ALMA MATER STUDIORUM
UNIVERSITY OF BOLOGNA

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

Mehmet Aydin [1], Michele Salvi [2], Francesco Prem Solidoro [3]

[1] University of Bologna, mehmet.aydin@studio.unibo.it
[2] University of Bologna, michele.salvi4@studio.unibo.it
[3] University of Bologna, francesco.solidoro@studio.unibo.it

The authors solemnly declare that no generative AI was used in creating code or producing text for this report.

1. Data

We used the monthly data from FRED-MD[1], going from January 1959 up to April 2025. It's a large macroeconomic database designed for empirical analysis of big data, reporting monthly and quarterly observations of many macroeconomic factors, such as indicators for industrial production (IPI) and inflation through the consumer price index (CPI).

1.1. Cleaning the Data

In order to use the data, we needed to make some adjustments:

- Covid months are tremendous outliers born off of an exogenous shock, which would offset most of our predictions in useless ways
- We needed the data to be stationary, to remove eventual outliers, and fill in missing values
- We needed to standardize the data

1.1.1. Covid

For the Covid months, we opted to remove them rather than try to control their effect. While this reduced the amount of data points we had, it improved the quality of our analysis which, being based on autocorrelative properties, would have been perverted by the strong exogenous shock of Covid.

1.1.2. Stationarization and Missing Values

Given that non-stationary data implies bias in time series estimation, we used the stationarized version of our original dataset. We also removed missing values from our dataset to permit our analyses.

1.1.3. Sampling and Standardization

We picked a sample size of length $T \simeq \frac{2}{3}N$, where N is the number of observations.

We then created a centered and standardized version of both the dataset and the training set by centering the data (by substracting the mean from it) and scaling it in relation to its standard deviation (by dividing it by its standard deviation), so we could pull vectors from them freely.

Taking the standardized training set, we can now extract a vector of all the values of the variable we wish to predict to formulate our models. In our case, we're going to extract IPI, from the INDPRO column, and CPI, from the PCEPI column.

2. Methods

As part of our report we tested different specifications of five models: two univariate models, three multivariate models, and one principal component model. The univariate models consist of an autoregressive model and a random walk model, whereas the multivariate models consist of an ordinary least squares (OLS) model, a ridge model, and a lasso model. In models that required hyperparameter tunings we used the Bayesian information criterion (BIC). The graphs regarding the BIC may be consulted in Section 4.1.

2.1. Univariate

In these models, the independent variables are just the lagged observations of the dependent variable. In this section, for each $y_{t-i}, i = 0, 1, ..., T-1$ we used the observations from the $p-i+1$-th to the $T-i$-th.

2.1.1. Autoregressive Model (AR)

We have estimated autoregressive models as below for all the possible number of lags p.

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_t, \forall p = 1, 2, ..., T-1$$

In order to select p, we calculated the BIC for all p models. Afterwards, we started from $p = 1$ and increased the number of lags until the BIC for the models stopped decreasing and increased. We chose the p as the one that gives the local minimum BIC. In other words, the p we selected is such that satisfies the equations below where BIC is taken as a function of the number of lags.

$$\mathrm{BIC}(i) > \mathrm{BIC}(i+1), \forall i = 1, 2, ..., p-1.$$
$$\mathrm{BIC}(p+1) > \mathrm{BIC}(p)$$

We calculated the BIC for the autoregressive model as below, where $n = T - p$ is the number of observations.

$$\mathrm{BIC} = \ln \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} + (p+1)\frac{\ln n}{n},$$
$$\hat{\varepsilon}_t = y_t - \hat{\beta}_0 - \sum_{i=1}^{p} \beta_i y_{t-i}$$

2.1.2. Random Walk (RW)

We have estimated the random walk model as below.

$$y_t = \beta_0 + y_{t-1} + \varepsilon_t$$

In order to estimate the model, we first created a column vector of differences of length $T-1$ every i-th row of which composed of the $t-i+1$-th value minus the $t-i$-th. Then, we ran an estimation with this vector as the dependent variable and a column vector of ones ι as the independent variable. Our estimated coefficient $\hat{\beta}_0$ was the mean of the values in our vector of differences.

