ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Quantitative Finance

Partial Least Squares and Penalized Regression in Time Series for Macroeconomic Forecasting

Jeroen van den Boogaard

435322

July 2019

Supervisor: A.M. Schnücker   Second assessor: D.J.C. van Dijk

---

Abstract

Many studies have applied factor models to reduce the dimension of the subspace spanned by the predictors through factor analysis. This paper revisits partial least squares to investigate the forecasting performance when this reduction is related to the forecast goal. The most well-known method to estimate the common factors, called principal components, is used for comparison. This study revisits three different approaches of partial least squares to investigate whether forecasting accuracy can be improved over this widely used factor forecasting method. In addition, a regularization and variable selection method, called the elastic net, is applied to the same data from the Stock and Watson database, as another method to forecast while the dimension among the predictors is reduced. One static and one dynamic partial least squares approach show good improvements over the principal components method. The elastic net method has relatively good forecast accuracy, but fails to improve the forecast performance of principal components in most cases.

Contents

1 Introduction

Multiple studies (for example, Stock and Watson, 2002a,b; De Mol et al., 2008; Matheson, 2006 and others) have shown that the performance in forecasting can be improved by incorporating a reduced set of the available predictors instead of forecasting with all the variables that are available. This is due to the fact that reducing the data set to the most informative predictors, removes the 'noise' when predicting the target variable. Incorporating a spacious amount of variables in one forecasting model while estimating with standard econometric methods appears to be inefficient or just impossible, because the added 'noisy' series cause the residual cross-correlation to exceed the amount allowed by the theory (Boivin and Ng, 2006). This can be due to the arising of high dimensionality and multicollinearity. Different methods and techniques have been found to take into account these flaws (for example, Stock and Watson, 2002a,b; De Mol et al., 2008; Matheson, 2006). Important examples of this are the factor models, which extract a small number of factors from a large amount of predictors by means of factor analysis and use these estimated factors to regress them over the target variable. The subspace that is spanned by the predictors is decreased by applying this method. For example, Chan et al. (1998) and Stock and Watson (2002b) extract three factors out of 150 series and with this improve the forecast accuracy of various macroeconomic variables. The most well-known and used method to obtain consistent estimates of the common factors are principal components (PC). This can be used in both static and dynamic factor models.

However, the current knowledge of improving the forecast performance by reducing the subspace spanned by the predictors is insufficient as it is not related to the forecast goal when this is done by means of factor methods. This is because when the common factors are estimated, the methodology does not consider the predictive ability of the predictors for the target variable.

Another method which relates the reduction of the subspace spanned by the predictors to the forecast goal is introduced by Wold (1996): partial least squares (PLS). This method deals with the problems of multicollinearity and high dimensionality in static applications. However, knowledge of this method in a dynamic context is still scarce. Fuentes et al. (2015), Groen and Kapetanios (2008) and Kelly and Pruitt (2012) revisit useful approaches in this context. Kelly and Pruitt (2012), for example, successfully obtain asymptotic efficiency when estimating common factors.

This paper revisits PLS, following Fuentes et al. (2015), and investigates whether the forecast performance can be improved in both static and dynamic applications. Compared to PLS, PC estimates common factors which "do not depend directly on the prediction purpose" (Fuentes et al., 2015). The reason for this is that when the factors are estimated, the methodology does not take into account the predictive power of the predictors over the target variable, such that there is no relation between the reduction of the dimension of the subspace spanned by the predictors and the goal of forecasting. PLS is a method that relates this reduction with the forecast goal. However, as PC is the most widely used method to estimate the common factors, I compare the forecast performances of this method with those of PLS. Many studies (Bai and Ng, 2002; Boivin and Ng, 2006 and others) show that not only the number of predictors, but also the quality of the predictors is of great importance to the estimation and prediction of the target variable. With the quality of the predictors, I refer to the informative content that they contain about the target variable. In this paper, I focus on choosing the predictors which contain the most useful information about the target variable to improve the forecast accuracy, instead of using all the predictors that are available. This means that for predictors with too little informative content, the factor loadings can have a value of zero.

Choosing a subset of variables can also be done by penalizing least squares methods which build on OLS. The lasso (Tibshirani, 1996) and the ridge (Hoerl and Kennard, 1988) regression are two of those penalizing techniques. The lasso performs an L_1-penalty on the regression coefficients whereas the ridge performs an L_2-penalty. A third method considered in this paper to reduce the influence of predictors that contain too little information about the target variable is the elastic net (EN). It is a regularization technique that performs automatic variable selection and simultaneously shrinks the estimates by my means of combining the penalties of the previously named penalizing techniques. The method is introduced by Zou and Hastie (2005) as a special case of the least angle regression (LARS) algortihm, introduced by Efron et al. (2004). The motivation for using the EN is because of the special properties it has: (1) the simultaneous variable selection and shrinkage of the estimates creates a parsimonious model; (2) it performs group selection for highly correlated predictors, a property that neither the lasso nor the ridge contains.

The forecasting performance of the EN method and three different PLS methods are compared to those of the PC to investigate whether forecasting performance can be improved upon the most widely used factor method. The focus of this comparison is on the US inflation over the timeframe from 1960 to 2003. To do so, I obtain data from the Stock and Watson database.

Extracting factors by means of PLS results in the best forecast accuracy when the lags of the target variable are directly included in the forecasting equation to capture the dynamic behaviour of the target. One static and one dynamic approach apply PLS with this property and show good improvements in prediction over PC.

Using the EN to select a subset of the predictors to forecast the target variable shows a strong forecast performance for the 24-month ahead prediction, but fails to improve forecast accuracy compared to the widely used PC method in most cases.

The paper proceeds as follows. Section 2 presents a review of the existing literature of this problem. Section 3 describes the data that is being used for the empirical comparison. The methods and techniques are explained and interpreted in section 4. Section 5 contains the results of the empirical comparisons. At last, conclusions are drawn in section 6.

2 Literature

A seminal study on PC is Stock and Watson (2002a), who use the first k principal components as factors in a linear regression to forecast the target variable. This is done by modelling the covariability of many predictors into a smaller number of factors. These forecasts are asymptotically efficient and the estimated factors are consistent. Boivin and Ng (2005) show a factor model which contains just a small number of auxiliary parameters that need specification, such that the model is robust to misspecification. However, it is based on static PC, such that it does not hold for dynamic applications. Bai and Ng (2008) investigate refinements to the factor models: (1) they allow a non-linear relation between the predictors and the target variable; (2) they relate the estimation with the forecast goal.

