lags.[1] These factors explain over 50% of total variation in FRED-MD panel. The default is also to include an autoregressive component in the forecasting target. The folder `data` contains the FRED-MD stationary-transformed data and the file containing the first 8 PCA factors.

Our main files allow for five different forecasting targets as in Coulombe et al. (2020): industrial production (INDPRO), unemployment rate (UNRATE), consumer price index (CPI), difference between 10-year treasury rate and federal funds rate (SPREAD), and housing starts (HOUST). The main files allows users to specify the forecasting target and forecasting horizon by defining the `YY` and `h` objects, respectively.

---

[*]Related code libraries in Python, MATLAB, and R can also be found at `github.com/vsellemi`

[†]PhD student at UC San Diego. Email: `vsellemi@ucsd.edu`

[1]This can be changed by modifying the definition of "factors" in the main files.

Main Files

The main files for the forecasting models are named `main_<model name>.m` and implement selection, estimation, and evaluation steps for each model.

Settings

These files allow users to specify the following settings.

1. Forecasting target and horizon: `YY` and `h`.

2. Model selection method: `model_selection`. We split data into three equally sized sequential subsets (training, validation, and testing samples) and allow users to choose from 4 model selection options. The options include:

   - `ictr`: maximizing in-sample information criteria (BIC, AIC, HQ) calculated on the training sample for models estimated on the training sample
   - `icval`: maximizing in-sample information criteria (BIC, AIC, HQ) calculated on the validation sample for models estimated on the training sample. The specific choice of information criteria can be specified via `icidx`.
   - `pooscv`: pseudo out-of-sample cross-validation with an expanding window. This option minimizes pseudo out-of-sample forecasting error on the validation sample. This is the most computationally intensive option.
   - `kfcv`: K-fold cross-validation. This option splits the combined validation and training samples into K+1 equally size sequential subsets (folds) and minimizes the average forecasting error in fold $i+1$ from models estimated on folds 1 to i where $i = 1, ..., 5$. The number of folds is specified by `K`.

   The default setting here is `icval` with Bayesian Information Criteria (BIC).

3. Out-of-sample procedure: users can specify if they want to do out-of-sample forecasting using a rolling or expanding window. The default is an expanding window forecast but a rolling window can be specified by setting `roll` equal to 1. Window length is set with `wL`.

4. Evaluation benchmark: users can specify the forecasting benchmark used in evaluation. The default is a random walk (`RW`) forecast but prevailing mean (`PM`) is also supported.

5. Hyperparameter space: each forecasting model is associated with a set of hyperparameters (HPs). Model selection iterates over all possible combinations of hyperparameters and users can define this set (e.g., maximum number of AR lags, maximum number of factors).

Evaluation

The main files also provide several standard point forecast evaluation metrics, namely Diebold and Mariano (1995) test, White (2000) and Hansen (2005) p-values, and the Hansen et al. (2011) model confidence set. These tests are implemented relative to the specified benchmark. We also provide code to implement the Mincer and Zarnowitz (1969) forecast evaluation test. Lastly, users generate time-series plots of the forecast, cumulative mean squared prediction error, and rolling root mean squared prediction error of the model and benchmark.

Models[2]

1. `main_AR.m`: autoregressive model

   HPs: lag length (`py`)

2. `main_ARDI.m`: factor augmented autoregressive model

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`)

3. `main_PLS.m`: partial least squares regression

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`)

4. `main_ENET.m`: elastic net regression

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`), lasso penalty (`lambda`), weight of lasso versus ridge penalty (`alpha`)

5. `main_LASSO.m`: lasso regression

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`), lasso penalty (`lambda`)

6. `main_KRR.m`: kernel ridge regression

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`), ridge penalty (`lambda`), kernel function (`kernel`), Gaussian (RBF) kernel parameter (`sigma`)

7. `main_SVR.m`: support vector regression

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`), kernel function (`kernel`), Gaussian (RBF) kernel parameter (`sigma`), SVR optimization parameters (`eps_max`, `C_max`)

8. `main_RF.m`: random forest ensemble regression

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`), bagging or boosting for aggregation (`method`), number of ensemble learning cycles (`numCycles`), resampling procedure (`resamp_ind`, `replac_ind`)

9. `main_NN.m`: feed-forward neural network regression

   HPs: AR lag length (`py`), factor lag length (`pf`), number of factors (`nf`), hidden unit activation (`hidden_activation`), output activation function (`output_activation`), optimizer (`solver`), hidden dimension (`hidden_dim`), number of layers (`nlayers`)

10. `main_TVPSV.m`: time-varying parameter stochastic volatility model

    HPs: see Pettenuzzo and Timmermann (2015)

11. `main_MS.m`: Markov-Switching model

    HPs: see Pettenuzzo and Timmermann (2015)

12. `main_CSR.m`: complete subset regression

    HPs: see Elliott et al. (2013)

[2]See Stock and Watson (2002) for details on model 2 and Hastie et al. (2001) or Coulombe et al. (2020) for more details on models 3-9.

References

P. G. Coulombe, M. Leroux, D. Stevanovic, and S. Surprenant. How is machine learning useful for macroeconomic forecasting?, 2020.

F. Diebold and R. Mariano. Comparing predictive accuracy. *Journal of Business Economic Statistics*, 13(3):253–63, 1995. URL `https://EconPapers.repec.org/RePEc:bes:jnlbes:v:13:y:1995:i:3:p:253-63`.

G. Elliott, A. Gargano, and A. Timmermann. Complete subset regressions. *Journal of Econometrics*, 177(2):357–373, 2013. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2013.04.017. URL `https://www.sciencedirect.com/science/article/pii/S0304407613000948`. Dynamic Econometric Modeling and Forecasting.

P. R. Hansen. A test for superior predictive ability. *Journal of Business Economic Statistics*, 23(4):365–380, 2005. ISSN 07350015. URL `http://www.jstor.org/stable/27638834`.

P. R. Hansen, A. Lunde, and J. M. Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011. doi: https://doi.org/10.3982/ECTA5771. URL `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA5771`.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

M. W. McCracken and S. Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016. doi: 10.1080/07350015.2015.1086655. URL `https://doi.org/10.1080/07350015.2015.1086655`.

J. A. Mincer and V. Zarnowitz. *The Evaluation of Economic Forecasts*, pages 3–46. NBER, 1969. URL `http://www.nber.org/chapters/c1214`.

D. Pettenuzzo and A. Timmermann. Forecasting macroeconomic variables under model instability. *Journal of Business Economic Statistics*, 35:0–0, 06 2015. doi: 10.1080/07350015.2015.1051183.

J. Stock and M. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business Economic Statistics*, 20(2):147–62, 2002. URL `https://EconPapers.repec.org/RePEc:bes:jnlbes:v:20:y:2002:i:2:p:147-62`.

H. White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000. doi: https://doi.org/10.1111/1468-0262.00152. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00152`.