2.2. Multivariate

In these models the independent variables are the once-lagged observations of all variables in the data set including the dependent variable. In this section, as the observations are lagged once, for y_t we used the observations from the 2nd to the T-th. For $\boldsymbol{X_{t-1}}$, on the other hand, we used the observations from the 1st to the $T-1$-th.

2.2.1. Ordinary Least Squares (OLS)

We estimated the ordinary least squares (OLS) model as below.

$$y_t = \beta_0 + \boldsymbol{X_{t-1}}\beta + \varepsilon_t,$$
$$\hat{\beta} = (\boldsymbol{X'_{t-1}}\boldsymbol{X_{t-1}})^{-1}\boldsymbol{X'_{t-1}}Y_t$$

2.2.2. Ridge

We estimated the ridge model as below.

$$\hat{\beta} = \arg\min(\hat{\varepsilon}'_t\hat{\varepsilon}_t + \lambda\beta'\beta)$$
$$= (\boldsymbol{X'_{t-1}}\boldsymbol{X_{t-1}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X'_{t-1}}Y_t,$$
$$\hat{\varepsilon}_t = y_t - \hat{\beta}_0 - \boldsymbol{X_{t-1}}\hat{\beta}$$

In order to select λ, we calculated the BICs for $\lambda \in [0.001, 10000]$ and selected the λ that had the lowest BIC associated to it. We calculated the BIC for the ridge model as below, where $n = T - 1$ is the number of observations and the effective degrees of freedom d is $\mathrm{tr}\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}'\right)$ and $\hat{\sigma}^2_\varepsilon$ is the estimated variance of the residuals.

$$\mathrm{BIC} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} + \frac{\ln n}{n} d\hat{\sigma}^2_\varepsilon$$

2.2.3. Lasso

We estimated the lasso model as below.

$$\hat{\beta} = \arg\min\left(\hat{\varepsilon}'_t\hat{\varepsilon}_t + \lambda\sum_{i=1}^{p}|\beta_i|\right),$$
$$\hat{\varepsilon}_t = y_t - \hat{\beta}_0 - \boldsymbol{X_{t-1}}\hat{\beta}$$

In order to select λ, we calculated the BIC for $\lambda \in [0.001, 1]$ and selected the λ that had the lowest BIC associated to it. We calculated the BIC for the lasso model as below, where $n = T - 1$ is the number of observations and p is the number of regressors with non-zero coefficients.

$$\mathrm{BIC} = \ln \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} + p\frac{\ln n}{n}$$

2.3. Principal Component Regression (PCR)

We estimated the principal component regression as below where $\boldsymbol{\hat{F}_{t-1}}$ is $\boldsymbol{X_t}$ multiplied by the matrix containing r eigenvectors with the highest eigenvalues of the matrix $\boldsymbol{X'_t X_t}$, divided by $\sqrt{T}$ and its last row removed.

$$\hat{y}_t = \beta_0 + \boldsymbol{\hat{F}_{t-1}}\beta + \varepsilon_t,$$
$$\hat{\beta} = \left(\boldsymbol{\hat{F}'_{t-1}}\boldsymbol{\hat{F}_{t-1}}^{-1}\right)\boldsymbol{\hat{F}'_{t-1}}Y_t$$

In order to select r, we calculated the BIC for $r = 1, 2, ..., p$ and selected the r that had the lowest BIC associated to it. We calculated the BIC for the principal component regression model as below where $n = T - 1$ is the number of observations.

$$\mathrm{BIC} = \ln \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} + (r+1)\frac{\ln n}{n}$$

