Forecasting with Factor Models: A Subset Problem

Leander Tijssen (334864)

July 1, 2013

Abstract

This paper evaluates a modification done on the DI framework of Stock and Watson (2002). The regressors used for PCA are first evaluated for their predictive power. As there are several techniques to create such subsets, five are used of which two are hard thresholding and three are soft thresholding methods. The 1, 6, 12 and 24 month ahead point-forecasts are then compared to a 3 factor benchmark with factors constructed from the complete dataset. Results show that selecting relevant predictors before applying principal component analysis yields improved forecasts.

1 Introduction

A common problem found when forecasting macro-economic variables is that a large set of predictors are potentially relevant but only a few, if there even are any, hold high predictive power.

When dealing with a large set of predictors often only a few are used for macro-economic forecasting. For example: if one wants to predict the inflation researchers historically would use the real economic activity, but perhaps several other variables are of importance as well. A way to decrease but still hold the information of the large set of predictors is, as Stock and Watson (2002) have shown, the use of diffusion indexes. These DI forecasts outperform univariate autoregressions, small vector autoregressions and leading indicator models. The diffusion indexes are first constructed by principal components analysis and then used to create a forecast on the targeted variable, y_t.

Although the method of Stock and Watson (2002) ensures that a huge amount of information from the predictors is stored in a small amount of factors, still some improvements on their method can be made. For example several combination schemes as shown by van Dongen et al. (2013) can improve the predictive power of the factor models. Another refinement to the DI framework is done by Bai and Ng (2007) as they state, just like Boivin and Ng (2006), that while the DI methodology captures a huge amount of the information of all the data some of these predictors might not enhance the predictive ability of the model when the factors are estimated. Therefore they suggest to perform principal components on a subset of predictors that have shown to have predictive power for the dependent variable.

As Bai and Ng (2007) focus on finding the best subset which enhances the predictive power of the model Inoue and Kilian (2005) use bagging to improve forecasts. Bagging is a weighting scheme that focuses on finding the subset with relevant predictors in the bootstrap re-samples. The forecasts made by the subsets created by the Bagging Dynamic Regression Model algorithm are combined by averaging them and so the factor based bagging estimation is formed.

The main interest in this paper will lie on finding a subset of relevant predictors to enhance forecasts of the diffusion index framework. There are several ways to find these subsets such as the method of Inoue and Kilian (2005) who used hard thresholding at each iteration of their algorithm to reduce forecasting variance. Under hard thresholding the decision of keeping or deleting a predictor from the subset is based on a pre-test. Another hard thresholding method is done by Bai and Ng (2007) who made a modification on the method of Bair et al. (2006). Both the techniques are applied on the original data before constructing factors. In their paper, Boivin and Ng (2006) also use a hard thresholding technique to narrow down the amount of predictors.

Bai and Ng (2007) also considered three soft thresholding methods. Under soft thresholding, the predictors are ordered by the particular technique and only the top ranked variables are kept.

2 Preliminaries

Over the years the diffusion index forecasts of Stock and Watson (2002) received deserved attention. The forecasts made by the DI framework have shown to result in a reduction in mean-squared forecast errors (MSFE) compared to other forecast methods. For example, the MSFE of inflation forecasts made by the DI method were smaller than MSFEs of forecasts made by using real variables such as the industrial production as shown by Boivin and Ng (2005).

Within the DI framework several improvements were made by combining forecasts made by a number of different factor models. Weighting schemes such as restricted least squares have shown to improve the predictive power of the DI framework, as can be seen in van Dongen et al. (2013).

One could wonder how much data is needed for the DI forecasts. As shown in Boivin and Ng (2006), a increase of data with predictors that hold little information about the factors does not always lead to an improvement in the forecasts. One method to improve principal component estimates is by using weighted principal components with GLS instead of OLS based principal components. Results in Stock and Watson (2004a) have shown that forecasts made by weighted principal components indeed outperform the 'ordinary' principal components.

A fine way to determine the number of predictors is by using hard and soft thresholdig techniques. The hard thresholding method used by Bair et al. (2006) is called 'supervised principal components' and was first described in a biological setting in Bair and Tibshirani (2004). The technique used by Bai and Ng (2007) is based on this method. Another hard thresholding method, which is based on the correlation coefficients of the errors, is done by Boivin and Ng (2006). Noteworthy is that the target variable was not taken into account in this procedure.

At every bootstrap sample of the bagging algorithm predictors are selected according to a pre-test. The term bagging is short for bootstrap aggregation (Breiman, 1996) and is a method that could reduce the MSFE of models selected by pre-tests. Research of Buhlmann and Yu (2002) shows that bagging has the ability to decrease the asymptotic prediction mean-squared error of regressions with a single regressor when using i.d.d. data. Their results show that bagging does not always resolve the instability of the pre-testing rule but has the potential to do so. Other research done by Inoue and Kilian (2005) show that baggging has the ability to improve factor model forecasts for the U.S. inflation with 26 predictors. However, it has never been tried on a larger amount of predictors.

