Econometrics IV - Final Assignment

Caio Garzeri, Guilherme Luz, Guilherme Masuko

August 15, 2023

Apart from this sheet with our answers we are sending you another file *ECO4GLMCode.pdf* with a compilation of our code.

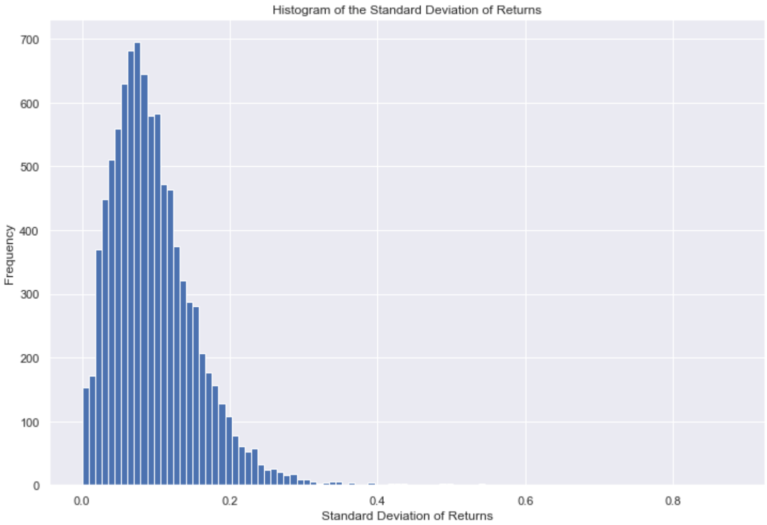
Question 1

We start by providing an overview of the data. The dataset consists of monthly excess returns from 9,456 firms traded in the New York Exchange Market, from November 1991 to December 2018. In addition to that we also have data on 16 monthly factors:

- Market (MKT) ; Small-minus-Big (SMB); High-minus-Low (HML); Conservative-minus-Aggressive (CMA); Robust-minus-Weak (RMW); Up-minus-down(UMD)
- Earning/Price ratio (EP); Cash-flow/Price ratio (CFP); Dividend/Price ratio (DY)
- Accruals (ACC); Net Share Issues (CHCSHO)
- Market Beta (BETA)
- Daily variance (RETVOL); Daily idiosyncratic variance (IDIOVOL)
- 1-month momentum (MOM1); 36-month momentum (MOM36)

Returns have relatively similar standard deviation (Figure 1). In following up with our computations we therefore use returns and factors just centered at their mean.

Figure 1: Standard Deviation of Excess Returns



Suppose we have the following model for excess returns:

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 X_{1t} + \cdots + \beta_p X_{nt} + U_t, \quad t = 1, \ldots, T \\
&= \boldsymbol{\beta}' \boldsymbol{X}_t + U_t \\
\boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U} \quad \text{(matrix notation)}.
\end{aligned}
$$

Because we have more than 9,000 variables and close to 360 observations, $n >>> T$ and the usual ordinary least squares (OLS) solution is obviously not valid.

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1} X'Y$$

The approach in this exercise will be postulating an alternative factor model:

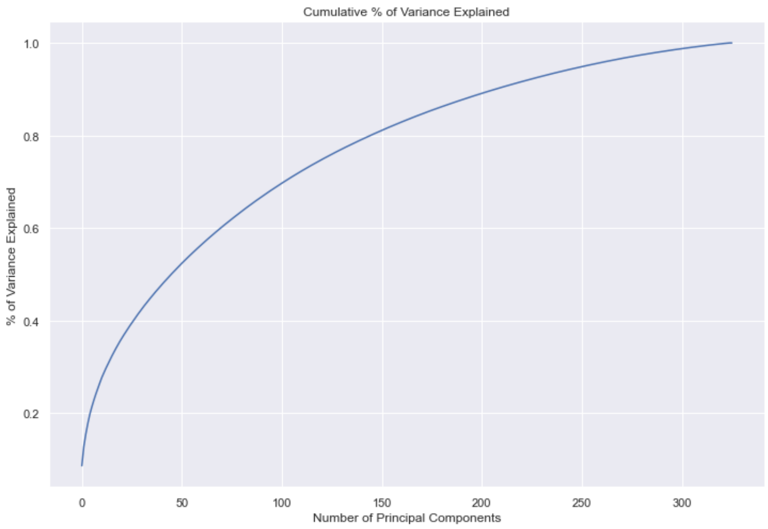
$$\underset{(n\times 1)}{\boldsymbol{X}_t} = \underset{(n\times k)}{\boldsymbol{\Lambda}} \underset{(k\times 1)}{\boldsymbol{F}_t} + \underset{(n\times 1)}{\boldsymbol{V}_t},$$

where: $\boldsymbol{F}_t$ is a set of $k << n$ unobserved factors; $\boldsymbol{V}_t$ is the vector of idiosyncratic errors and; $\boldsymbol{\Lambda}$ is the matrix of unobserved factor loadings.

Item a

We begin by computing the principal components of the returns and determining the optimal number of principal factors. Figure 2 displays the cumulative variance in returns explained by adding consecutive principal components. Explaining a lot of the variance requires a reasonable amount of components: with 100 components only 70% of the variance is explained.

Figure 2: Cumulative Variance



Choosing the number of factors in these models is a notoriously open issue in the literature (Bai and Ng, 2002). We evaluate below the optimal number of factors using three commonly used criteria.

- **Rule of Thumb**: Stop at a k such that the $(k+1)$-th PC does not add much to the already explained variance (say $< 3\%$).

  In this case we would choose only the first 2 PCs, which explain 12.82% of the variance.

- **Cutoff**: Choose the number of components such that a large portion (say 90%) of the variance is explained.

  Under this criterion the first 207 PCs would be chosen as they together explain 89.91% of the variance.

- **Biggest Drop**: Choose the biggest drop in the eigenvalues of the PCs, λ_j (Onatski, 2010), that is:

$$r := \arg \max_{1 \leq j < n} \frac{\lambda_j}{\lambda_{j+1}}.$$

In this case only the first PC would be used. After the first PC we observe the biggest drop (8.68%) in the amount of explained variance.

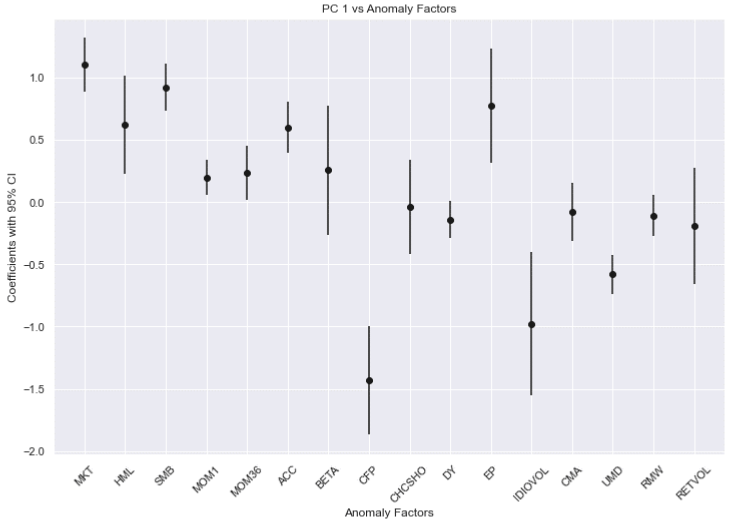
In summary for the rest of this question we will adopt the Rule of Thumb, which implies picking the first 2 PCs. Together they explain 12.82% of the variance.

Item b

We now regress the first 2 PCs on the 16 factors listed before. Factors have been standardized.

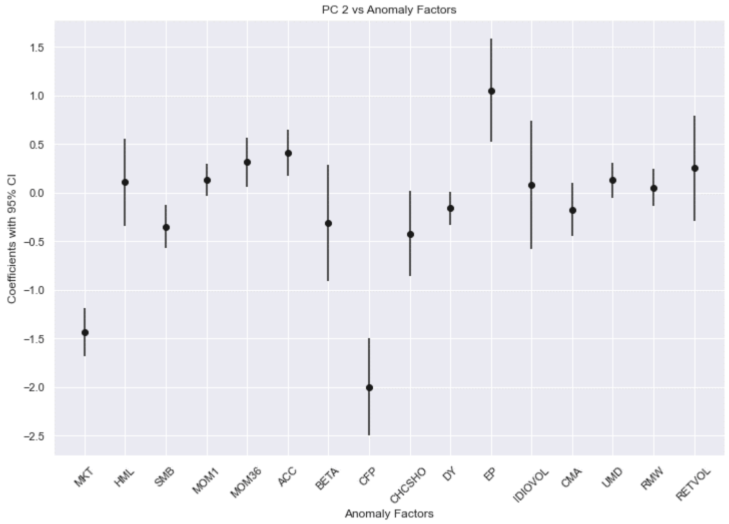
For the first PC (PC1) estimated coefficients along with their confidence interval are shown in Figure 3. One can see how this PC is strongly and positively correlated with the Market (MKT), High-minus-Low(HML), Small-minus-Big (SMB), Accruals (ACC) and Earning/Price Ratio factors. Most of the other factors are either weakly correlated or not statistically different from zero except for Cash-flow/price ratio (CFP) and Idiosyncratic volatility (IDIOVOL) which show significant and negative coefficients.

Figure 3: PC1 and Anomaly Factors



Repeating the exercise for PC2, we get a somewhat less clear result (Figure 4). Most of the factors are not related to PC2. For two factors the pattern is the same as before: Earning/Price Ratio is once again positively correlated with PC2 with a high coefficient while Cash-flow/price ratio is negatively correlated.

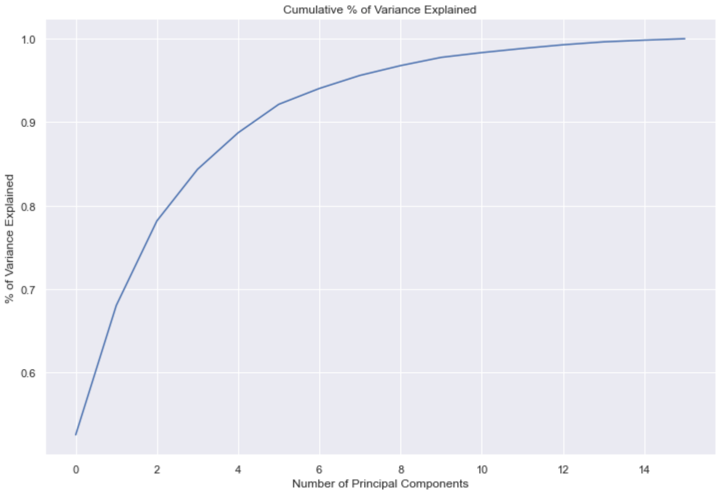
Figure 4: PC2 and Anomaly Factors



Item c

By running a principal component analysis on the 16 anomaly factors one can see how much of the variance is explained by incorporating new PCs (Figure 5). As we did before, we compute the optimal number of PCs under the three criteria.