PLS tries to improve the forecast accuracy while reducing the dimension of the subspace spanned by the predictors, where there is a relationship between the forecast goal and this reduction. PLS is introduced by Wold (1966) for static applications. Wold (1966) shows that this method is valid even in the situations where the amount of predictors exceeds the sample size and for multicollinearity. Groen and Kapetanios (2008), (2015) are one of the first to investigate PLS in a dynamic context. They conclude that PLS regression is comparable to PC regression when the target variable and the predictors relate via a factor structure. However, PLS regression surpasses PC regression in case of weaker factor structure in the data. Also, Eickmeier and Ng (2011) acknowledge that the strength of the factor structure in the data is of great importance to the forecast accuracy and the comparison of different methods such as PC and PLS. With this, they mention that their New Zealand data set has a stronger factor structure than the international data sets. Following Fuentes et al. (2015), who efficiently improve forecasting while reducing the dimension of the subspace spanned by the predictors using different approaches of PLS, I revisit PLS in both static and dynamic applications, where I use the same United States (US) macroeconomic time series.

The EN is a regularization and variable selection method introduced by Zou and Hastie (2005), which is based on the LARS algorithm, proposed by Efron et al. (2004). The use of the EN method next to PLS is because of the different ways of reducing the dimension of the subspace spanned by the predictors. The EN method performs pure selection of the predictors to incorporate in the forecasting model, wehereas PLS extracts a small number of factors that weight all the variables. Zou and Hastie (2005) combine the L_1-penalty on the coefficients from the lasso regression (Tibshirani 1996) and the L_2-penalty on the coefficients from the ridge regression (Hoerl and Kennard, 1988) to construct the EN regression and find that it outperfoms both techniques. Bai and Ng (2008) consider the EN *soft-thresholding rules* to select the variables from the original data set and extract factors from this subset to forecast the target variable. This paper uses the EN method as a pure selection technique, whereas the selection is just an intermediate step for Bai and Ng (2008) before extracting the factors.

3 Data

The target variable is the logarithm of the US consumer price index (CPI). The choice of inflation as the target variable is because of the difficulty in forecasting it. In particular, it is difficult to improve the forecast performance with multivariate forecasting models, over a univariate benchmark. This is due to the changes in the inflation process since the mid-1980s (Stock and Watson, 2006). Following Fuentes et al. (2015), at least the second differences of the target variable has to be taken to obtain a covariance stationary series. This paper defines the target variable as defined in Fuentes et al. (2015):

$$y_{t+h}^{h} = \frac{1200}{h}(y_{t+h} - y_t) - 1200(y_t - y_{t-1}) \tag{1}$$

and

$$z_t = 1200(y_t - y_{t-1}) - 1200(y_{t-1} - y_{t-2}) \tag{2}$$

where z_t is considered as the lag of the target variable.

The Stock and Watson (2005) database is used to compute the forecast performance of PLS and PC. This data set contains 132 monthly US macroeconomic time series over the time frame from January 1959 until December 2003. The data set is not complete for every variable, such that the sample starts at January 1960 with in total 528 observations. To achieve stationarity, the series are transformed by taking logarithms and/or first or second differences. This is done in the same way as in Bai and Ng (2008) and Stock and Watson (2006). Following Fuentes et al. (2015) and Bai and Ng (2008), I divide the total sample into seven different forecast subsamples.

If the relation between the predictors and the target variable changes over time, then this is considered by using these different subsamples. In this way the dynamic features of the data are taken into account. The subsamples are shown in table 1.

Table 1: Estimation and forecast subsamples, h is the forecast horizon

SS	Estimation subsample	Forecast subsample
M1	1960:03 to 1970:03-h	1970:03 to 1980:12
M2	1960:03 to 1980:03-h	1980:03 to 1990:12
M3	1960:03 to 1990:03-h	1990:03 to 2000:12
M4	1960:03 to 1970:03-h	1970:03 to 1990:12
M5	1960:03 to 1970:03-h	1970:03 to 2000:12
M6	1960:03 to 1980:03-h	1980:03 to 2000:12
M7	1960:03 to 1970:03-h	1970:03 to 2003:12

4 Methodology

This section describes the framework and techniques that are used to forecast the target variable. First, the general forecasting framework is discussed. Next, for both the PLS and the EN method is explained how these are applied to this framework.

4.1 Partial Least Squares

The forecast of y_{t+h}^h, depends on the information up to time t of the target variable itself, predictors X_t and their lags. Following Fuentes et al. (2015), the general forecasting framework is defined as follows:

$$y_{t+h}^h = \mu + \phi(L)z_t + \beta'(L)\hat{F}_t + \eta_{t+h} \tag{3}$$

where y_{t+h}^h is the target variable to be forecasted h-step ahead, $\phi(L)$ and $\beta'(L)$ denote the lag polynomials such that $\phi(L)z_t$ and $\beta'(L)\hat{F}_t$ are the lags of the target variable and the estimated factors, $\hat{F}_t$, respectively, with their corresponding coefficients. The h-step-ahead prediction error is denoted by η_{t+h}.

The set of predictors can contain a huge amount of variables. However, I only use a small (relevant) part of these variables that contain the most useful information about the target variable. Following Fuentes et al. (2015), I use PLS which relates the reduction of the dimension among the predictors to the prediction of the target variable. The factors that are extracted by means of PLS are based on the covariance between the set of predictors and the target variable to be forecasted. To obtain these factors, I perform the eigenvalue decomposition on the following matrix:

$$M = X'YY'X \tag{4}$$

where Y is the vector of dimension T containing the target variable and X is the set of predictors with dimension TxN. Both include the information up to time $T = t$.
Making a linear combination of the first eigenvector of the matrix M and the variables in X, gives the first PLS factor $\hat{f}_{jt}^{PLS}$. For the second PLS factor, the matrices X and Y are different such that the second factor obtains information that is not contained in the first one. For simplicity I denote this new matrix as:

$$M = v'uu'v \tag{5}$$

where u and v are the vectors of residuals from the regressions of the target variable and the predictors, respectively, on the first PLS factor. For every next PLS factor, I perform the eigenvalue decomposition on the vectors of residuals from the regressions of the target variable and the predictors, respectively, on the previous PLS factors. Whereas PC only focuses on the variance between the predictors and the target variable, PLS also focuses on the correlation.

This technique can be used for static and dynamic applications, where this paper examines one static approach (a) and two dynamic approaches (b and c), as in Fuentes et al. (2015). The forecasting model consists of the forecasting equation (3) and

$$\hat{F}_t = WX_t, \tag{6}$$

which denotes the factors as a linear combination of the predictors as to be substituted in the forecasting equation. Following Fuentes et al. (2015), the static and dynamic approaches for the forecasting process are listed and elaborated below.

Static approach:

(a) For the static approach, I apply PLS as stated before to the matrix $M = X'Y_hY_hX$ where $Y_h = (y_{1+h}^h, ..., y_{T+h}^h)$ contains the target variable h periods ahead for every point of time t. Stated otherwise: if the forecast horizon is equal to one ($h = 1$) and the estimation sample is from March 1960 to February 1970, then Y_h contains the target variable y^h for the period April 1960 to March 1970. The matrix X contains the original set of predictors such that, using the previous example, it contains all variables for the period from March 1960 to February 1970. For extracting the factors, the lags of the target variable are not included whereas they are included in the forecasting equation, such that the dynamics of the target variable are not taken into account when constructing the factors. However, by including these lags in the forecasting equation, the dynamics of the target variable are taken into account directly.