A classical soft thresholding technique for data containing many correlated predictors is ridge regression (Hoerl and Kennard, 1970). In the research of Donoho and Johnstone (1994), many soft thresholdig methods yield optimal results and LASSO asymptotically comes close to being the ideal subset selector in terms of its oracle function. However, they and many other researchers assume an i.d.d. data. Many soft thresholding procedures are, as shown in Efron et al. (2004), special cases of the least angle regression algorithm. Noteworthy is that LASSO and LARS, according to results in Efron et al. (2004), choose similar subsets.

3 Subset selection

3.1 Factor model

Assume one has a large set of data of N predictors $X_t = (X_{1,t}, ...X_{N,t})$ for $t = 1, ., T$ and wants to predict a certain variable, y^h_{t+h}. Let W_t be a matrix containing a vector of ones and lags of y_t and consider the model:

$$y_{t+h} = \alpha^{'} W_t + \beta \hat{f}_t + \epsilon_{t+h} \tag{1}$$

$$X_t = \Lambda F_t + \varepsilon_t \tag{2}$$

With $F_t = (f_{1,t}, f_{2,t}, ..., f_{N,t})$ containing all the common factors. Then let $\bar{r} \leq N$ be the optimal number of common factors such that $f_t = (f_{1,t}, f_{2,t}, ...f_{\bar{r},t})$. Forecasts made by the above model are then given by:

$$\hat{y}_{t+h} = \hat{\alpha}^{\prime}_h W_t + \hat{\beta}_h \hat{f}_t \tag{3}$$

The optimal number of common factors, $\bar{r}$, can be estimated by an information criterion such as the bayesian information criterion (BIC).

3.2 Hard thresholding

The first method used to select relevant predictors is simply based on a statistical test. Whether to include or exclude a variable is based on a pre-test without considering other predictors.

The hard thresholding method used in this paper is the same as used by Bai and Ng (2007). Their method comes close to the implementation of hard thresholding by Bair et al. (2006), who state that the principal components of a subset could dominate principal components of a large group of regressors. However, the method suggested by Bair et al. (2006) is used on a very different kind of dataset from an other field (genes) and therefore the adjustment on their implementation as done by Bai and Ng (2007) also holds for the dataset used in this paper. The reason behind this adjustment is that because of the dependent nature of my data the relation between the targeted variable and the regressors needs to be controlled by lags of the dependent variable.

Consider the following model;

$$y_{t+h} = \alpha^{'} W_t + \omega^{'} X_{i,t} + \varepsilon_{t+h} \text{ for } i = 1, ..., N \tag{4}$$

with W_t as a matrix containing a vector of ones and lags of y_t and $X_{i,t}$ as possible regressor. Let Ω be the vector containing the parameters. Let t_i be the t-statistic for the null that ω_i is zero and ω_i is the i-th element of the vector Ω. Then let X^*_t be the set of predictors of which the absolute t value exceeded the critical value, c, on which, by principal component analyses, the factors, $\hat{F}_t$, can be constructed. From this set of factors a subset of factors, $\hat{f}_t$, can be estimated by using the BIC. These factors can be used to construct a h period ahead forecast as done in equation (3). As shown by van Dongen et al. (2013), for the target variables: IP, PI and EMP, lags do not enhance forecasts therefore W_t does not have to contain lags of the specific targeted variable but only a vector of ones. This reduces the above equation into:

$$\hat{y}_{t+h} = \hat{\alpha} + \hat{\beta}\hat{f}_t \tag{5}$$

When constructing the t-values I use standard errors that allow for serial correlation and conditional heteroskedasticity. The errors are the nonparametric robust standard errors of the HAC estimator (Newey and West, 1987).

3.3 Bagging

The idea behind bagging is to first include all the predictors in the model and then create a large number of bootstrap resamples. The third step is to apply the pre-test rule on each of the resamples followed by creating forecasts with these pre-test models after which a bagging forecast can be constructed by averaging over the pre-test models. The bagging forecasts removes the instability of the pre-test rule.

Consider the unrestricted model;

$$y_{t+h} = \beta^{'} f_t + \varepsilon_{t+h} \tag{6}$$

with f_t as vector containing the M largest factors formed by principal component analysis and β is the vector containing the parameters. Let t_j be the t-statistic for the null that β_j is zero and β_j is the j-th element of the vector β. Then let f_t^* be the set containing the factors for which the related β_j had a t-statistic such that: $|t_j| > c$. This gives us the pre-test model;

$$y_{t+h} = \gamma^{'} f_t^* + \epsilon_{t+h} \tag{7}$$

The above method focuses on the factors rather then first selecting regressors and apply PCA after the shrinkage as the following procedure does. Consider the unrestricted model;