Figure 5: Cumulative Variance - PCA on Anomaly Factors



- **Rule of Thumb**: In this case, we would choose the first 6 PCs, which explain 92.14% of the variance.
- **Cutoff**: Under this criterion the first 5 PCs would be chosen as they together explain 88.73% of the variance.
- **Biggest Drop**: In this case only the first PC would be used. After the first PC we observe the biggest drop (52.52%) in the amount of explained variance.

We will choose the optimal number of PCs using the rule of thumb again, which will give us 6 PCs explaining 92.14% of the variance.

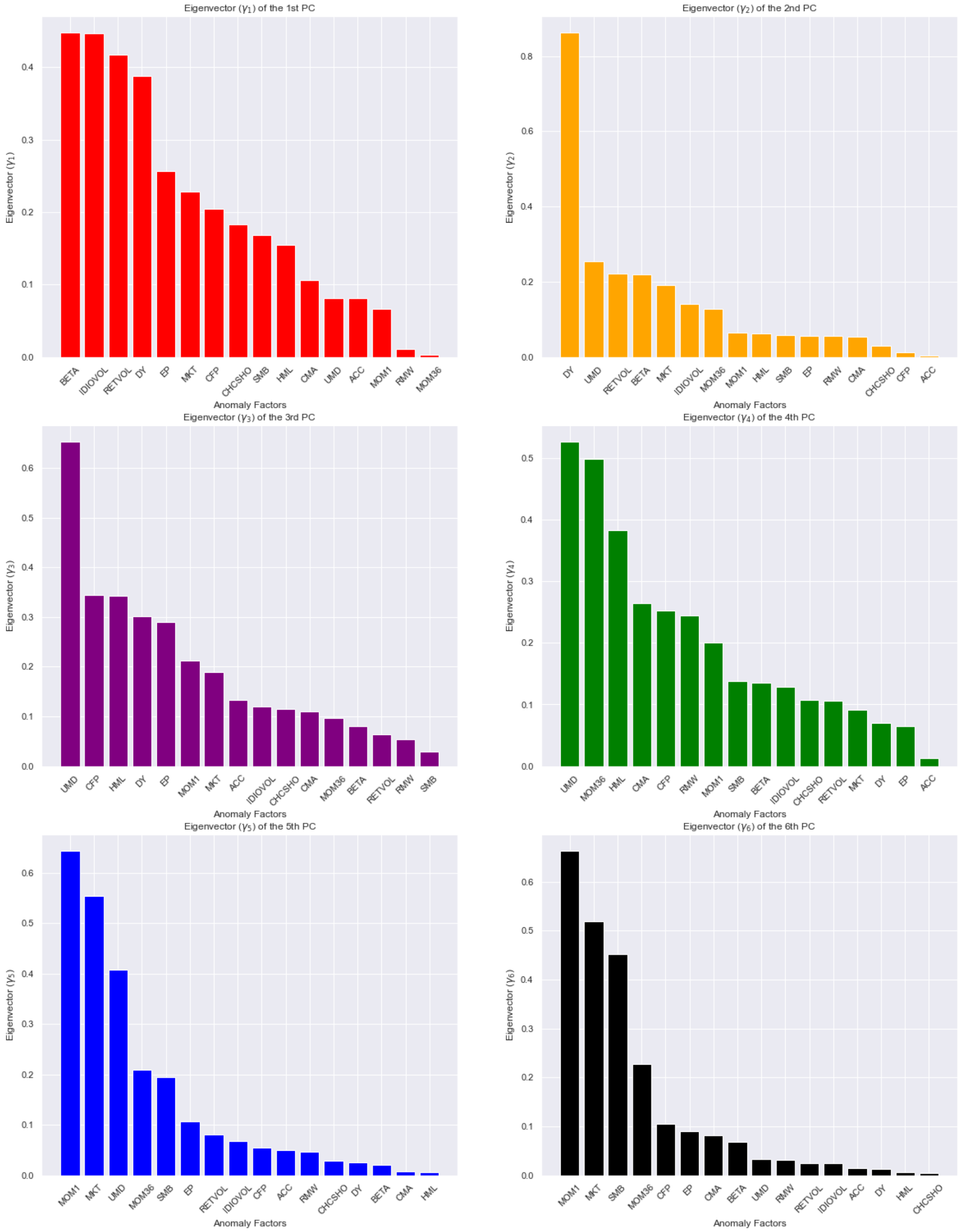
The eigenvectors ($\boldsymbol{\Gamma}_k$) give us how the factors ($\boldsymbol{X}$) and the k-th PC ($\boldsymbol{Z}_{(k)}$) are related since the PCs are obtained by the equation.

$$\begin{aligned} \boldsymbol{Z}_{(k)} &:= \boldsymbol{X}\boldsymbol{\Gamma}_k \\ &:= (Z_1, \ldots, Z_k) \end{aligned}$$

Therefore, in order to find a dominating "anomaly" we will look through the eigenvectors ($\boldsymbol{\Gamma}_k$) of the selected first six PCs by the rule of thumb.

In Figure 6, we display the coefficients ($\boldsymbol{\Gamma}_6$) of the factors on the 6 PCs, in absolute value.

Figure 6: First six eigenvectors ($\mathbf{\Gamma}_6$) of the Anomaly Factors



Through an analysis of Figure 6, it becomes evident that the first Principal Component (PC) is characterized by four dominant Anomaly Factors, each displaying γ values hovering around 0.4. These factors are identified as Market Beta (BETA), Daily idiosyncratic variance (IDIVOL), Daily variance (RETVOL), and Dividend/Price ratio (DY). In the subsequent second and third PCs, the highlight goes to Dividend/Price ratio (DY) and Up-minus-down (UMD), holding coefficients of 0.86 and 0.65 respectively, as the predominant Anomaly Factors.

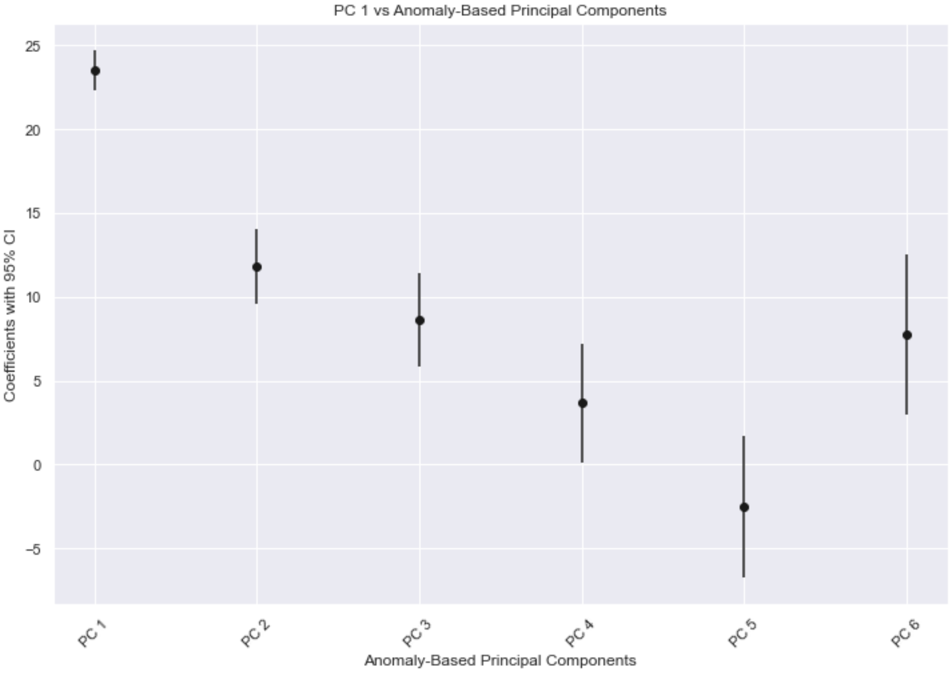
Moving forward, the remaining Principal Eigenvectors tell us three major Anomaly Factors for each Principal Component. The fourth PC is shaped by Up-minus-down (UMD), 36-month momentum (MOM36), and High-minus-Low (HML). Lastly, the fifth and sixth PCs, have its composition formed by 1-month momentum (MOM1) and Market (MKT), as first and second main Anomaly Factors, while the third Anomaly Factors are Up-minus-down (UMD) and Small-minus-Big (SMB) for the fifth and sixth

PCs respectively.

Item d

Since the Anomaly Factors are also constructed using the returns of the companies, we would expect a strong relationship between the first Principal Components (PCs) from both data sets, if there is indeed some underlying factor in the data.

Figure 7: Return-based PC1 and Anomaly-based PCs



This expectation finds validation in Figure 7, which shows the coeficients of a regression of the return-based PC on the 6 anomaly-based PCs. We can see that the first anomaly-based PC has a dominant coefficient in absolute value terms and its corresponding confidence interval. Even though the relation between the first anomaly-based PC and the first return-based PC is at least twice as pronounced as with other PCs, the second and third PCs also exhibit significance. Notably, the confident intervals consistently increase monotonically as the PC index advances.

Figure 8: Return-based PC2 and Anomaly-based PCs

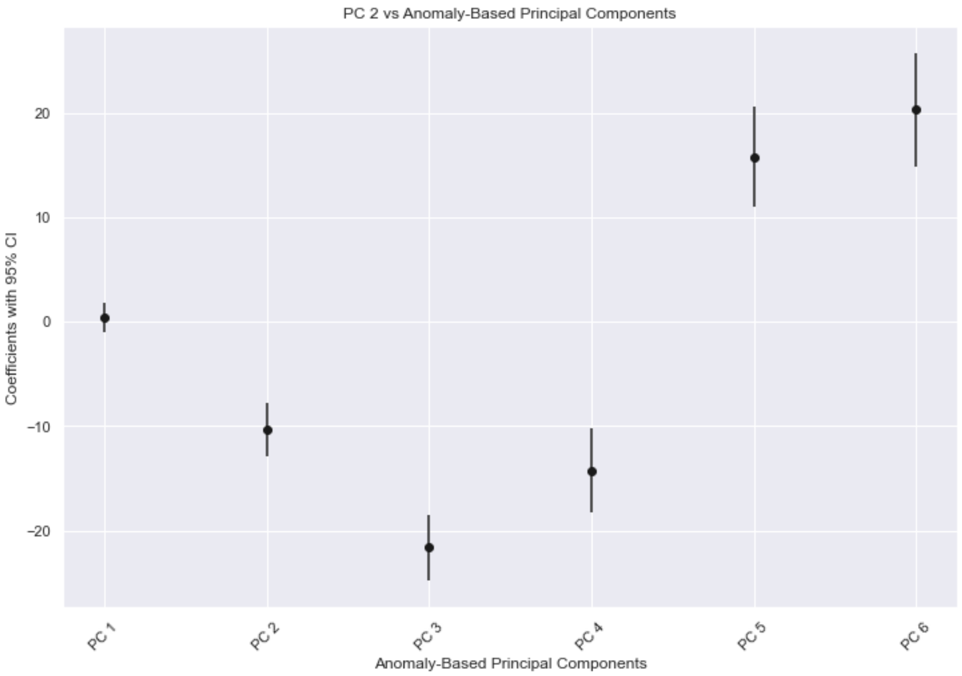


Figure 8 displays the coefficients of the second Return-based PC. As expected, the coefficient of the first Anomaly-based PC is zero due to the way the PCs are built. The math behind PC's methodology builds components to be uncorrelated between themselves. Given the strong correlation between the first Return-based PC and the first Anomaly-based PC, it is unsurprising that the coefficient of the second Anomaly-based PC is zero. However, the remaining Anomaly-based PCs seem to maintain a significant relationship with the second Return-based PC.