Dynamic approaches:

(b) For the first dynamic approach, I apply PLS to the matrix $M = X_e' Y_h Y_h' X_e$ where Y_h is the target variable as explained in static approach (a) and X_e is the set of all original predictors and augmented with q lags of the target variable $(z_t, ..., z_{t-q+1})$. By setting $q = 1$ for the example as used above, the set X is augmented with the lags of the target variable for the observations March 1960 to February 1970. In this way the dynamics of the target variable are taken into account while extracting the factors. The lags of the target variable are not included in the forecasting equation.

(c) For the second dynamic approach, I first perform an AR(p) process on the target variable for the specific estimation sample. Next, I apply PLS to the matrix $M = X'\epsilon\epsilon'X$ where X is again the original set of predictors and ϵ is the vector of residuals from the estimated AR(p) process of the previous step. For this dynamic approach, I include lags of the target variable in the forecasting equation. In this way, the effect of the AR(p) process is considered solely before making use of the PLS estimation.

After extracting the factors by means of one of the approaches above, it is substituted in the forecasting equation (3) as a reduced set of explanatory variables.

The final forecasts are estimated by equation (3), where $\phi(L)z_t$ denotes the effect of the lags of the target variable and $\beta'(L)\hat{F}_t$ denotes the effect of the estimated factors and their lags. I estimate the factors and parameters with the information available up until time t. Following Fuentes et al. (2015) and Bai and Ng (2008), I use the Bayesian information Criterion (BIC) to determine the amount of lags of the factors and the target variable, with a maximum lag length of six when allowed by the sample size, and set to four otherwise. For both the static and dynamic approaches, I investigate the forecast horizons $h = 1$, 6, 12 and 24. I consider the forecast performance for different numbers of factors and choose the final number that coincides with the best performance.

4.2 Elastic Net

Another method to reduce the number of predictors that are used to forecast the target variable y^h_{t+h} is penalized regression. Multiple penalizing techniques together with their drawbacks and advantages are discussed in this section. The penalizing technique used in this paper is the elastic net, which is a combination between the ridge and the lasso penalizing techniques. The resulting problem is solved by means of the LARS-EN algortihm.

A different forecasting framework is used to predict the target variable by means of the EN:

$$y^h_{t+h} = \mu + \beta'(L)X_{e,t} + \eta_{t+h} \tag{7}$$

where $\beta'(L)$ denotes the lag polynomials such that $\beta'(L)X_{e,t}$ are the predictors and their lags, with their corresponding coefficients. To take into account the dynamics of the target variable, the lags of the target, z_t, as defined in equation (2), are included in the set of predictors, resulting in the enlarged matrix X_e. The EN penalizing technique selects the predictors to include in the forecasting model.

The ridge regression applies a L_2-penalty ($\sum_{j=1}^{N} \beta_j^2 = \|\beta\|_2^2$ where N is the number of predictors including z_t) on the regression coefficients and solves the following minimization problem:

$$\min_{\beta} \quad RSS + \lambda_{ridge} \sum_{j=1}^{N} \beta_j^2 \tag{8}$$

where RSS is the sum of squared residuals from a standard OLS regression with y^h_{t+h} as a target, and $\lambda_{ridge} \in [0, \infty)$ is the ridge penalty. The drawback regarding this technique is that the ridge penalty function only shrinks the regression coefficients towards zero, but never sets them equal to zero. Hence, it does not result in a parsimonious model.

The lasso regression substitutes the L_2-penalty with the L_1-penalty ($\sum_{j=1}^{N} |\beta_j| = \|\beta\|_1$), resulting in the problem:

$$\min_{\beta} \quad RSS + \lambda_{lasso} \sum_{j=1}^{N} |\beta_j| \tag{9}$$

where $\lambda_{lasso} \in [0, \infty)$ denotes the lasso penalty. An advantage of the lasso over the ridge regression is that it can set the values of the estimates equal to zero, such that it simultaneously shrinks the estimates and performs variable selection, implying that a parsimonious model can be created. However, Zou and Hastie (2005) address two drawbacks regarding the lasso regression: (1) in the case of $N > T$, the convex minimization problem constitutes to the fact that the lasso cannot select more than T predictors; (2) in the case of high pairwise correlations within a group of predictors, the lasso does not perform group selection but selects one predictor and makes no distinguish in which one to select.

The solution to the drawbacks as stated above is the EN regression, because besides the fact that it shrinks the coefficient estimates and simultaneously performs automatic variable selection, it also performs group selection for a group of predictors with high pairwise correlations. As stated by Zou and Hastie (2005) it is "a stretchable fishing net that retains 'all the big fish'". The EN regression combines the penalties of the lasso and the ridge regression by means of a convex combination. This results in the following minimization problem:

$$\min_{\beta} \quad RSS + \lambda_{EN} P_{\alpha}(\beta) \tag{10}$$

where

$$P_{\alpha}(\beta) = (1-\alpha)\sum_{j=1}^{N} |\beta_j| + \alpha \sum_{j=1}^{N} \beta_j^2 \tag{11}$$

where $\lambda_{EN} \in [0, \infty)$ and $\alpha \in (0, 1]$ are regularization parameters, following Zou and Hastie (2005) and Hastie et al. (2008). When α is strictly between 0 and 1, the EN regression is performed. When α is equal to one, the problem is equal to the lasso. When α moves towards zero, the problem approaches the ridge regression. To follow the denotation of Bai and Ng (2008), I rewrite the equation and substitute $\lambda_1 = \lambda_{EN}(1-\alpha)$ and $\lambda_2 = \lambda_{EN}\alpha$ to obtain the following problem:

$$\min_{\beta} \quad RSS + \lambda_1 \sum_{j=1}^{N} |\beta_j| + \lambda_2 \sum_{j=1}^{N} \beta_j^2 \tag{12}$$

from which the penalty is strictly convex if $\frac{\lambda_2}{\lambda_1+\lambda_2} > 0$.