2.4. Forecasting

All the aforementioned models were used to forecast for $i = 1, 2, ..., N - T$, i.e., the entire length of the test dataset. All forecasts were conditional on the immediately previous observation. Given that the data used for the estimates was standardized, the forecasts were first multiplied by the standard deviation and then summed with the mean of the respective dependent variable obtained prior to standardization. The resulting forecasts

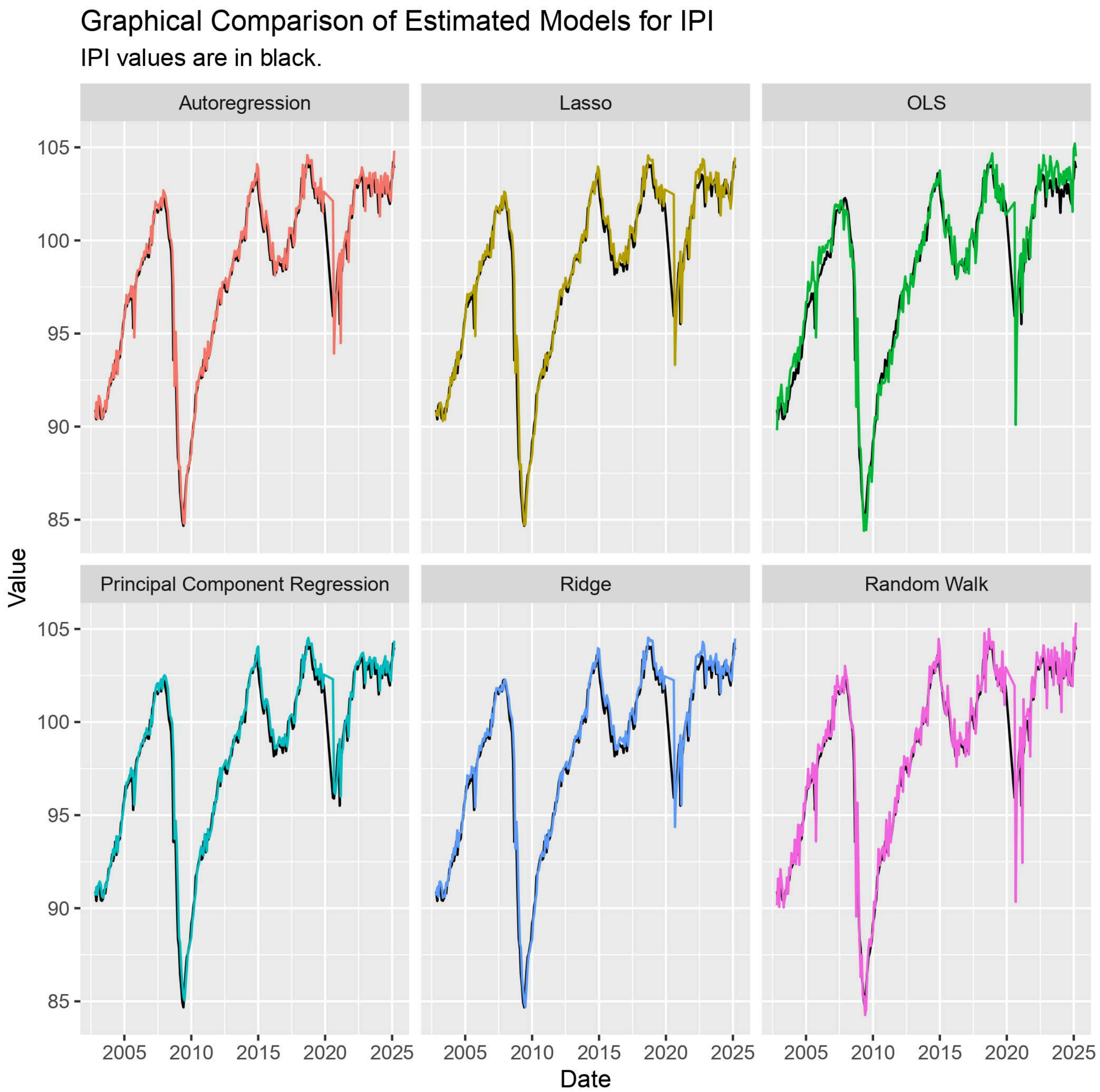

Figure 1: Comparison of estimated models against true IPI values

$\hat{w}_{T+i|T+i-1}$ were then leveled using the original dataset as below.