Therefore, there seems to be one important dimension of the data that is being captured by the first principal factors in both cases (anomaly-based and return-based principal components). While the relation between the remaining PCs is not as clear.

Question 2

In this question we compare four different models for inflation forecasting: (i) AR; (ii) PCR; (iii) Ridge and (iv) LASSO.

Our database is the FRED-MD, a macroeconomic dataset for the US Economy with monthly observations from January 1959 to November 2021 (vintage data as of December 2021). Variables are divided in eight groups: (i) output and income; (ii) labor market; (iii) housing; (iv) consumption, orders and inventories; (v) money and credit; (vi) interest and exchange rates; (vii) prices; and (viii) stock market.

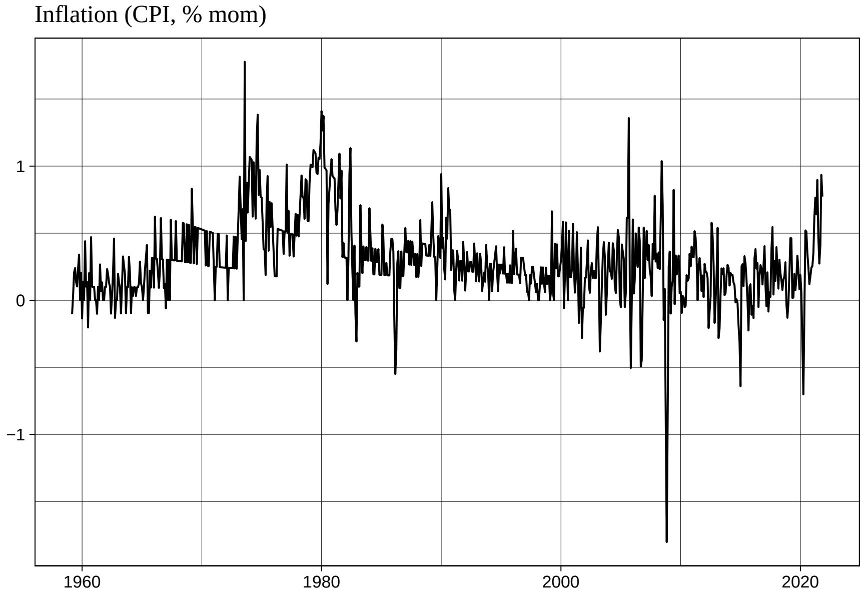
We begin by treating the data. Dropping series with missing variables leaves us with 104 series for the whole period. We adopt the transformations suggested by the Fed (McCracken and Ng, 2016) for all variables, as described in Table 1, except for the CPI. We use the inflation definition suggested in this exercise, i.e., $\pi_t = \frac{\Delta y_t}{y_t}$.

Table 1: FRED-MD - Transformations codes

Code	Transformation
1	None
2	Δx_t
3	$\Delta^2 x_t$
4	$\log(x_t)$
5	$\Delta \log(x_t)$
6	$\Delta^2 \log(x_t)$
7	$\Delta \left(\frac{x_t}{x_{t-1}} - 1 \right)$

The transformed CPI series which we want to forecast is depicted in Figure 9 below:

Figure 9: Inflation measured by CPI



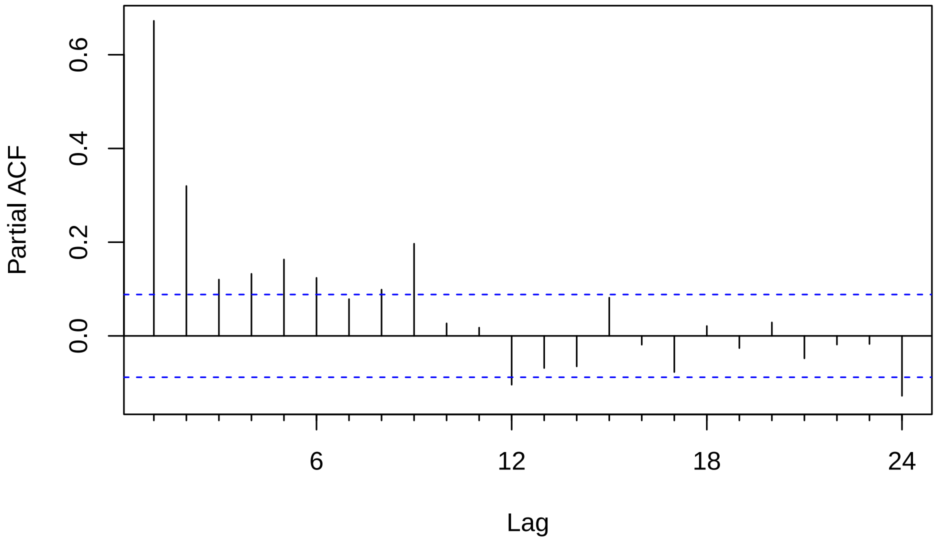
In all cases we will build one-step ahead forecasts for inflation, using a fixed rolling window of 492 observations. Our forecasts start in March 2000 and end in November 2021.

Autogressive (AR) Model:

$$\hat{\pi}^{(AR)}_{t+1|t} = \hat{\phi}_0 + \hat{\phi}_1 \pi_t + ... + \hat{\phi}_p \pi_{t-p+1} \tag{1}$$

Coefficients $\hat{\phi}_0, \hat{\phi}_1, ..., \hat{\phi}_p$ will be estimated by OLS. The order (p) of the AR will be determined by BIC criterion. We start by plotting the Partial Autocorrelation of the CPI for the first window (Figure 10) in order to get an idea of what should be the order p.

Figure 10: Partial Autocorrelation of CPI



After the 9th lag coefficients for the autocorrelation are mostly irrelevant. We nevertheless will let the BIC choose a maximum number of lags of 24 months, which is well above standard. The optimal number of lags will be chosen for each forecasting window using the BIC. The modal value for the optimal number of lags is 4. In Table 1A in the Appendix we show the count of optimal number of lags per window.

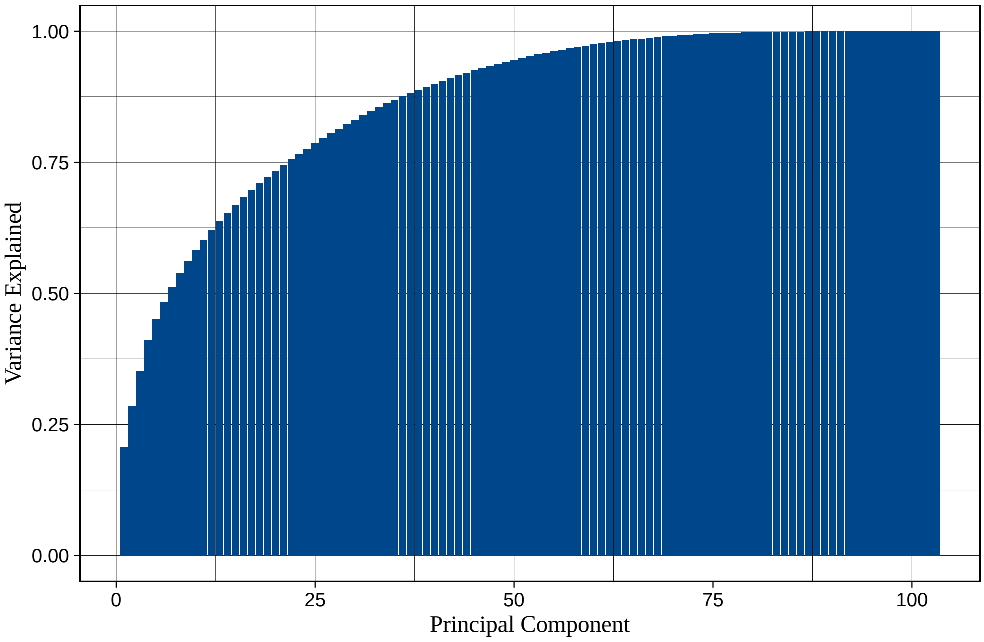
AR + Principal Component Regression (PCR) Model::

In this specification apart from an autogressive component, we add a vector of factors F_t which are given by principal component analysis:

$$\hat{\pi}^{(PCR)}_{t+1|t} = \hat{\phi}_0 + \hat{\phi}_1 \pi_t + ... + \hat{\phi}_p \pi_{t-p+1} + \hat{\lambda}' \hat{F}_t \tag{2}$$

Starting with the Principal Component Analysis (PCA) we center and scale the data and the choose the optimal number of factors k. Figure 11 below depicts the cumulative variance explained by adding consecutive principal components.

Figure 11: Cumulative Variance - Principal Components



As we discussed in Question 1, there are different ways to choose the number of factors. We look at the same 3 criteria and opt for the "Rule of Thumb" as it seems to be the most parsimonious in this case. We stop adding factor when it contributes to less than 3% of the variance. This results in an optimal $k = 6$ factors.

Given the optimal number of factor, we once again let the BIC decide on the optimal number of lags for the AR component of the model. We keep the maximum number of lags at 24. Table 3 shows the count of optimal lags chosen under the BIC for the estimation windows. The modal optimal lag is 15 lags.

Ridge:

We also use a Ridge model following the specification below:

$$\hat{\pi}^{(Ridge)}_{t+1|t} = \hat{\phi}_0 + \hat{\phi}_1 \pi_t + ... + \hat{\phi}_p \pi_{t-p+1} + \hat{\beta}'_1 X_t + ... + \hat{\beta}'_p X_{t-p+1} \tag{3}$$

We choose the penalty term according to the BIC[1]. It is important to note, however, that this criterion is obviously silent with respect to the optimal number of lags in X or π.