An advantage of the EN is that by adjusting the data set by

$$X_e^* = (1+\lambda_2)^{-1/2} \begin{pmatrix} X_e \\ \sqrt{\lambda_2} I_N \end{pmatrix} \quad \text{and} \quad y^* = \begin{pmatrix} y \\ 0_N \end{pmatrix}, \tag{13}$$

such that X_e^* is a $(T+N)$xN matrix and y^* is a vector of length $(T+N)$, the EN criterion can be rewritten as a lasso problem which can be solved by means of the LARS-EN algorithm. For $\gamma = \frac{\lambda_1}{\sqrt{1+\lambda_2}}$ the EN estimator can be rewritten as

$$\beta^{**} = \underset{\beta}{\operatorname{argmin}}\, RSS^* + \gamma \sum_{j=1}^{N} |\beta_j|. \tag{14}$$

Before the LARS-EN algorithm can be performed, the double-shrinkage effect has to be deleted. The double-shrinkage effect occurs because both the lasso and the ridge penalty perform shrinkage to the coefficient estimates. Zou and Hastie (2005) introduce $\beta^* = (1+\lambda_2)\beta^{**}$ as the EN estimator that solves this problem.

4.3 LARS-EN Algorithm

The LARS ('least angle regression') algorithm proposed by Efron et al. (2004) iteratively estimates the target variable in a piecewise linearly way taking into account the maximum correlation with the residuals. The LARS-EN algortihm used in this paper, proposed by Zou and Hastie (2005), is based on the LARS algorithm.

Let $\hat{\mu}_k$ be the estimate of the target variable in the k^{th}-step, with starting value $\mu_0 = 0$ in the first step. In every step, one predictor is added to the active set, such that k is the number of predictors after k steps. This implies that the coefficients of the other $N - k$ predictors are equal to zero. Let $\hat{c} = X_e^{*\prime}(y^* - \hat{\mu}_k)$ be the vector of dimension N of correlations between the set of predictors and the residual vector in the k^{th}-step. The maximum correlation is defined as

$$\hat{C} = \max_j |\hat{c}_j| \quad \text{such that} \quad K = \{j : |\hat{c}_j| = |\hat{C}|\} \tag{15}$$

where K is the set of indices that correspond with the predictors that contain the largest correlations in absolute value. With this, the active variable set corresponding to K, in the k^{th}-step is denoted by

$$X^*_{e,K_k} = (\text{sign}(\hat{c}_j)x^*_j)_{j \in K}. \tag{16}$$

Let $G_{K_k} = X^{*'}_{e,K_k} X^*_{e,K_k}$ such that with the transformed data set this is

$$G_{K_k} = \frac{1}{1 + \lambda_2}(X^{*'}_{e,K_k} X^*_{e,K_k} + \lambda_2 I_N). \tag{17}$$

The inverse of this matrix is needed to construct $A_{K_k} = (1'_K G^{-1}_{K_k} 1_K)^{-1/2}$ (where 1_K is a Kx1 vector of ones), which is done by up- or downdating the Cholesky factorization of the $G_{K_{k-1}}$ matrix from the previous step. Zou and Hastie (2005) found that the up- or downdating of the Cholesky factorization of $X^{*'}_{e,K_{k-1}} X^*_{e,K_{k-1}} + \lambda_2 I_N$ is done in the exact same way as is done on $X^{*'}_{e,K_{k-1}} X^*_{e,K_{k-1}}$ by Golub and Van Loan (1983). When A_{K_k} is constructed, the equiangular vector u_{K_k} is denoted by

$$u_{K_k} = X^*_{e,K_k} w_{K_k} \quad \text{where} \quad w_{K_k} = A_{K_k} G^{-1}_{K_k} 1_K \quad \text{and} \quad a_{K_k} = X^{*'}_{e,K_k} u_{K_k}. \tag{18}$$

The LARS-EN algorithm then updates the target variable in every step:

$$\hat{\mu}_{k+1} = \hat{\mu}_k + \hat{\gamma} u_{K_k} \tag{19}$$

where

$$\hat{\gamma} = \min_{j \in K^c}{}^{+} \left(\frac{\hat{C} - \hat{c}_j}{A_K - a_j}, \frac{\hat{C} + \hat{c}_j}{A_K + a_j} \right). \tag{20}$$

4.4 Choice of tuning parameters

The parameters λ_{EN} en α from equation (10) and (11) are considered as the tuning parameters of the EN regression. A widely used method, and also used by Zou and Hastie (2005), is the 10-fold cross-validation (CV) to compute the prediction error of the estimation sample. I use the 10-fold CV and choose to use the largest value of λ_{EN} for which the prediction error is within one standard error from the minimum mean squared error (MSE). Choosing the tuning parameter α implies to what extent the EN penalty interpolates between the ridge penalty and the lasso penalty (Hastie et al., 2008). To choose a value for the tuning parameter α, I compare forecasts for three different values ($\alpha = 0.25$, 0.5 and 0.75) and allow the final value to be determined by the forecasting performance.

5 Forecast Results

The forecasting performance of the PC method and the three PLS approaches for the forecast horizons $h = 1$, 6, 12 and 24 are shown in tables 2, 3, 4 and 5, respectively. Following Bai and Ng (2008), I use $k = 10$ factors for the PC regression. Regarding PLS, I investigate both the forecast performance for $k = 1$ and $k = 2$ factors and let the number of factors be determined by the best forecasting performance. To compare the predictive ability of the different methods, I use the *relative mean-squared error* (RMSE) over a univariate benchmark, as in Fuentes et al. (2015). The univariate benchmark used for the forecast horizon $h = 1$ is an AR(4) process. For the horizons $h = 6$, 12 and 24, I fit the target variable to an AR(4) process, z_t and three lags. This gives the following RMSE:

$$\text{RMSE (method)} = \frac{\text{MSE (method)}}{\text{MSE (AR(4))}}. \tag{21}$$

From the RMSE I can clearly observe whether a method performs better or worse than the benchmark. Namely, when the RMSE has a value of less than one, the method has a better forecast performance than the AR(4) model.

5.1 Fixed Estimation Sample

Tables 2-5 present the results of the forecasting performances of the PC method and the three different approaches of PLS as discussed in section 4.1 together with the EN method for the different forecast horizons. For these forecasts a fixed estimation sample is used. With a fixed estimation sample, I refer to the fixed number of observations used to estimate the factors and regression coefficients for the forecast sample. For example, for forecasting 1970:03 to 1980:12, all forecasts use the factors and parameters estimated for the sample 1960:03 to 1970:03-h. This is done differently than in Fuentes et al. (2015), where an expanding estimation sample is used. Regarding the same example, for the first forecast (1970:03) an expanding estimation uses factors and parameters estimated for the sample 1960:03 to 1970:03-h. However, the last forecast (1980:12) is based on estimation up to 1980:12-h. The difference in the estimation samples is taken into account when interpreting the results. The results in tables 2-5 give multiple remarkable insights to the forecasting models regarding the increasing horizon and when comparing the results with those of Fuentes et al. (2015).

