Replication of
De Mol et al. (2008, JoE)
Roadmap

Matteo Barigozzi

Università di Bologna
matteo.barigozzi@unibo.it

May 15, 2025

1. Download monthly data FRED-MD up to April 2025 from https://www.stlouisfed.org/research/economists/mccracken/fred-databases.

2. Remove rows from January 2020 to July 2020 from the csv file.

3. Run the code downloaded with data to transform to stationarity, remove outliers, and fill in missing values. Check autocorrelations for raw and transformed data for some variables.

4. Define T as the length of the training sample so $T \simeq 2/3$ of total number of time periods.

5. Let $\mathbf{Z}$ be the $T \times N$ training sample and copy the column of $\mathbf{Z}$ that we want to forecast and call it $\mathbf{w}$ also of length T.

6. Center and standardize all columns of $\mathbf{Z}$ denoted as $\mathbf{z}_i$, save the N means in $\mathbf{M}_Z = (M_{1z} \cdots M_{NZ})'$ which is $N \times 1$ and the std.dev. as $\mathbf{S}_Z = (S_{1z} \cdots S_{NZ})'$ which is $N \times 1$. Define

$$\mathbf{x}_i = \frac{\mathbf{z}_i - M_{Zi}}{S_{Zi}}, \quad i = 1, \dots, N.$$

   which is $T \times 1$, and

$$\mathbf{X} = (\mathbf{x}_1 \cdots \mathbf{x}_N)$$

   which is $T \times N$.

7. Center and standardize $\mathbf{w}$, save the mean in M_w which is 1×1 and the std.dev. as S_w which is 1×1. Define,

$$\mathbf{y} = \frac{\mathbf{w} - M_w}{S_w}$$

   which is $T \times 1$.

8. The models to be estimated are

   (a) AR(1)

$$y_t = \beta_0 + y_{t-1}\beta_1 + e_t, \quad t = 2, \dots, T$$

   estimate via OLS and forecast as

$$\widehat{y}_{T+1|T} = \widehat{\beta}_0 + y_T\widehat{\beta}_1$$

(b) AR(p), with $p > 1$

$$y_t = \beta_0 + y_{t-1}\beta_1 + \ldots + y_{t-p}\beta_p + e_t, \quad t = p+1, \ldots, T$$

use BIC to find p then estimate via OLS with optimal p and forecast as

$$\widehat{y}_{T+1|T} = \widehat{\beta}_0 + y_T\widehat{\beta}_1 + \ldots + y_{T-p+1}\widehat{\beta}_p$$

(c) Random walk

$$y_t = \beta_0 + y_{t-1} + e_t, \quad t = 2, \ldots, T$$

estimate via OLS (careful here!) and forecast as

$$\widehat{y}_{T+1|T} = \widehat{\beta}_0 + y_T$$

(d) Multivariate (note that $\mathbf{X}_{t-1}$ contains y_{t-1})

$$y_t = \beta_0 + \mathbf{X}'_{t-1}\boldsymbol{\beta} + \boldsymbol{e}_t, \quad t = 2, \ldots, T$$

i. estimate via OLS (here $N < T$ so it can be done)
ii. estimate via ridge use BIC to find penalization
iii. estimate via lasso use BIC to find penalization

and forecast as

$$\widehat{y}_{T+1|T} = \widehat{\beta}_0 + \mathbf{X}'_T\widehat{\boldsymbol{\beta}}$$

(e) Factor, run PCA on $\mathbf{X}$ to retrieve r PCs $\widehat{\mathbf{F}}_t$ which is $r \times 1$ use BIC to find number of factors r. Then

$$\widehat{y}_t = \beta_0 + \widehat{\mathbf{F}}'_{t-1}\boldsymbol{\beta} + \boldsymbol{e}_t, \quad t = 2, \ldots, T$$

estimate via OLS and forecast as

$$\widehat{y}_{T+1|T} = \widehat{\beta}_0 + \widehat{\mathbf{F}}'_T\widehat{\boldsymbol{\beta}}$$

9. The final forecast is

$$\widehat{w}_{T+1|T} = M_w + S_w * \widehat{y}_{T+1|T}$$

for all models.

10. Forecast the target variable in levels.

(a) If the variable to forecast is Industrial production IPI then $w_t = \Delta \log IPI_t$ so

$$\widehat{w}_{T+1|T} = \log \widehat{IPI}_{T+1|T} - \log IPI_T$$

so the forecast is

$$\widehat{IPI}_{T+1|T} = \exp\left\{\log IPI_T + \widehat{w}_{T+1|T}\right\}$$

(b) If the variable to forecast are prices CPI then $w_t = \Delta\pi_t$ with $\pi_t = \Delta \log CPI_t$ so

$$\widehat{w}_{T+1|T} = \widehat{\pi}_{T+1|T} - \pi_T = \log \widehat{CPI}_{T+1|T} - \log CPI_T - \log CPI_T + \log CPI_{T-1}$$

so the forecast is

$$\log \widehat{CPI}_{T+1|T} = \widehat{w}_{T+1|T} + 2\log CPI_T - \log CPI_{T-1}$$

11. Repeat 5 to 10 for an expanding window of training data points, so if T_1 is the total number of observations repeat using data up to $\tau = T, \dots, T_1 - 1$. This will give a series of $T_1 - T$ forecasts for each model.

$$\widehat{IPI}_{T+1|T}, \dots, \widehat{IPI}_{T_1|T_1-1}, \qquad \widehat{CPI}_{T+1|T}, \dots, \widehat{CPI}_{T_1|T_1-1}$$

12. Compute RMSE for each model

$$RMSE_{IPI} = \sqrt{MSE_{IPI}}, \quad MSE_{IPI} = \frac{1}{T_1 - T} \sum_{\tau=T}^{T_1-1} \left(\widehat{IPI}_{\tau+1|\tau} - IPI_{\tau+1}\right)^2$$

and

$$RMSE_{CPI} = \sqrt{MSE_{CPI}}, \quad MSE_{CPI} = \frac{1}{T_1 - T} \sum_{\tau=T}^{T_1-1} \left(\widehat{CPI}_{\tau+1|\tau} - CPI_{\tau+1}\right)^2$$

13. Compare and discuss.