Initially, we intended to use an order of $p = 4$, but the predictive power of the Ridge regression model, notably displayed a poor fit (see Appendix). It has come to our attention that due to the Ridge's inability to yield a sparse solution, excessive inclusion of variables results in an estimated model that effectively becomes an intercept, with nearly all other coefficients converging towards, albeit not reaching, zero. In light of this observation, we proceeded to explore alternative, more parsimonious model specifications.

We experimented with different specifications both for the lags in independent variables and in CPI lags. We examined a configuration encompassing all macroeconomic variables, excluding any lags, in addition to incorporating lags of the CPI. This refined approach yields a more coherent outcome. Importantly, the results exhibit robustness across varying counts of CPI lags, reaffirming our initial intent to maintain four lags of the CPI. Henceforth, we call "Ridge" this improved specification and "Ridge (4 lags)" the specification with 4 lags of all variables.

LASSO:

Finally we use a LASSO model following the specification below:

$$\hat{\pi}^{(LASSO)}_{t+1|t} = \hat{\phi}_0 + \hat{\phi}_1 \pi_t + ... + \hat{\phi}_p \pi_{t-p+1} + \hat{\beta}'_1 X_t + ... + \hat{\beta}'_p X_{t-p+1} \tag{4}$$

Once again, we choose the penalty term based on the BIC. We also tested different specifications for the number of lags (p) and decided to settle for the best fit which in this case four lags ($p = 4$). Notably, in contrast to the Ridge model, the LASSO exhibits the remarkable capability to accommodate four lags for all variables, owing to its inherent ability to yield sparse solutions.

Item a

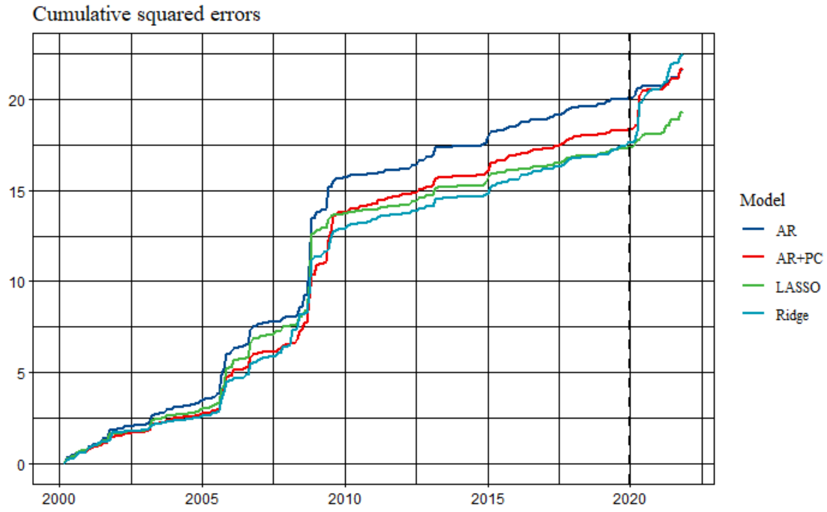
Using the 4 models described above we computed the squared one-step-ahead forecasting error for the next observation. Figure 12 depicts the cumulative value of these squared errors.

From the graph it is clear that LASSO (green line) is the winner model from the perspective of mean squared errors. The other 3 models (Ridge, AR, AR+PC) produced similar results in terms of predictive power, with Ridge being the worst overall.

Overtime, we can see a first big jump in forecast errors around the 2008 Great Recession. In general terms, COVID-19 (represented by the black dashed line) also resulted in bigger MSE for all models but affected the Ridge predictions in a particularly strong way. In fact, between 2010 and 2017, Rigde was actually the best model. Then, it was outperformed by the LASSO during the pandemic period. Moreover, the ARPC model also performed poorly after Covid-19. Up until 2020, this model presented significant improvement relative to the benchmark AR model. Nevertheless, when we consider the whole forecasting period the two models have a similar performance.

[1] Implementation of the BIC criterion for LASSO and Ridge was done with HDEconometrics Package by Gabriel Vasconcelos.

Figure 12: Cumulative MSE - AR, ARPC, Ridge, LASSO



Item b

We now proceed to analyze the importance of different variables to the LASSO, Ridge and ARPC models. We start by computing the importance of each variable in each window. For LASSO and Ridge the importance is given by the coefficient (ϕ for CPI lags ; β for macro variables) multiplied by the standard deviation of the variable.

In the case of ARPC, we first multiply the coefficients estimated for the factors (λ) in our equation model (2) by the (α) in $F_{it} = \alpha' X_t$. Because factors are linear combinations of the original macro variables, this will result in a measure of importance for the variables underlying the factor model.

We analyze both the importance of individual variables and the importance of the groups of variables. We follow the division into eight groups suggested by the Fed and add a ninth one with the lagged values of the CPI itself, ending up with the following groups: (i) output and income; (ii) labor market; (iii) housing; (iv) consumption, orders and inventories; (v) money and credit; (vi) interest and exchange rates; (vii) prices; (viii) stock market; (ix) lagged values of the CPI. In measuring group importance over time, we compute the importance of each group as the sum of the absolute values of the coefficients belonging to that group over the sum of the absolute values of all the coefficients.

We start with the Ridge model. As it will become clear it is notably different from the other models when it comes to variable importance. In Figure 13, we can see that the most relevant variables were a mixture from various groups including: RPI (Real Personal Income); M1SL (M1 Money Stock); lags of the CPI (CPIUCSL, CPIUCSL.1); prices (OILPRICEx); labor market variables (USWTRADE - All Employees Wholesale Trade; PAYEMS - All Employees Total nonfarm; SRVPRD - All Employees Service-Providing Industries; CES0600000007 - Avg Weekly Hours Goods-Producing). For comparative reasons we set the importance of the most important variable at 100. When looking at the importance of groups over time (Figure 14) one can see the prevalence of Labor Market variables (Group 2).

Figure 13: Importance of Variables RIGDE - Top 10

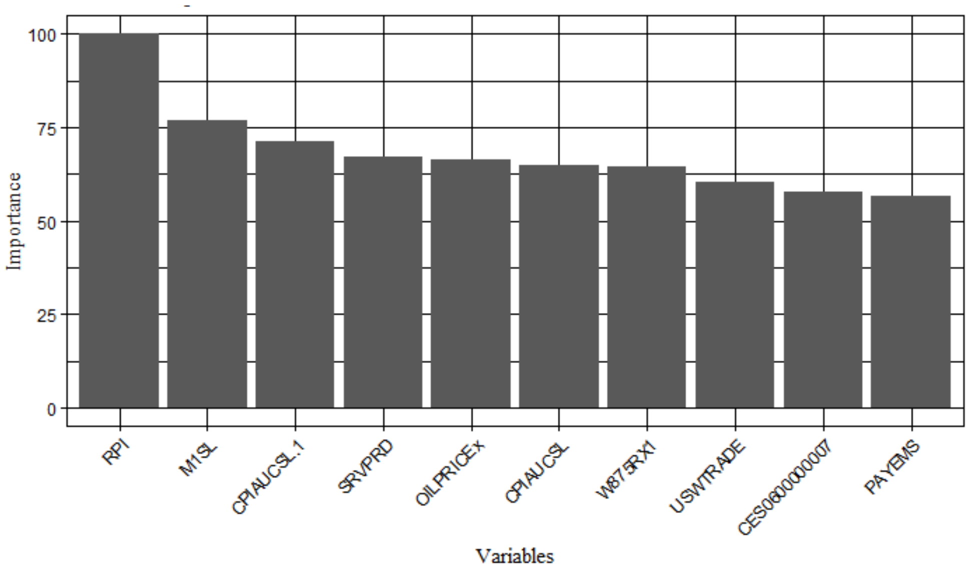
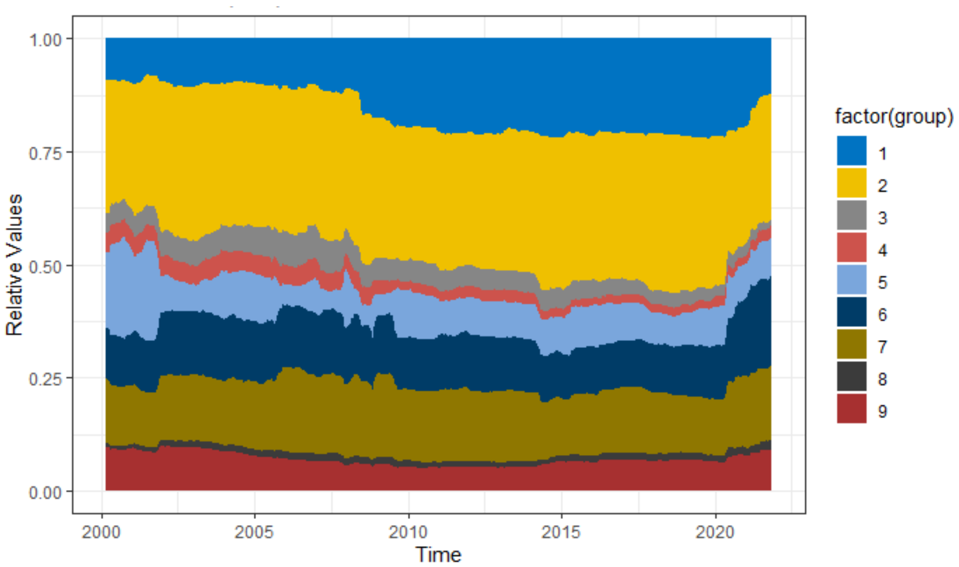


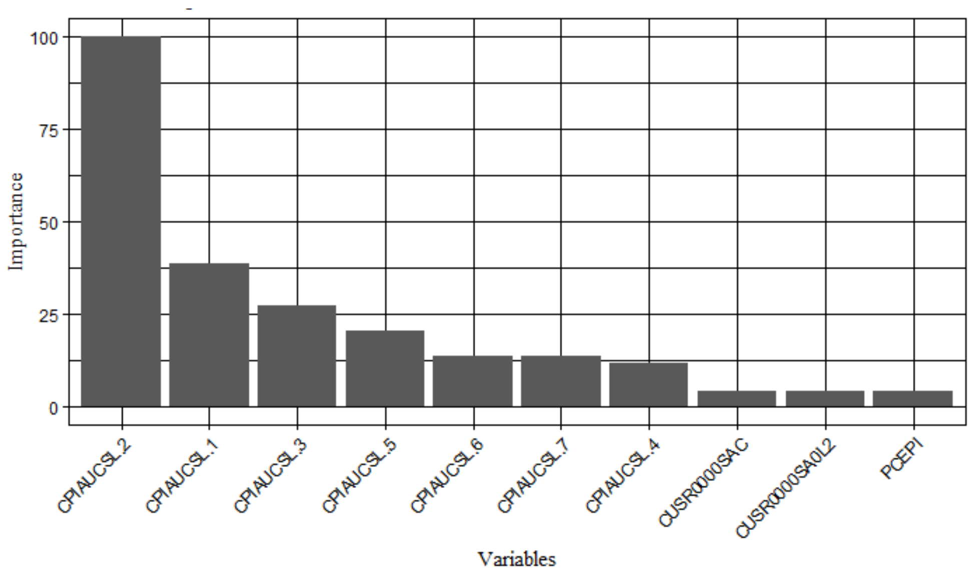
Figure 14: Importance of Variables RIGDE - Groups over time



Now, for the other two models a more similar pattern will be seen. Both for the ARPC and LASSO models, most relevant group of variables will be the CPI lags themselves.