Table 2: RMSE, $h = 1$

	PC	PLS			EN
Period	($k = 10$)	a ($k = 1$)	b ($k = 2$)	c ($k = 2$)	($k =$ max. 10)
M1: 70.3-80.12	0.991	1.052	1.080*	0.988	**0.872**$^{\alpha=0.5}$
M2: 80.3-90.12	1.073	1.002	1.045*	1.077	**0.959**$^{\alpha=0.75}$
M3: 90.3-00.12	0.964	1.013	1.078	**0.924**	0.947$^{\alpha=0.5}$
M4: 70.3-90.12	0.970	1.015	1.132*	0.992	**0.945**$^{\alpha=0.75}$
M5: 70.3-00.12	0.958	1.007	1.135*	0.984	**0.946**$^{\alpha=0.5}$
M6: 80.3-00.12	1.049	0.992	1.105	0.967	**0.934**$^{\alpha=0.75}$
M7: 70.3-03.12	0.983	1.014	1.139*	0.988	**0.930**$^{\alpha=0.75}$

Note: The table presents the RMSE of PC, PLS and EN over the benchmark for the 1-month forecast horizon. An asterisk means $k = 1$. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN.

Table 3: RMSE, $h = 6$

Period	PC	PLS			EN
	($k = 10$)	a ($k = 1$)	b ($k = 2$)	c ($k = 2$)	($k =$ max. 10)
M1: 70.3-80.12	0.713	0.592	1.012	**0.591**	$0.731^{\alpha=0.75}$
M2: 80.3-90.12	0.646	**0.594***	1.189	0.619	$0.887^{\alpha=0.75}$
M3: 90.3-00.12	0.669	**0.564***	1.029	0.651	$0.715^{\alpha=0.5}$
M4: 70.3-90.12	0.668	**0.568**	1.092	0.631	$0.843^{\alpha=0.75}$
M5: 70.3-00.12	0.675	**0.539**	1.068	0.612	$0.821^{\alpha=0.5}$
M6: 80.3-00.12	0.648	**0.563***	1.197	0.610	$0.815^{\alpha=0.75}$
M7: 70.3-03.12	0.671	**0.526**	1.067	0.597	$0.826^{\alpha=0.75}$

Note: The table presents the RMSE of PC, PLS and EN over the benchmark for the 6-month forecast horizon. An asterisk means k=1. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN.

Table 4: RMSE, $h = 12$

Period	PC	PLS			EN
	($k = 10$)	a ($k = 1$)	b ($k = 2$)	c ($k = 2$)	($k =$ max. 10)
M1: 70.3-80.12	0.648	0.558	0.810	**0.552**	$0.578^{\alpha=0.75}$
M2: 80.3-90.12	0.613	**0.497**	1.232	0.521	$0.787^{\alpha=0.75}$
M3: 90.3-00.12	0.739	**0.511**	1.185	0.649	$0.691^{\alpha=0.75}$
M4: 70.3-90.12	0.664	**0.564**	0.887	0.565	$0.601^{\alpha=0.75}$
M5: 70.3-00.12	0.668	**0.544**	0.871	0.549	$0.607^{\alpha=0.75}$
M6: 80.3-00.12	0.620	0.512	1.329	**0.509**	$0.737^{\alpha=0.75}$
M7: 70.3-03.12	0.647	**0.531**	0.881	0.532	$0.625^{\alpha=0.75}$

Note: The table presents the RMSE of PC, PLS and EN over the benchmark for the 12-month forecast horizon. An asterisk means k=1. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN.

To start, I find that there is improvement with respect to the widely used PC method. For the 1-month forecast horizon the forecast performances of the static PLS approach (a) are weaker then those of the PC method. However, for the remaining forecast horizons this PLS approach performs better than the PC method for almost every subsample.

Table 5: RMSE, $h = 24$

Period	PC	PLS			EN
	($k=10$)	a ($k=1$)	b ($k=2$)	c ($k=2$)	($k=$ max. 10)
M1: 70.3-80.12	0.536	**0.526**	0.649	0.532	$0.634^{\alpha=0.75}$
M2: 80.3-90.12	0.550	0.628*	1.223	**0.515**	$0.734^{\alpha=0.75}$
M3: 90.3-00.12	0.694	1.518	3.487	**0.593**	$0.641^{\alpha=0.75}$
M4: 70.3-90.12	0.539	**0.531**	0.844	0.547	$0.638^{\alpha=0.75}$
M5: 70.3-00.12	0.547	**0.502**	0.849	0.525	$0.622^{\alpha=0.75}$
M6: 80.3-00.12	0.554	0.702*	1.825	**0.506**	$0.689^{\alpha=0.75}$
M7: 70.3-03.12	0.530	**0.515**	0.853	0.521	$0.626^{\alpha=0.5}$

Note: The table presents the RMSE of PC, PLS and EN over the benchmark for the 24-month forecast horizon. An asterisk means k=1. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN.

Second, the general pattern of the results show that the forecast accuracy keeps improving over to the univariate benchmark as the forecast horizon gets larger. This holds for all methods. This is also found by Fuentes et al. (2015) and before in the factor model literature. However, the results found by Fuentes et al. (2015) show better forecast performances for all methods when the forecast horizon increases. These differences may be due to the fact that I am using a fixed estimation sample, whereas Fuentes et al. (2015) uses an expanding estimation sample. The estimation sample expands with the size of the horizon every time a forecast is made, such that the number of observations for estimation increases for every forecast. The estimation sample used in this paper is fixed, such that there is a fixed number of observations to estimate the parameters and factors for the complete forecast sample. This might explain the better results obtained in Fuentes et al. (2015) and the increasing difference in results for larger forecast horizons.

Third, the results show that the PLS approach (b) performs (much) worse than the other methods and in many cases even worse than the benchmark. This is due to the way of incorporating the dynamic behaviour of the target variable into the forecasting model. In this approach the dynamics of the target variable are taken into account by incorporating the lags of the target variable as additional predictors for extracting the factors, whereas they are not included directly into the forecasting equation. PLS assigns a weight to all the predictors, in this case including the lags of the target variable. Because the number of variables is large, the weights associated with the lags of the target are considerably smaller compared to when they are incorporated directly into the forecasting equation (as in approaches (a) and (c)), such that it does not sufficiently capture the dynamics of the variable to forecast. Besides this, I observe that the performances for the subsamples M2 and M6 for all forecast horizons are much lower compared to those of Fuentes et al. (2015). This might imply that using PLS approach (b) with a fixed estimation sample does not account for the changing relation over time between the predictors and the target variable as is done with the expanding estimation sample. This is also the case for the results of PLS approach (a) considered for subsample M2 and M6 for the forecast horizon $h = 24$.

Another observation regarding the results from PLS approach (b) is that some forecasts are better than in Fuentes et al. (2015) considered for the horizons $h = 1$, 6 and 12. This difference may arise from the choice of the lags of the target variable in the enlarged matrix of predictors X_e. This choice is not clearly defined in Fuentes et al. (2015), although they may have chosen to use y_t and several lag, whereas I augment the matrix with z_t as defined in equation (2). The choice of z_t as the lags is because they are defined as the lags of the transformed target variable, whereas y_t are the lags of the natural logarithm of the CPI. Besides this, I also use z_t as lags in the forecasting equation for approaches (a) and (c), such that this research consistently uses z_t as lags of the target variable. With this, I obtain considerably better results than Fuentes et al. (2015), in particular for the 12-month forecast horizon.