$$\widehat{\text{IPI}}_{T+i|T+i-1} = \exp\left(\hat{w}_{T+i|T+i-1} + \ln \text{IPI}_{T+i-1}\right)$$

$$\widehat{\text{CPI}}_{T+i|T+i-1} = \frac{\exp\left(\hat{w}_{T+i|T+i-1} + 2\ln \text{CPI}_{T+i-1}\right)}{\text{CPI}_{T+i-2}}$$

3. Results

We forecasted on the models we estimated in the previous section on an expanding window of training data points, obtaining a complete series of forecasts for each model.

We then estimated the goodness of fit of the estimated models by taking the RMSE for both industrial production and the consumer price index.

3.1. Industrial Production

For industrial production, the observed RMSEs are displayed in Table 1.

AR	0.855	**RW**	1.11
OLS	1.04	**Ridge**	0.819
Lasso	0.876	**PCR**	0.813

Table 1: RMSE values for the models predicting industrial production

Figure 1 offers a comparison of the models estimating IPI.

The best performing one, according to RMSE, is PCR, followed by ridge, autoregression, and lasso. OLS's and random walk's inferior performance isn't surprising, due to the stochastic component of random walks and tendency to overfitting of OLS in such a complex dataset. Additionally the volatility of industrial production is damaging the forecasting ability of the autoregression model, thus explained the increased variance.

It is exactly the dataset's scope that characterizes the best models. In fact PCR, minimizing distances between observations and their projections, reduces dimensions, recasting the data along the one that explains the most variance. Industrial production, being quite volatile, benefitted of estimation from shrinkage models such as PCR and ridge, being essentially a penalized PCR that includes all principal components but with decreasing weights.

3.2. Consumer Price Index

For CPI the observed RMSE's are displayed in Table 2.