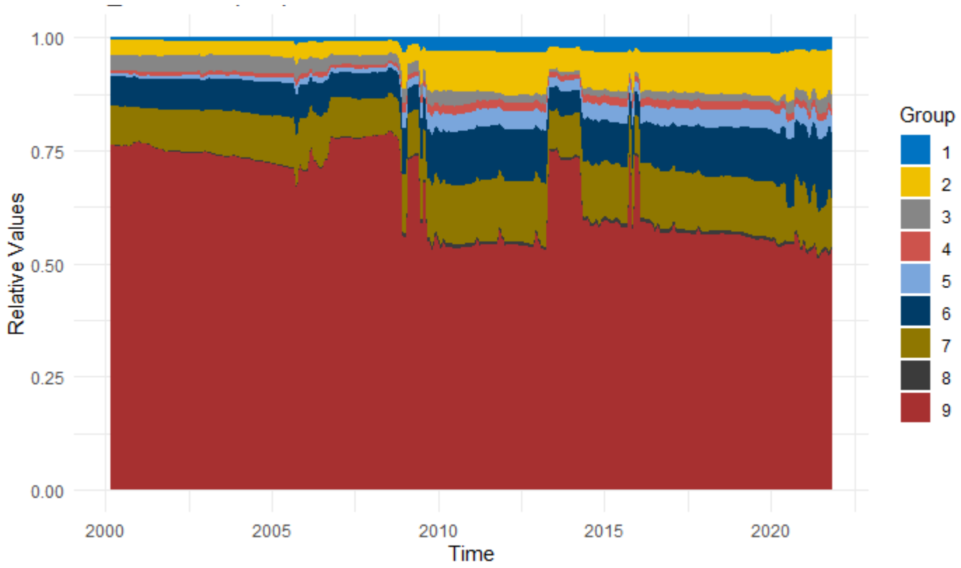
Figure 15 depicts the 10 most important variables over all windows for the ARPC. We can see how the first seven most important variables represent lagged values of the CPI itself. The other three variables are also measures related to prices: CUSR0000SAC is the commodities component of the CPI; CUSR0000SA0L2 is the CPI excluding shelter items and PCEPI is the Personal Consumption Expenditures Price Index another inflation index produced by the Bureau of Economic Analysis (BEA).

Figure 15: Importance of Variables ARPC - Top 10



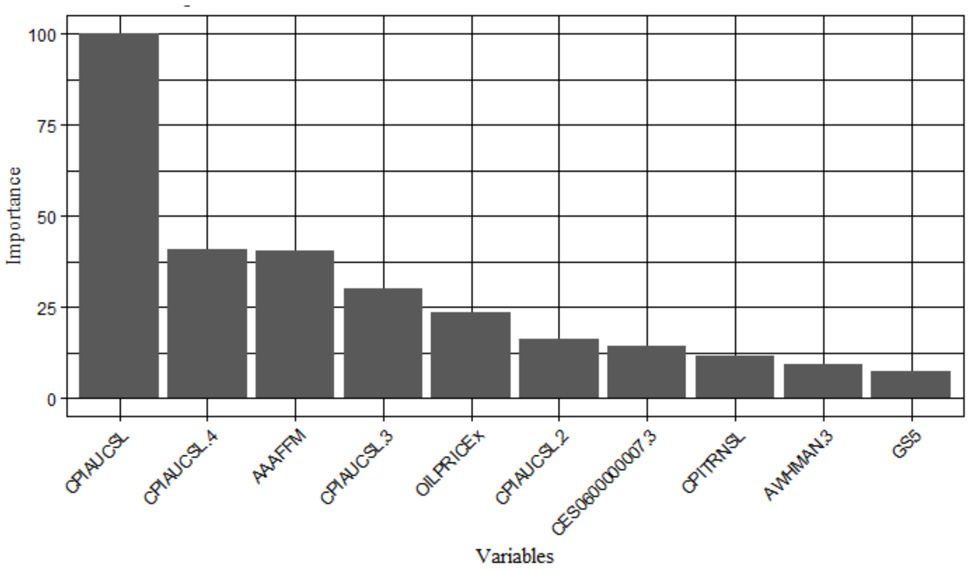
In Figure 16 we plot the evolution of Group importance overtime for the ARPC. One can see how Group 9 (CPI lags) is the dominating group over all estimation windows followed by Group 7 (Prices). Together they account for more 60% of overall importance throughout the whole period. In the last years, Groups 6 (Interest and Exchange Rates) and Group 2 (Labor Market) have become more important and represent a little more than 20% of overall importance in the last windows.

Figure 16: Importance of Variables ARPC - Groups over time



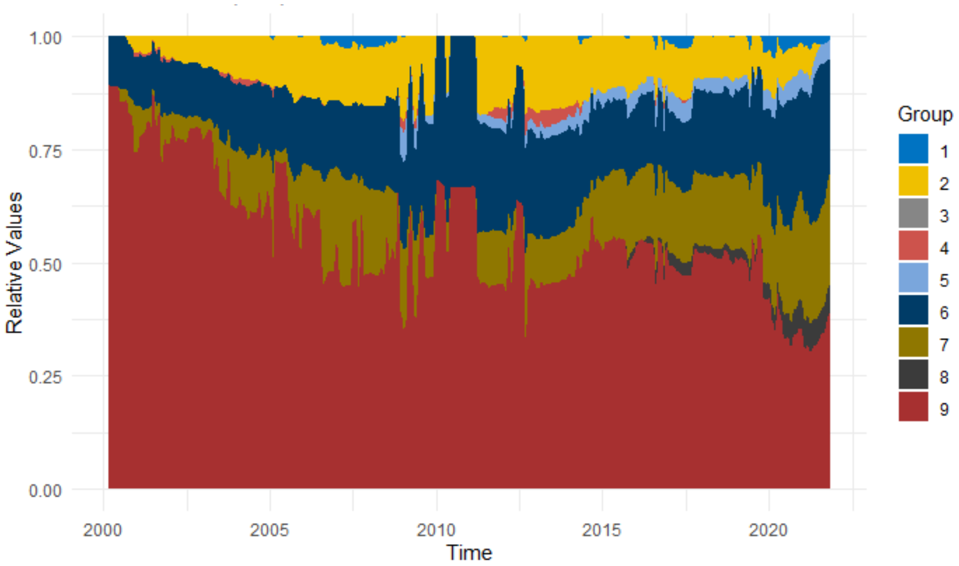
Lastly, for the LASSO model, Figure 17 portrays the 10 most important variables over the estimation windows. Once again measures of previous CPI and other variables related to pricing are the most relevant ones. For the LASSO, however we see other kinds of variables as well. This is the case of AAAFFM (Moody's Aaa Corporate Bond Minus FEDFUNDS), AWHMAN (Avg Weekly Hours : Manufacturing) and GS5 5-Year Treasury Rate.

Figure 17: Importance of Variables LASSO - Top 10



When looking at group importance (Figure 18), we get back to a similar conclusion. Group 9 (CPI lags) has been the most important group overtime, although with a perceptible downward trend. Group 7 (Prices) and Group 6 (Interest and Exchange Rates) come in second and third in importance. These three together account for more than 90% of the overall importance in the last windows.

Figure 18: Importance of Variables LASSO - Groups over time



Question 3

Item a

Given that the Convolution model primarily captures seasonal components and is extensively employed in image recognition, we have opted to explore two Neural Network methods: the Deep Neural Network and Long Short-Term Memory (LSTM).

Our choice of the Deep Neural Network is influenced by recent literature, specifically the paper Medeiros, Schütte and Soussi (2022). This paper showcases improvements in inflation forecasting across multiple countries through the utilization of non-linear models like Random Forest and feedforward Neural Networks. Our Neural Network architecture follows the feedforward scheme outlined in the paper, where the number of neurons goes down as the model progresses through layers. We have adopted a three-layer design, like the approach presented in the aforementioned paper. The number of neurons for each layer is determined using guidelines from Masters (1993), resulting in 100 neurons for the first layer, 60 for the second, and 30 for the third. We employ the adaptive moment (Adam) optimizer Kingma and Ba (2014) and utilize the standard relu activation function ($f(x) = \max(x, 0)$).

The LSTM model also employs the same Adam optimizer and the default for activation and recurrent activation functions, which are hyperbolic tangent $f(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$ and sigmoid $f(x) = \frac{1}{1+\exp(-x)}$. The architecture is the same as in the Deep Neural Network model, just differs with the LSTM acting like neuron.

Item b

In Figure 19 we plot the forecasts for the two specifications of Neural Networks along with the four linear models we estimated before. As hinted by the first graph, the Deep Neural Network model presents a poorer fit than any other model. This result coincides with the ones found in Medeiros, Vasconcelos, Veiga and Zilberman (2021). The model seems to be particularly bad in predicting the biggest spikes and dips in inflation and lost considerable predictive power after COVID-19. The LSTM model on the other hand seems to produce better forecasts. When comparing the MSE we will be able to confirm that.