Fourth, the results regarding approach (c) outperform the widely used PC method in 82% of all subsamples. This implies that considering the effect of the AR(p) process before PLS estimation seems appropriate to capture the dynamic relationship of the target variable. However, the results are adjusted such that they are comparable to the other approaches and to the results from Fuentes et al. (2015). These adjustments are made by changing the number of lags of the target variable and the factors in the forecasting equation. Defining the number of lags by means of the BIC results in much better forecasting performance than those obtained by Fuentes et al. (2015). Such strong forecast performance seems non applicable, so that adjustment had to be made. The results regarding approach (c) are obtained by fitting an AR(1) process to the target variable to obtain the residuals for the eigenvalue decomposition, whereas Fuentes et al. (2015) does not state which AR(p) process to apply. I have tried multiple alternatives to obtain the residuals, from which the results are shown in Appendix A. None of the alternatives results in the RMSE comparable to Fuentes et al. (2015).

Fifth, the EN method shows the best forecast performance considering the 1-month forecast horizon for six out of seven subsamples. However, when the forecast horizon increases, the forecast performance of the EN improves over the benchmark, but not with the same amplitude as the other methods. This results in the fact that the EN shows weak forecast performance for the remaining forecast horizons. The EN also shows weaker forecast performances considered for the 24-month forecast horizon compared to the 12-month horizon.

Sixth, PLS approaches (a) and (b) for the subsample M3 considered for the 24-month forecast horizon show significantly weak forecast performance. This implies that these approaches cannot capture the dynamics of the target variable for this specific sample. In particular, PLS approach (b) performs more than three times worse than the benchmark. Approach (a) accounts for some more of the temporal instability because of the incorporated lags of the target in the forecasting equation.

At last, PLS approach (a) performs the best in almost all samples for the forecast horizons $h = 6$, 12 and 24. This indicates that applying PLS with a static approach extracts the most relevant information about the target variable when using a fixed estimation sample.

5.2 Expanding Estimation Sample

Tables 6-9 show the forecast performance of PC, LARS and the EN over the univariate benchmark. For this empirical comparison an expanding estimation sample is used as described in the previous section. The forecasting results with respect to PC and LARS are obtained from Fuentes et al. (2015), for which PC uses $k = 10$ for the number of factors and LARS for both $k = 5$ and 10. I have decided to keep a maximum of ten variables selected by the EN for a good comparison with LARS.

Table 6: RMSE, $h = 1$

	PC	LARS		EN
Period	$k = 10$	$k = 5$	$k = 10$	$k = \text{max. } 10$
M1: 70.3-80.12	**1.015**	1.102	1.190	$1.370^{\alpha=0.75}$
M2: 80.3-90.12	0.982	1.018	1.022	$\mathbf{0.910}^{\alpha=0.75}$
M3: 90.3-00.12	**0.963**	1.015	1.025	$1.192^{\alpha=0.75}$
M4: 70.3-90.12	**0.998**	1.067	1.112	$1.118^{\alpha=0.75}$
M5: 70.3-00.12	**0.990**	1.059	1.098	$1.130^{\alpha=0.75}$
M6: 80.3-00.12	**0.972**	1.019	1.025	$0.974^{\alpha=0.75}$
M7: 70.3-03.12	**0.979**	1.047	1.092	$1.138^{\alpha=0.75}$

Note: The table presents the RMSE of PC, LARS and the EN over the benchmark for the 1-month forecast horizon. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN. Results of PC and LARS are obtained from Fuentes et al. (2015).

Table 7: RMSE, $h = 6$

	PC	LARS		EN
Period	$k = 10$	$k = 5$	$k = 10$	$k = \text{max. } 10$
M1: 70.3-80.12	0.712	0.786	0.719	$\mathbf{0.608}^{\alpha=0.75}$
M2: 80.3-90.12	**0.654**	0.789	0.794	$0.925^{\alpha=0.75}$
M3: 90.3-00.12	0.660	0.986	1.066	$\mathbf{0.591}^{\alpha=0.75}$
M4: 70.3-90.12	**0.675**	0.815	0.789	$0.699^{\alpha=0.75}$
M5: 70.3-00.12	0.671	0.825	0.810	$\mathbf{0.658}^{\alpha=0.75}$
M6: 80.3-00.12	**0.652**	0.808	0.826	$0.800^{\alpha=0.75}$
M7: 70.3-03.12	0.670	0.817	0.803	$\mathbf{0.635}^{\alpha=0.75}$

Note: The table presents the RMSE of PC, LARS and the EN over the benchmark for the 6-month forecast horizon. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN. Results of PC and LARS are obtained from Fuentes et al. (2015).

The first observation regarding the EN method for the tables 6-9 is the weak forecast performance considered for the 1-month forecast horizon. The EN performs worse than the widely

Table 8: RMSE, $h = 12$

Period	PC	LARS		EN
	$k = 10$	$k = 5$	$k = 10$	$k =$ max. 10
M1: 70.3-80.12	0.631	0.606	**0.554**	$0.586^{\alpha=0.75}$
M2: 80.3-90.12	**0.575**	0.641	0.710	$0.869^{\alpha=0.75}$
M3: 90.3-00.12	0.723	1.032	0.989	$\mathbf{0.694}^{\alpha=0.75}$
M4: 70.3-90.12	**0.603**	0.624	0.626	$0.650^{\alpha=0.75}$
M5: 70.3-00.12	**0.611**	0.670	0.666	$0.638^{\alpha=0.75}$
M6: 80.3-00.12	**0.594**	0.717	0.765	$0.800^{\alpha=0.75}$
M7: 70.3-03.12	**0.609**	0.680	0.671	$0.630^{\alpha=0.75}$

Note: The table presents the RMSE of PC, LARS and the EN over the benchmark for the 12-month forecast horizon. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN. Results of PC and LARS are obtained from Fuentes et al. (2015).

Table 9: RMSE, $h = 24$

Period	PC	LARS		EN
	$k = 10$	$k = 5$	$k = 10$	$k =$ max. 10
M1: 70.3-80.12	0.532	0.539	0.542	$\mathbf{0.464}^{\alpha=0.75}$
M2: 80.3-90.12	**0.506**	0.535	0.545	$0.977^{\alpha=0.75}$
M3: 90.3-00.12	**0.546**	0.975	0.767	$0.679^{\alpha=0.75}$
M4: 70.3-90.12	0.522	0.537	0.547	$\mathbf{0.340}^{\alpha=0.75}$
M5: 70.3-00.12	0.523	0.572	0.564	$\mathbf{0.482}^{\alpha=0.75}$
M6: 80.3-00.12	**0.512**	0.599	0.576	$0.886^{\alpha=0.75}$
M7: 70.3-03.12	0.523	0.574	0.565	$\mathbf{0.495}^{\alpha=0.75}$

Note: The table presents the RMSE of PC, LARS and the EN over the benchmark for the 24-month forecast horizon. The best forecasting performance for each subsample is indicated in bold. The value of the tuning parameter α is indicated for the EN. Results of PC and LARS are obtained from Fuentes et al. (2015).

used PC method and both options of LARS, for five out of seven subsamples. However, when the forecast horizon increases to 6 months, the EN performs much better and yields the best forecast performance for four subsamples.