AR	0.192	**RW**	0.300
OLS	0.216	**Ridge**	0.208
Lasso	0.210	**PCR**	0.213

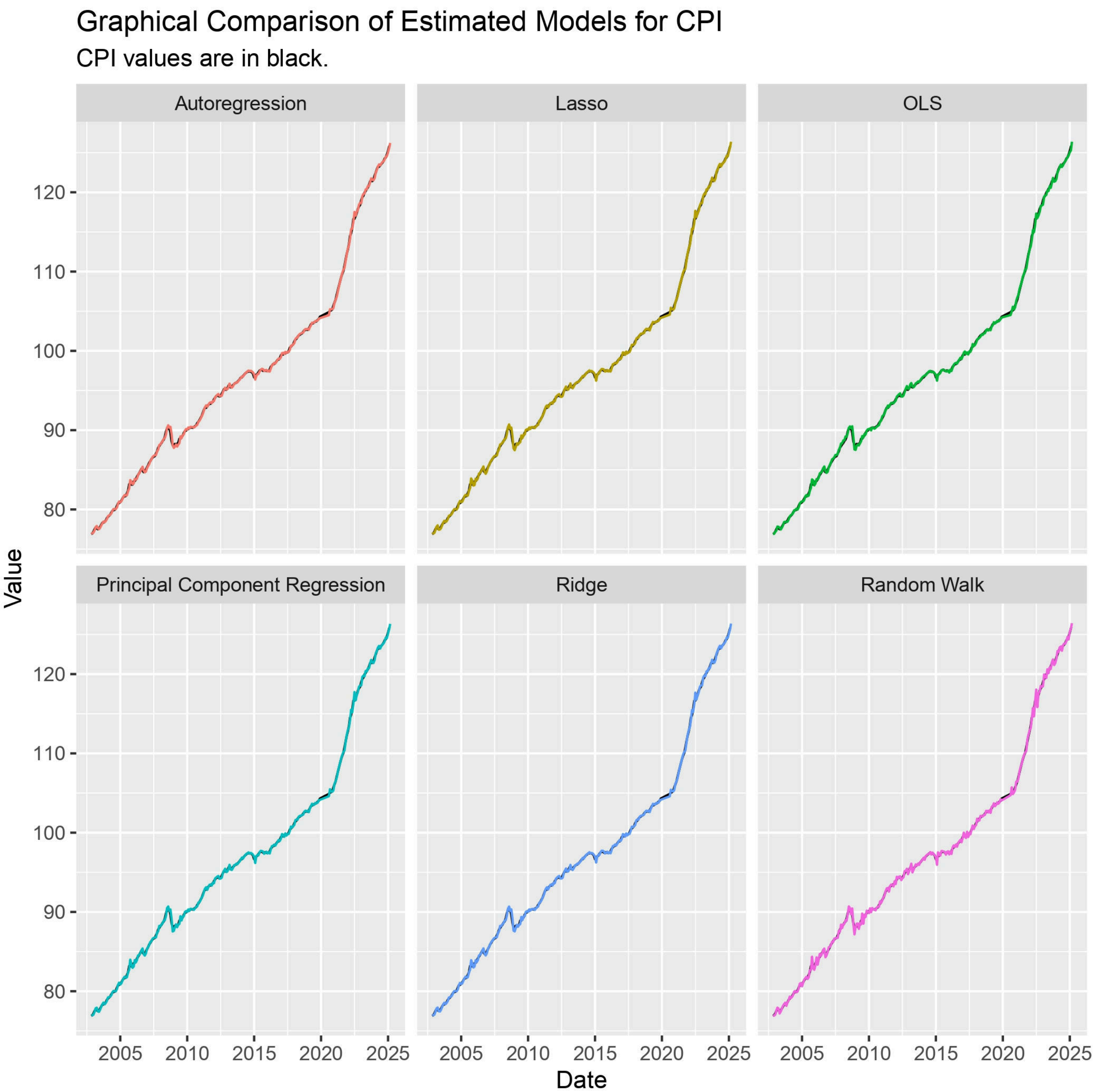

Figure 2: Comparison of estimated models against true CPI values

Table 2: RMSE values for the models predicting consumer price index

Figure 2 offers a graphical comparison of the models estimating CPI.

CPI, or better in this case PCEPI, due to its almost linear increase, presents a different situation to what was shown for industrial production. The remarks on stochastic components and overfitting made towards random walks and OLS still stand, but to a smaller degree, as the aforementioned lack of volatility in the data is captured by closer and smaller RMSEs.

The best performing model in this case is the autoregressive one, as the data points reveal a pattern that is almost steadily increasing, meaning that past values of CPI make good predictors for forecasts, which is exactly what the autoregression does, a regression against the data's own lagged values.