Figure 19: Inflation forecasts

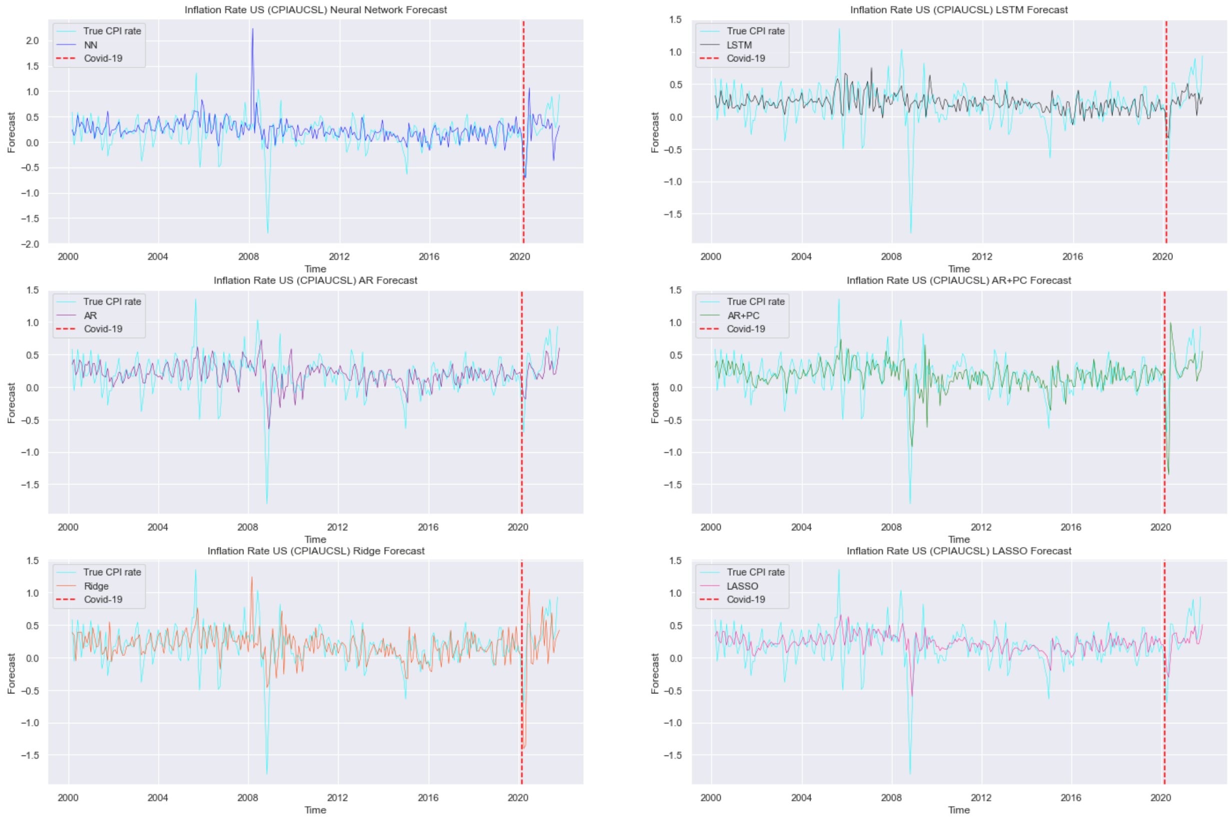


Table 2: MSE - NN and linear models

Model	MSE
NN	1.2403
Ridge	1.0615
AR+PC	1.0238
AR	1.0000
LSTM	0.9733
LASSO	0.9130

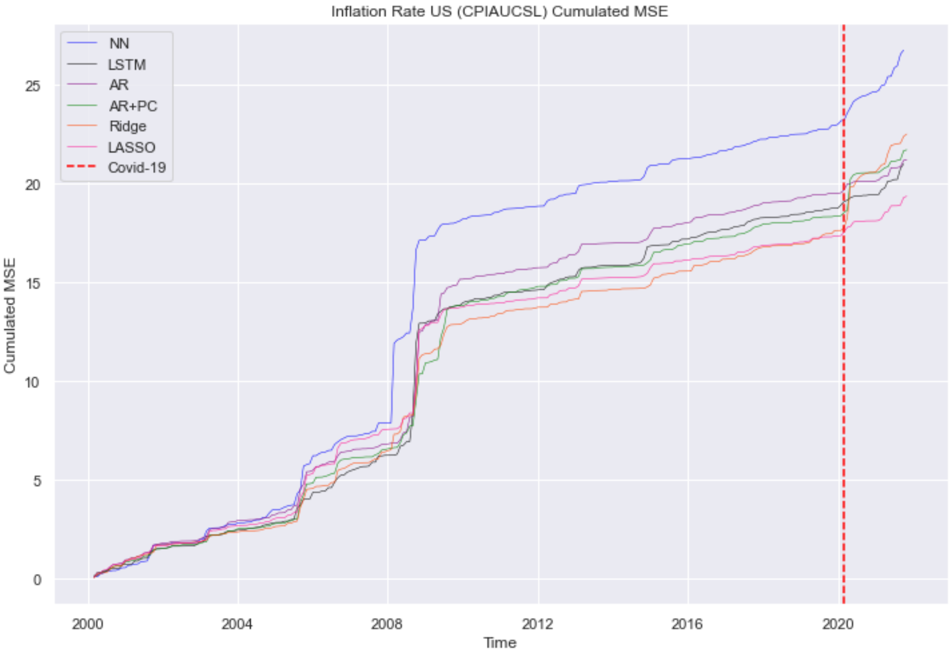
Note: The MSE relative to the benchmark AR (=1)

Item c

We plot cumulative MSE for all models in Figure 20. It is clear that the Deep Neural Network is considerably worse than any other model, throughout the whole period. Even though the LSTM specification is not able to outperform the LASSO, it was the second best performing model when considering all windows. Notably, for the period before the Global Financial crisis (GFC), the LSTM did better than the LASSO.

Table 2 compares the MSE of the linear models, taking the AR model as benchmark, with the MSE of the Neural Network models. The LSTM specification represents an improvement of close to 2.5% when compared to the benchmark, but as we mentioned fails to beat LASSO. The Deep NN is considerably worse than the linear models.

Figure 20: Cumulative MSE



Item d

For the model based on regression trees, we choose the Random Forest, because the literature has shown that it can perform well (improving forecasts relative to linear benchmarks and other ML methods) in this type of application (Medeiros et al., 2021).

The Random Forest, proposed by Breiman (2001), relies on bagging (bootstrap aggregating) in order to reduce the variance of regression trees, which are unstable estimators. It is important to notice that Breiman's method is not designed for time series and uses I.I.D. bootstrap. However, since we are interested in direct forecasts, we can simply embed the lags of the variables we want to use in our model, that is, include them in our matrix of covariates, and then treat the observations independently. This way, we can apply the usual RF method.[2]

In choosing the parameters for the RF, we follow Medeiros et al. (2021) - we tried alternative values for each parameter and the results are fairly robust. The proportion of variables selected randomly in each split is 1/3; the number of trees is 500; and each tree is grown until there are 5 observations in each leaf (terminal node).

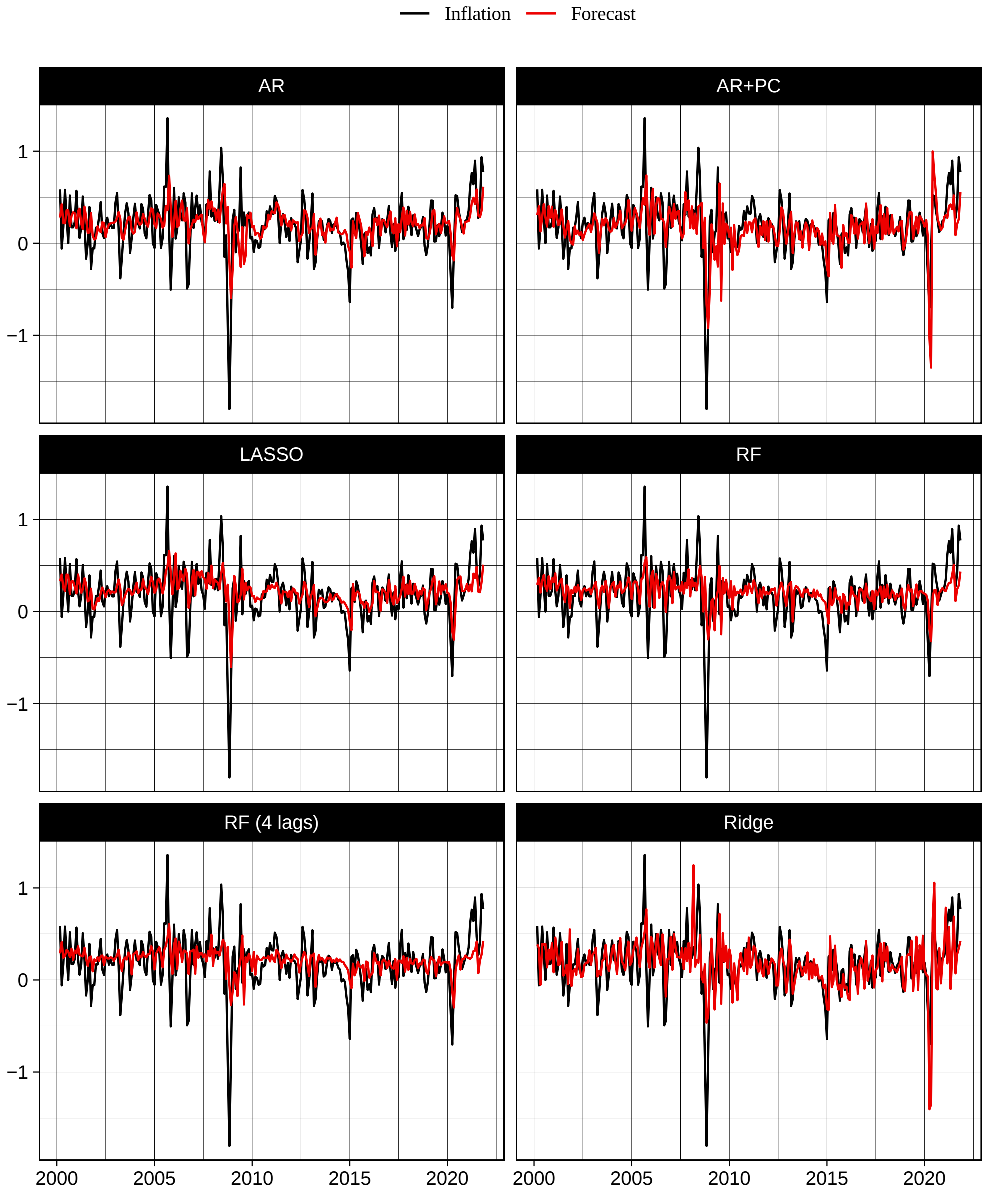
Similarly to the Ridge regression, we started using 4 lags of all variables, but later noticed that the model did not handle well that many variables. Hence, we also present the results for the estimation using 4 lags of the CPI and no lags of the other variables (only current values).

Item e

In Figure 21, we plot the forecasts of the four linear models from Question 2 along with the forecasts produced by the Random Forest models. As the analysis of MSE will confirm, the Random Forest models perform well and better than most of the linear models.