A second interesting observation is the value for the tuning parameter α for the EN. This value is equal to 0.75 for every subsample over all forecast horizons, implying that the EN penalization interpolates more towards the lasso penalty than the ridge penalty. Because of the fact that the ridge penalty never sets the coefficients equal to zero, the value of α indicates that there are many zero coefficients in the true model.

Third, the results of the EN show inconsistent improvement with the increasing forecast horizon. With this, I refer to the improvements made over the benchmark when the forecast horizon increases. Namely, when the horizon increases to $h = 12$, the EN yields better forecasts than for $h = 6$, but the improvement is much less compared to the previous improvement when the forecast horizon increases from $h = 1$ to $h = 6$. Also, the improvement of the EN compared to the other methods is much lower, such that the EN has a weaker forecast performance than PC and LARS for the 12-month forecast horizon. However, when the forecast horizon increases further to $h = 24$, the EN performs again much better for four subsamples, yielding the four best forecasts.

Third, the forecast accuracy regarding the EN method considered for the subsamples M2 and M6 are considerably lower compared to the other subsamples. The results of the two subsamples for the 24-month forecast horizon are even worse than for the 12-month horizon, something that is also observed for the EN with a fixed estimation sample.

Fourth, solely comparing LARS with the EN shows that the EN performs better than LARS for only 57% of the time. This implies that none of the two penalizing techniques consistently dominates the other. However, the subsamples for which the EN has a stronger forecast performance the results differ substantially from those of LARS. This might be due to the property of group selection that is performed by the EN. Groups of variables with high pairwise correlations might be present in those subsamples, such that the EN has a better forecast accuracy than LARS.

6 Conclusions

This paper revisits partial least squares (PLS), following Fuentes et al. (2015), to investigate whether the forecast performance can be improved over the widely used principal components (PC) method when reducing the dimension of the subspace spanned by the predictors and relating this feature to the forecast goal. One static and two dynamic approaches of the PLS method are applied to 132 monthly US macroeconomic time series. Another method to reduce the dimension among the predictors applied in this study is the elastic net (EN), a regularization technique that combines the penalties of two penalizing techniques to simultaneously perform shrinkage and variable selection. The target is the US inflation where the estimation sample regarding PLS is fixed, whereas the EN is considered for both a fixed estimation sample and an expanding estimation sample.

Regarding the three different approaches of PLS, results show that including the lags directly into the forecasting equation yields much better results than incorporating these lags to the original set of predictors. This is due to the fact that PLS weights all the predictors. The weights associated with the lags of the target are much less compared to when they are incorporated directly into the forecasting equation, such that it does not sufficiently capture the dynamics of the target variable. Static approach (a) includes the lags of the target directly into the forecasting equation, which yields better forecasting results than the widely used PC method in 71% of all subsamples.

Dynamic PLS approach (c) also includes the lags of the target variable directly into the forecasting equation, but isolates the AR(p) process before PLS estimation. This approach outperforms the PC method for almost all subsamples, implying that the alternative way of incorporating the dynamics of the target variable seems to capture the dynamic relationship appropriately. However, these results should be interpreted carefully, since the lags in the forecasting equation had to be adjusted to obtain comparable results to Fuentes et al. (2015).

The EN with a fixed estimation sample performs good for the very short forecast horizon ($h = 1$), but shows no improvement over PC when the forecast horizon increases. When the EN is performed with an expanding estimation sample, it shows weak forecast performance for the very short forecast horizon, but the forecast accuracy improves when the forecast horizon gets longer. In particular, the EN has the strongest forecasting performance for four subsamples considered for the 24-month forecast horizon. Yet, PC shows better forecast accuracy than the EN for 61% of the subsamples. Hence, improvements over PC are found, but the EN cannot be regarded as an consistent improvement over PC.

Overall, the empirical comparison shows that reducing the subspace among predictors by extracting factors from the most informative predictors to predict a specific target variable improves forecasting performance of the most widely used factor method.

Additionally, the EN has the property to select variables for every period a forecast is made. It can be of great interest for policy makers to know which predictors consist of the most informative content about the target variable. These could be observed and interpreted so that the EN could be used as an exploratory tool for an additional gain to this study.

References

[1] Bai J, Ng S. 2002. Determining the number of factors in approximate factor models. *Econometrica* **70**: 191-221.

[2] Bai J, Ng S. 2008. Forecasting economic time series using targeted predictors. *Journal of Econometrics* **146**: 304–317.

[3] Bernanke B, Boivin J, Eliasz P. 2002. Factor augmented vector autoregressions (FVARs) and the analysis of monetary policy. *Quarterly Journal of Economics* **120**: 387-422.

[4] Bernanke B, Boivin J. 2003. Monetary policy in a data rich environment. *Journal of Monetary Economics* **50**(3): 525–546.

[5] Boivin J, Ng S. 2006. Are more data always better for factor analysis? *Journal of Econometrics* **132**: 169-194.

[6] Chan Y, Stock J, Watson M.W. 1998. A dynamic factor model framework for forecast combinations. *Journal of Forecasting* **23**: 405-430.

[7] De Mol C, Giannone D, Riechlin L. 2008. Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components? *Journal of Econometrics* **146**(2): 318-328.

[8] Efron B, Hastie T, Johnstone I, Tibshirani R. 2004. Least angle regression. *Annals of Statistics* **32**(2): 407-499.

[9] Eickmeier S, Ng T. 2011. Forecasting national activity using lots of international predictors: an application to New Zealand. *International Journal of Forecasting* **27**(2): 496–511.

[10] Forni M, Hallin M, Lippi M, Reichlin L. 2005. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association* **100**(471): 830–840.

[11] Friedman J, Tibshirani R, Hastie T. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**, No. 1.

[12] Fuentes J, Poncela P, Rodriquez J. 2015. Sparse partial least squares in time series for macroeconomic forecasting. *Journal of Applied Econometrics* **30**: 576-595.

[13] Golub G, Van Loan C. 1983. Matrix Computations, 3rd edition. *Johns Hopkins University Press.*

[14] Groen J, Kapetanios G. 2008. Revisiting useful approaches to data-rich macroeconomic forecasting. *Federal Reserve Bank of New York Staff Report* 327.