4. Appendix

4.1. Extra Graphs

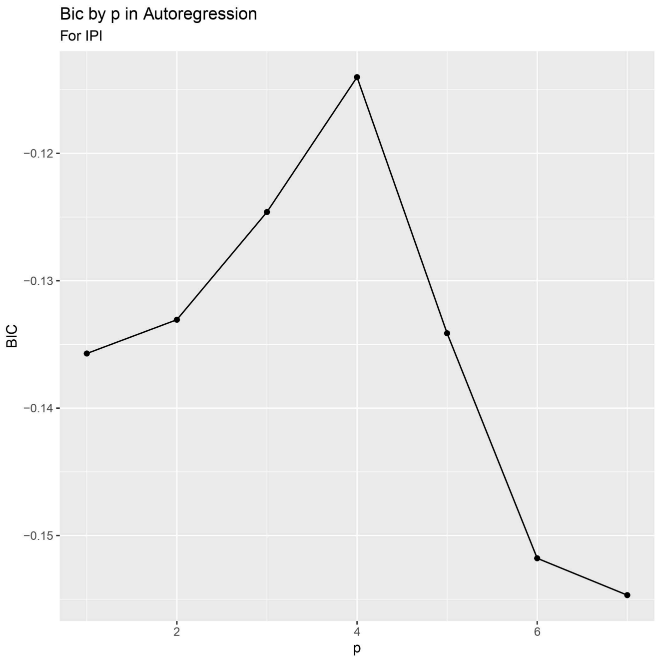

Figure 3: BIC by the number of lags for the IPI AR model

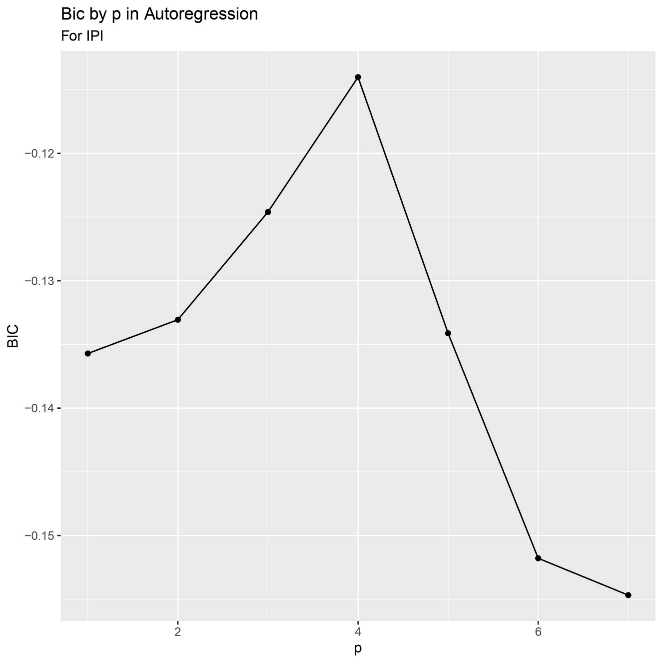

Figure 4: BIC by the number of lags for the CPI AR model

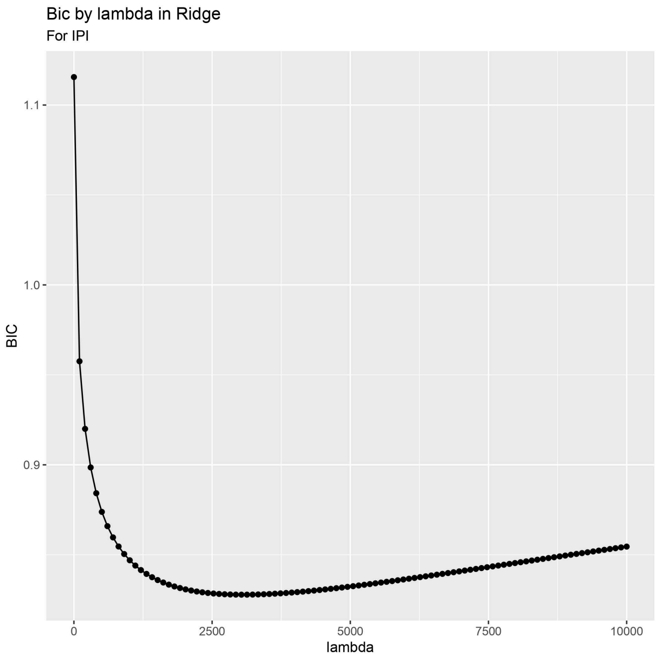

Figure 5: BIC by lambda for the IPI ridge model

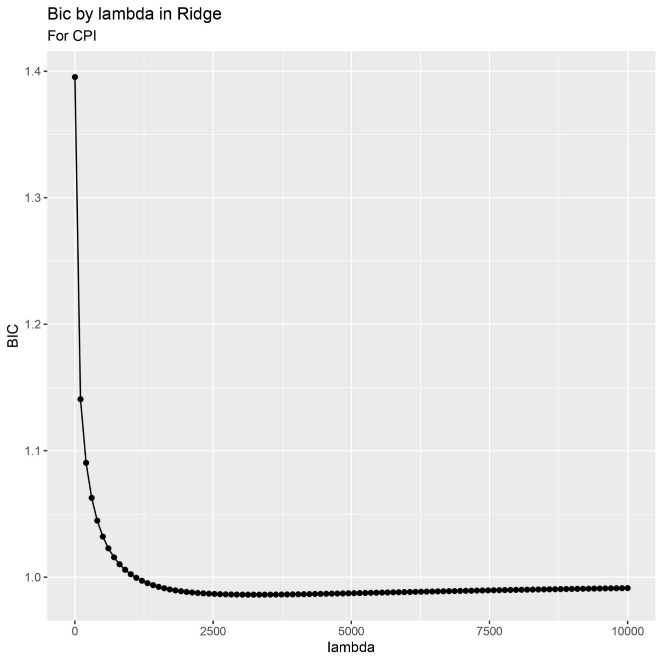