[2] This embedding process is done inside which forecast window, to take into account the loss of observations when including lags, and have a fair comparison to the other models

Figure 21: Inflation forecasts



Item f

The Random Forest outperforms all the linear models from Question 2. The best specification of the RF presents an improvement in cumulative MSE of almost 10% relative to the AR model (Table 3). Even the specification with 4 lags of all variables still shows improvement over the AR model. This is consistent with the results in the literature.

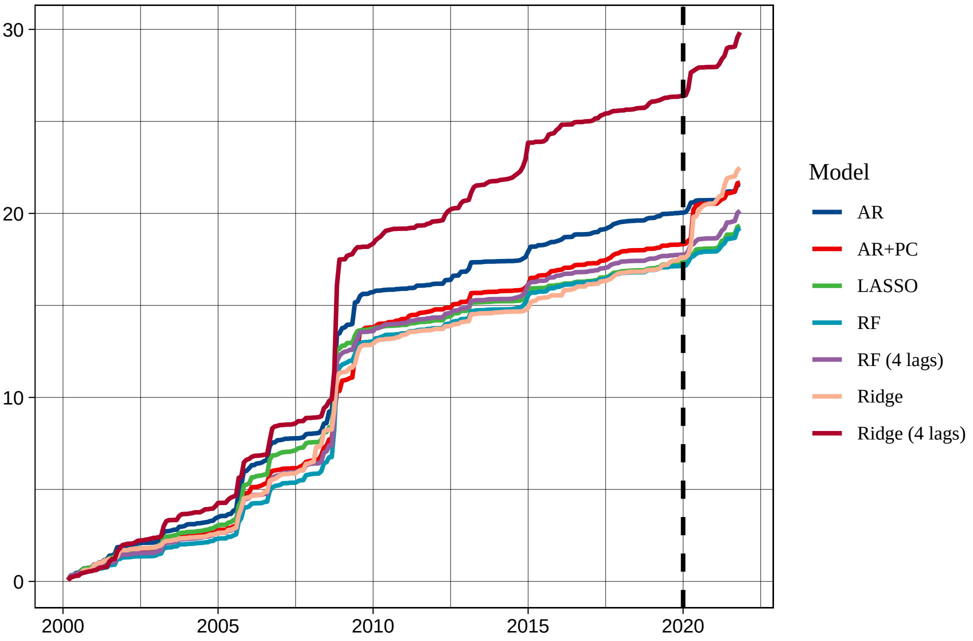
Having said that, it is worth noting that the LASSO, which is the best linear model, attains an improvement of 8.70% relative to the AR model, outperforming the RF with 4 lags.

Table 3: MSE - RF and Linear Models

Model	MSE
AR	1.0000
AR+PC	1.0238
Ridge (4 lags)	1.4092
Ridge	1.0615
LASSO	0.9130
RF (4 lags)	0.9511
RF	0.9070

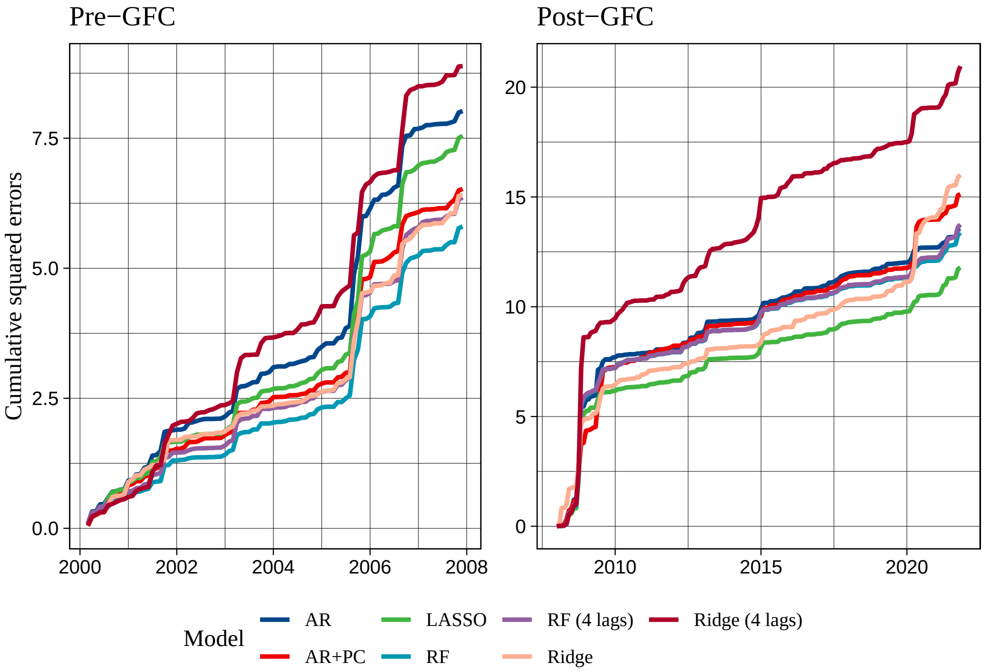
Note: The MSE relative to the benchmark AR (=1)

Figure 22: Cumulative MSE



Upon a closer examination of the performance exhibited by the LASSO and the RF, it becomes evident that a singular superior model does not universally prevail. Although the RF outperforms the LASSO across the entirety of the forecasting period, there are sub-samples in which the LASSO has a smaller cumulative error. In particular, one relevant split of the sample is the GFC of 2008. As illustrated in Figure 23, before the GFC, the best model is the RF, followed by the Ridge and the AR+PC. Yet, during the post-GFC period, the LASSO takes the lead in terms of performance, followed by the RF. Moreover, we can see that the RF's advantage over the benchmark AR model stems mainly from the pre-GFC period.

Figure 23: Cumulative MSE over different sub-samples



Item g

Figure 24 combines all models we estimated in Questions 2 and 3. We had already shown that among linear models, LASSO yielded the best results in predictive power and that Neural Networks models were not capable of beating LASSO. The two Random Forest model specifications on the other hand beat LASSO and all the other models including Neural Network specifications.

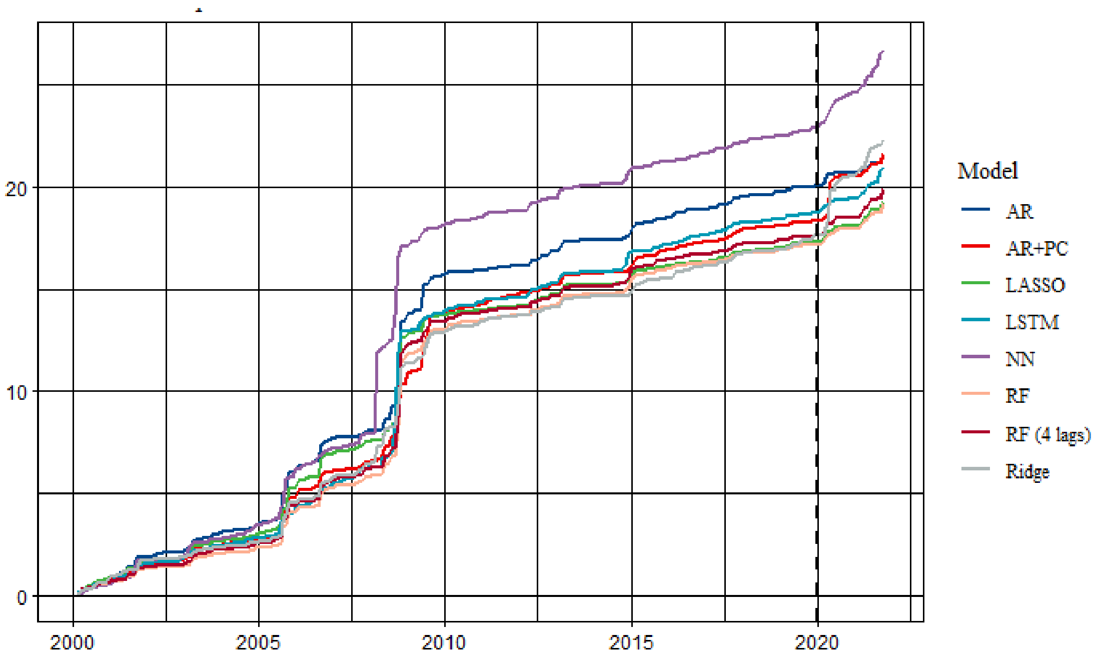
In Table 4 we compare for the last time the MSE of all models. MSE compared to the Benchmark (AR model) are shown in descending order. Compared to the Benchmark, both LASSO and RF were able to generate considerable gains, with a MSE around 10 % smaller. The Neural Network with LSTM specification also delivered some improvement as well as the RF model with 4 lags. All other models produced no predictive gains when compared to the AR, with the deep NN and Ridge with 4 lags being considerably worse.

Table 4: MSE

Model	MSE
Ridge (4 lags)	1.4092
NN	1.2403
Ridge	1.0615
AR+PC	1.0238
AR	1.0000
LSTM	0.9733
RF (4 lags)	0.9511
LASSO	0.9130
RF	0.9070

Note: The MSE relative to the benchmark AR (=1)

Figure 24: Cumulative MSE - All models



Appendix A

In this appendix we provide additional information on the models estimated for Question 2. Tables A1 and A2 below depict the count of optimal lags chosen for the estimation windows following the Bayesian Information Criterion (BIC). For the AR model, the most common optimal number of lags is 4. For the AR+PC model, the most common optimal number of lags is 15.

Table A1: AR - Optimal Lags under BIC

Optimal Lag	Number of Windows
4	75
9	64
12	60
3	22
15	11
5	7
6	6
13	3
16	2
2	1
17	1
18	1
19	1
20	1

Table A2: AR + PC - Optimal Lags under BIC

Optimal no. Lags	Number of Windows
15	69
9	65
18	43
8	37
10	23
7	9
16	6
11	2
17	2
21	2
19	1
22	1
23	1

The four figures below depict the CPI and the predicted values for each one of the estimated models: AR; AR+PC; LASSO and Ridge. The mean squared errors for these models are compared in Question 2 above.