[15] Groen J, Kapetanios G. 2008; revised 2015. Revisiting useful approaches to data-rich macroeconomic forecasting. *Federal Reserve Bank of New York Staff Report* 327.

[16] Hastie T, Tibshirani R, Friedman J. 2008. The Elements of Statistical Learning, 2nd edition. *Springer, New York.*

[17] Hoerl A, Kennard R. 1988. Ridge regression. *Encyclopedia of Statistical Sciences* **8**: 129–136.

[18] Inoue A, Kilian L. 2008. How useful is bagging in forecasting economic time series? A case study of U.S. CPI inflation. *Journal of the American Statistical Association* **103**: 511–522.

[19] Kelly B, Pruitt S. 2012. The three-pass regression filter: a new approach to forecasting using many predictors. *Chicago Booth Paper* No. 11-19.

[20] Matheson T.D. 2006. Factor model forecasts for New Zealand. *International Journal of Central Banking* **2**(2): 169-237.

[21] Stock J, Watson M. 2002a. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97**(460): 1167-1179.

[22] Stock J, Watson M. 2002b. Macroeconomic forecasting using duffusion indexes. *Journal of Business and Economic Statistics* **20**: 147-163.

[23] Stock J, Watson M. 2005. Implications of dynamic factor models for VAR analysis. *NBER Working Paper* No. 11467.

[24] Stock J, Watson M. 2006. Forecasting with many predictors. *Handbook of Economic Forecasting. Elsevier Science* **1**: 516–542.

[25] Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society, Series B* **58**: 267–288.

[26] Wold H. 1996. Estimation of principal components and related models by iterative least squares. *Multivariate Analysis, Krishnaiaah PR (ed.). Academic Press: New York*: 391-420.

[27] Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *Journal of Royal Statistical Society, Series B* **67**(2): 301-320.

Appendix A: Alternative results PLS approach (c)

Table A1: RMSE PLS approach (c) for different alternatives of residuals, $h = 1$

Period	Residual alternatives		
	$\epsilon = \mathrm{y}^h_{t+h} - \phi \mathrm{y}^h_t$	$\epsilon = \mathrm{y}^h_{t+h} - \phi \mathrm{y}^h_{t+h}(BIC)$	$\epsilon = \mathrm{y}^h_{t+h} - \phi \mathrm{z}_t$
M1: 70.3-80.12	0.612*	0.687*	0.653*
M2: 80.3-90.12	0.763*	0.842*	0.766*
M3: 90.3-00.12	0.674*	0.756*	0.815*
M4: 70.3-90.12	0.815	0.857	0.809
M5: 70.3-00.12	0.790*	0.787*	0.790*
M6: 80.3-00.12	0.967*	0.801*	0.803*
M7: 70.3-03.12	0.785	0.786	0.791

Note: The table presents the RMSE of PLS approach (c) ($k = 2$) over the benchmark for the 1-month forecast horizon. The second row denotes the different alternatives to obtain the residuals, as tried for PLS approach (c). BIC denotes the number lags p used for the AR(p) process as determined by the BIC. An asterisk means k=1.

Table A2: RMSE PLS approach (c) for different alternatives of residuals, $h = 6$

Period	Residual alternatives		
	$\epsilon = \mathrm{y}^h_{t+h} - \phi \mathrm{y}^h_t$	$\epsilon = \mathrm{y}^h_{t+h} - \phi \mathrm{y}^h_{t+h}(BIC)$	$\epsilon = \mathrm{y}^h_{t+h} - \phi \mathrm{z}_t$
M1: 70.3-80.12	0.591	0.563	0.585
M2: 80.3-90.12	0.458*	0.482*	0.518*
M3: 90.3-00.12	0.322*	0.536*	0.577*
M4: 70.3-90.12	0.515	0.551	0.546
M5: 70.3-00.12	0.488	0.529	0.530
M6: 80.3-00.12	0.430*	0.498*	0.517*
M7: 70.3-03.12	0.496*	0.514*	0.523*

Note: The table presents the RMSE of PLS approach (c) ($k = 2$) over the benchmark for the 6-month forecast horizon. The second row denotes the different alternatives to obtain the residuals, as tried for PLS approach (c). BIC denotes the number lags p used for the AR(p) process as determined by the BIC. An asterisk means k=1.

Table A3: RMSE PLS approach (c) for different alternatives of residuals, $h = 12$

Period	Residual alternatives		
	$\epsilon = y^h_{t+h} - \phi y^h_t$	$\epsilon = y^h_{t+h} - \phi y^h_{t+h}(BIC)$	$\epsilon = y^h_{t+h} - \phi z_t$
M1: 70.3-80.12	0.552	0.584	0.579
M2: 80.3-90.12	0.521	0.542	0.535
M3: 90.3-00.12	0.349*	0.537*	0.560*
M4: 70.3-90.12	0.565	0.577	0.578
M5: 70.3-00.12	0.549	0.555	0.563
M6: 80.3-00.12	0.509	0.562	0.540
M7: 70.3-03.12	0.532	0.539	0.548

Note: The table presents the RMSE of PLS approach (c) ($k = 2$) over the benchmark for the 12-month forecast horizon. The second row denotes the different alternatives to obtain the residuals, as tried for PLS approach (c). BIC denotes the number lags p used for the AR(p) process as determined by the BIC. An asterisk means k=1.

Table A4: RMSE PLS approach (c) for different alternatives of residuals, $h = 24$

Period	Residual alternatives		
	$\epsilon = y^h_{t+h} - \phi y^h_t$	$\epsilon = y^h_{t+h} - \phi y^h_{t+h}(BIC)$	$\epsilon = y^h_{t+h} - \phi z_t$
M1: 70.3-80.12	0.523	0.544	0.563
M2: 80.3-90.12	0.424	0.496	0.507
M3: 90.3-00.12	0.536*	0.527*	0.534*
M4: 70.3-90.12	0.511	0.554	0.585
M5: 70.3-00.12	0.488	0.519	0.547
M6: 80.3-00.12	0.412	0.531	0.527
M7: 70.3-03.12	0.456	0.492	0.522

Note: The table presents the RMSE of PLS approach (c) ($k = 2$) over the benchmark for the 24-month forecast horizon. The second row denotes the different alternatives to obtain the residuals, as tried for PLS approach (c). BIC denotes the number lags p used for the AR(p) process as determined by the BIC. An asterisk means k=1.

Appendix B: Programs

- **EN_Fix_Est_Sample**: This program calculates the mean squared errors for the Elastic Net with a fixed estimation sample.
- **EN_Exp_Est_Sample**: This program calculates the mean squared errors for the Elastic Net with an expanding estimation sample.
- **PLS**: This program estimates the factors by means of the three different approaches of PLS, which are exported to EViews to forecast the target variable.
- **PC**: This program estimates the factors by means of the PC method, which are exported to EViews to forecast the target variable.