Figure 6: BIC by lambda for the CPI ridge model

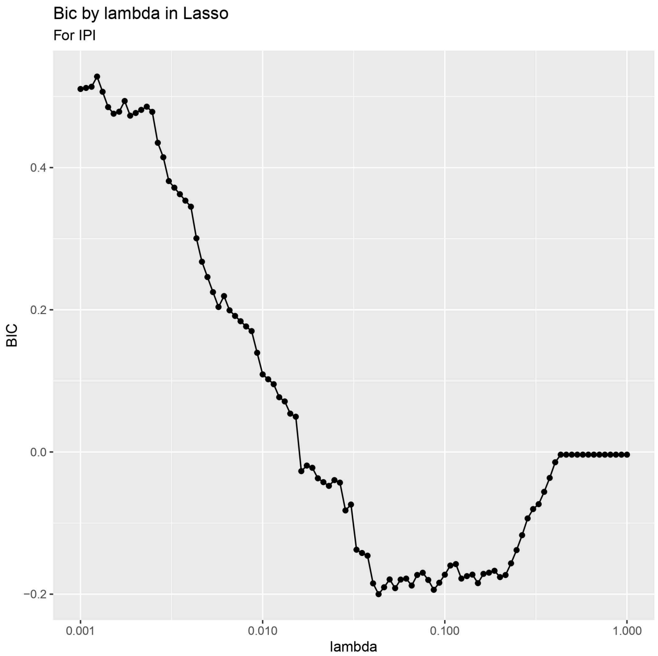

Figure 7: BIC by lambda for the IPI lasso model

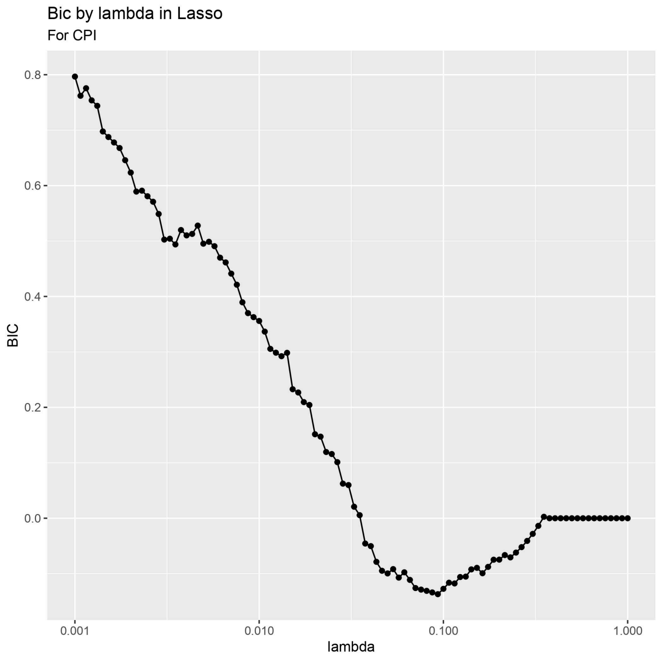

Figure 8: BIC by lambda for the CPI lasso model

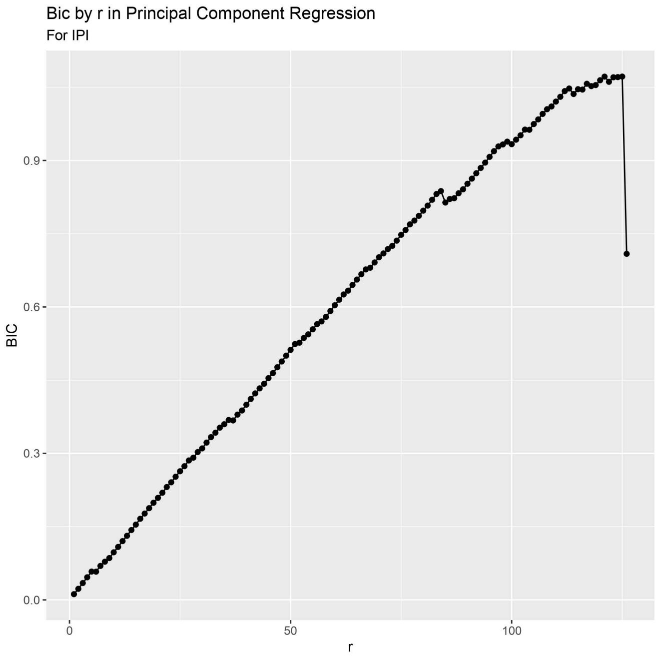

Figure 9: BIC by the number of principal components for the IPI PCR

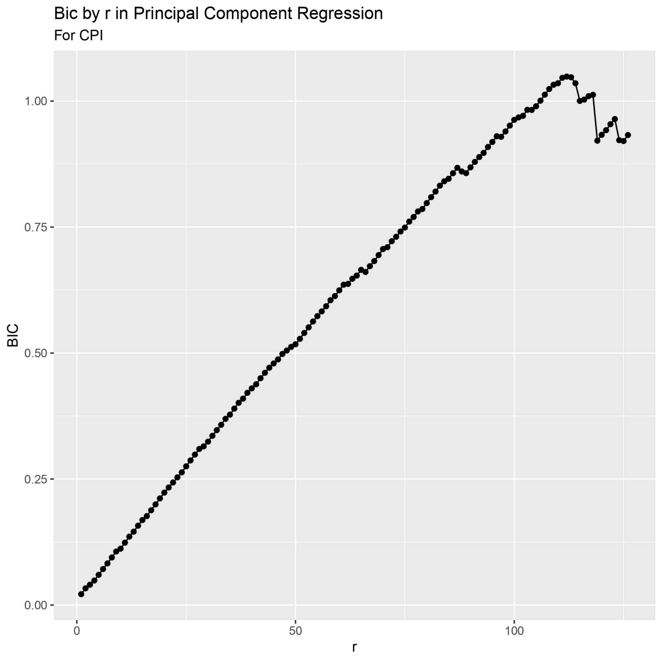

Figure 10: BIC by the number of principal components for the CPI PCR

4.2. The Code

The GitHub repository containing the R code used to perform the analyses and all the files can be reached by clicking here.

Bibliography

[1] "FRED-MD and FRED-QD: Monthly and Quarterly Databases for Macroeconomic Research." [Online]. Available: https://www.stlouisfed.org/research/economists/mccracken/fred-databases