Figure A1: AR Forecast

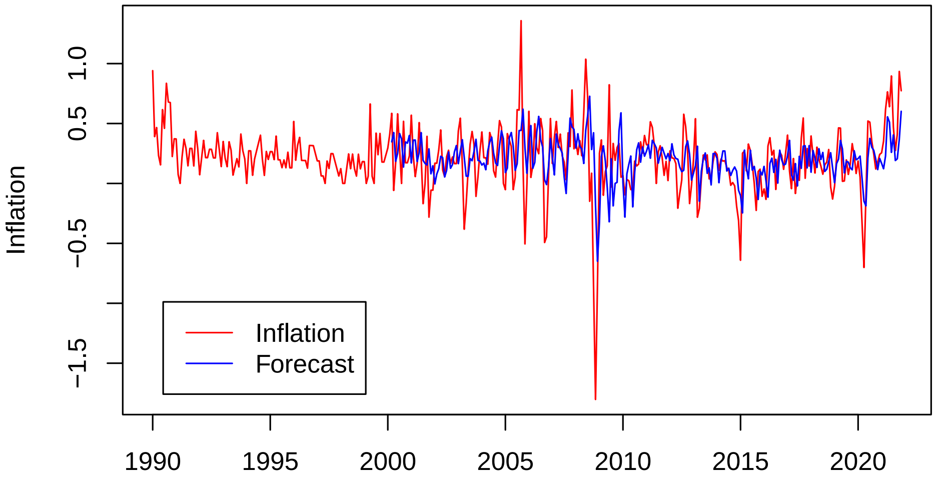


Figure A2: AR + PC Forecast

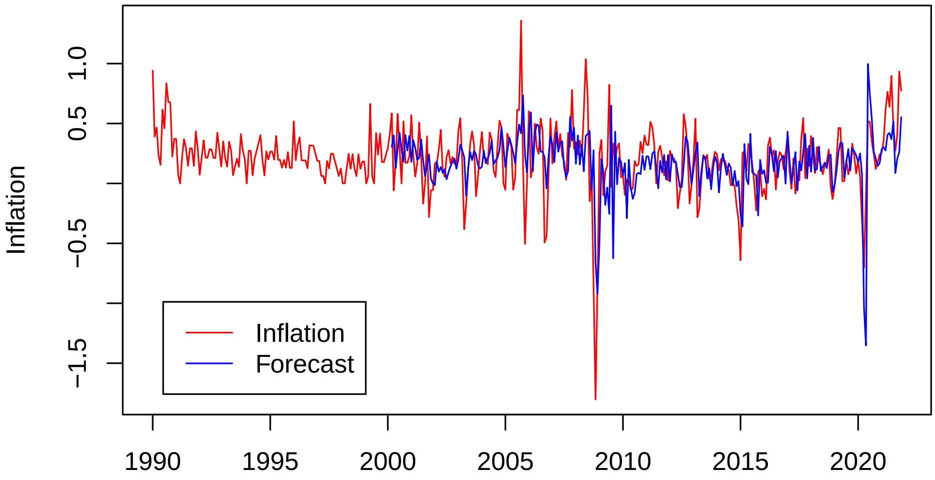
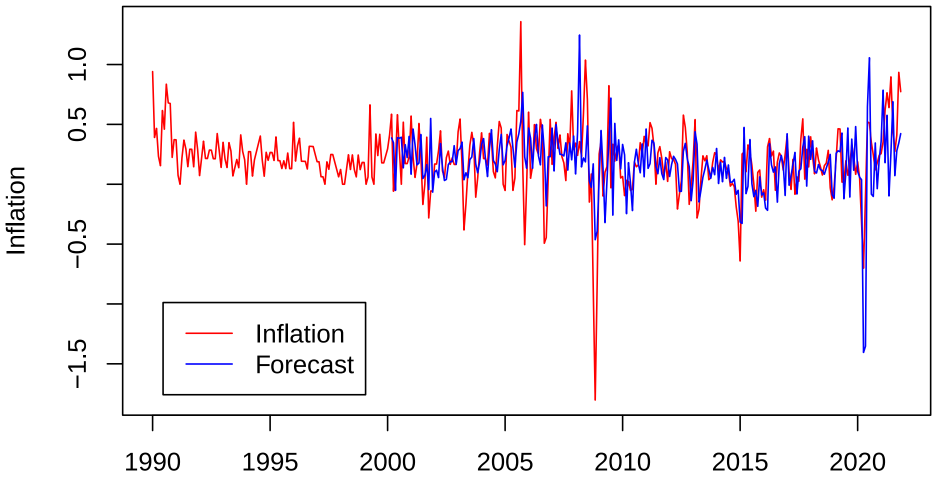
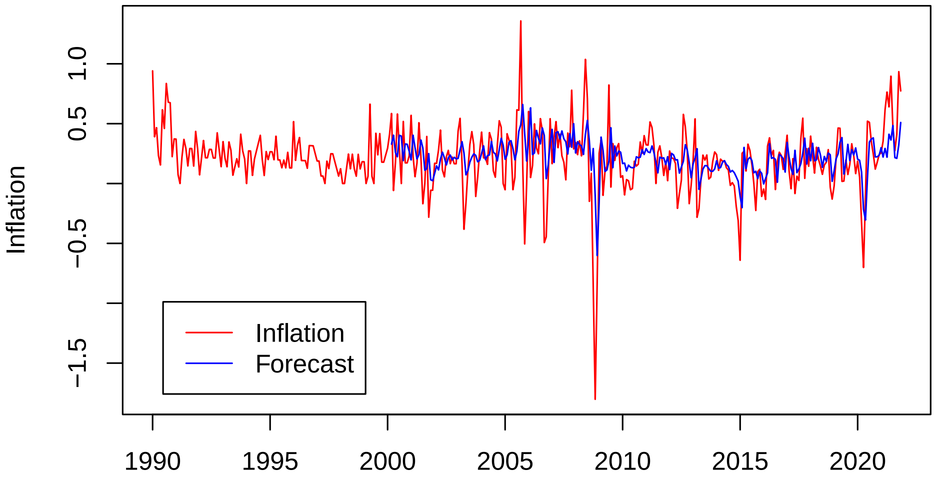


Figure A3: Ridge Forecast



Using 4 lags of CPI and no lags of other variables

Figure A4: LASSO Forecast



Appendix B

In this appendix we show the alternative specifications for the Ridge model we estimated on Question 2, i.e., a model with 4 lags for all macro variables and the CPI. It is easy to see how this models fares much worse than the more parsimonious Ridge with no lags for the macro variables.

Figure B1: Ridge Forecast - Alternative Specification

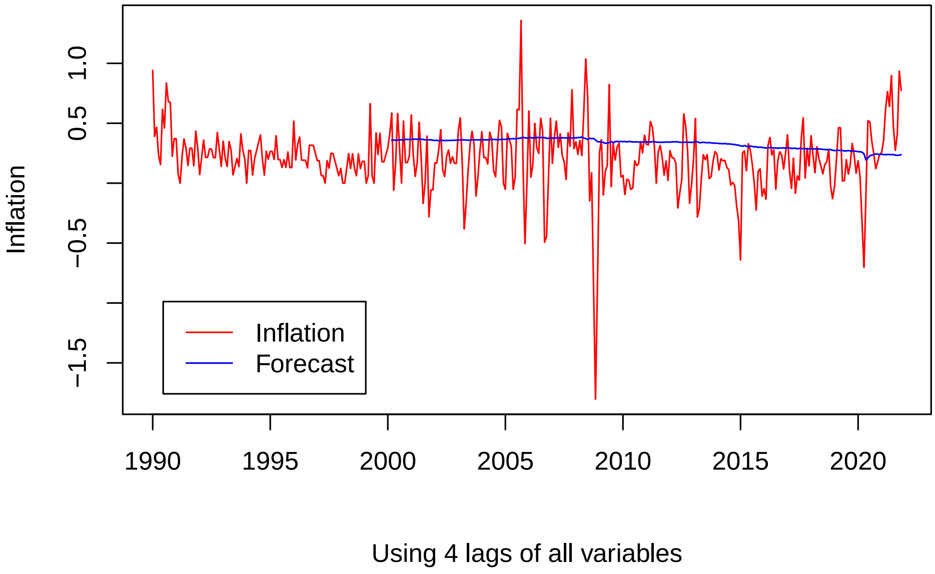
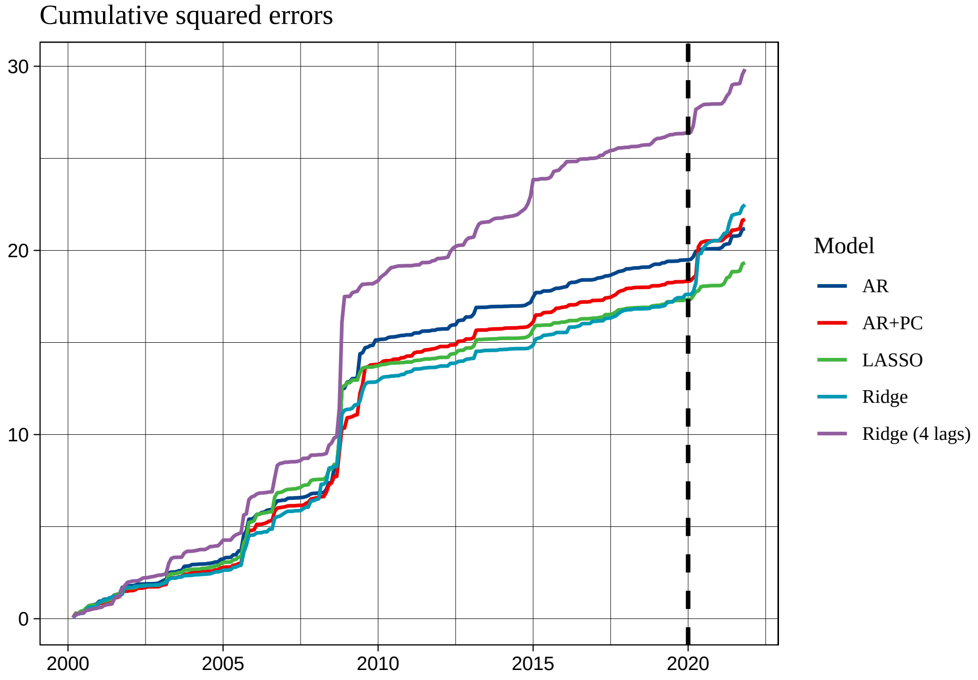


Figure B2: LASSO Forecast



References

Bai, Jushan and Serena Ng, "Determining the number of factors in approximate factor models," *Econometrica*, 2002, *70* (1), 191–221.

Breiman, Leo, "Random forests," *Machine learning*, 2001, *45*, 5–32.

Kingma, Diederik P and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Masters, Timothy, *Practical neural network recipes in C++*, Morgan Kaufmann, 1993.

McCracken, Michael W and Serena Ng, "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 2016, *34* (4), 574–589.

Medeiros, Marcelo C, Erik Christian Montes Schütte, and Tobias Skipper Soussi, "Global inflation forecasting: Benefits from machine learning methods," *Available at SSRN 4145665*, 2022.

Medeiros, Marcelo C., Gabriel F. R. Vasconcelos, Álvaro Veiga, and Eduardo Zilberman, "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods," *Journal of Business & Economic Statistics*, 2021, *39* (1), 98–119.

Onatski, Alexei, "Determining the number of factors from empirical distribution of eigenvalues," *The Review of Economics and Statistics*, 2010, *92* (4), 1004–1016.