$$y_{t+h} = \beta^{'} X_t + \epsilon_{t+h} \tag{8}$$

With X_t containing the full set of regressors and let β be the vector containing the coefficients. Let t_j be the t-statistic for the null that β_j is zero and β_j is the j-th element of the vector β. Then let X_t^* be the subset containing the predictors for which the related β_j had a t-statistic such that: $|t_j| > c$. This subset of predictors is submitted to PCA to estimate factors F_t. By using the bayesian information criterium a subset $f_t \subset F_t$ is extracted. This gives us the pre-test model;

$$y_{t+h} = \gamma^{'} f_t^* + \varepsilon_{t+h} \tag{9}$$

The bagging forecast is then constructed by averaging the pre-test forecasts across the iterations of the bagging algorithm. In practice a number of 100 bootstrap samples is sufficient and just as the hard threshold method the nonparametric robust standard errors of the HAC estimator (Newey and West, 1987) are used to construct the t-values.

3.4 Soft thresholding

The hard thresholding method mentioned above might be sensitive to small changes in the data due to the crudeness of the decision rule. The bagging algorithm might resolve this, but then again both methods have a drawback when it comes to selecting predictors. Hard thresholding does not take other variables into account when selecting a predictor, which might result in a subset with 'similar' variables. Predictor selection at the bootstrap re-samples might happen too abrupt as all the insignificant variables are removed at once, even though they still might hold valuable information for other predictors. For these reasons several other methods are mentioned by Bai and Ng (2007) as these techniques perform subset selection and shrinkage simultaneously. All of the methods are also implementations of the Least Angle Regressions Algorithm also known as LARS as shown by Efron et al. (2004).

3.4.1 Lasso

The first of the three is the least absolute shrinkage selection operator also known as LASSO estimator of Tibshirani (1996). Let $\text{RSS}(\alpha, \beta)$ be the sum of squared residuals from a regression of y_{t+h}^h on the full set of regressors. Then the LASSO estimator is the solution to:

$$\min_{\beta,\alpha} \quad RSS + \lambda \sum_{j=1}^{N} |\beta_j| \tag{10}$$

With $0 \leq \lambda < \infty$ as parameter that shrinks the least squares estimates of β_j or sets them to zero. This is close to the ridge estimator which is the solution to:

$$\min_{\beta,\alpha} \quad RSS + \lambda \sum_{j=1}^{N} \beta_j^2 \tag{11}$$

However, an advantage of the LASSO estimator over the ridge estimator is that thanks to the L_1 penalty, $\sum_{j=1}^{N} |\beta_j|$, instead of the L_2 penalty, $\sum_{j=1}^{N} \beta_j^2$, some of the β_j can be zero.

As I focus on finding the best subset to enhance factor model forecasts little attention is paid to the LASSO estimators. Solely the selection of predictors is of interest. Let X_t^* be a subset of X_t containing the predictors of which the LASSO estimators are non-zero. From the set X_t^* factors can be estimated via principal component analysis. Again the BIC will be used to determine the optimal number of common factors. These factors will be used to construct a h-step ahead forecast.

3.4.2 The Elastic Net

Even though the LASSO estimator is an improvement over the ridge estimator as it sets coefficients to zero, LASSO still has a few disadvantages. First of all, when there is a high correlation between the regressors, LASSO seems to be dominated by the ridge. Secondly as shown by Zou and Hastie (2005), if $N > T$ LASSO can at most select a T number of predictors. Finally, in case there is a group of variables with high pairwise coefficients, LASSO is inclined to choose just one of these variables and has no concerns about which one.

As result of these imperfections a combination of ridge and LASSO was constructed by Zou and Hastie (2005) called: 'the elastic net' estimator. The EN estimator is the solution to:

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

A distinctive feature of EN is that it can be reformulated as a LASSO problem as shown by Bai and Ng (2007).

One way to set λ_2 for the second penalty is by assigning a predetermined value to it and keep it unchanged for the full forecast period. The problem with a predetermined and unaltered λ_2 is that both assumptions might not hold and therefore influence the forecasts in a bad way. A grid search for an optimal λ_2 solves the 'predetermined' problem and doing so at every t in the forecast period also allows λ_2 to change over time.

Even though a grid search might enhance the performance of the EN, it has a major drawback as it suffers from the curse of dimensionality which makes it an exhaustive method. As random search (Rastrigin, 1963) suffers less from the curse of dimensionality it is a good alternative for the grid search method. The random search method works as follows:

1. Start with a random λ_2 in the search space
2. Until the stopping criterion is met:
   (a) Chose a random λ_2^* surrounding λ_2 given a specific radius.
   (b) If $f(\lambda_2^*) < f(\lambda_2)$ then $\lambda_2 = \lambda_2^*$
3. Last found λ_2 is the best found.

The stopping criterion could be anything like a number of iterations or when λ_2^* reaches a certain value. The stopping criterium used in this paper is that if no better candidate has been found for 5 times on a row, the last found λ_2 is the best found.
Just as with the LASSO estimates the main interest is the selection of predictors.

3.4.3 Least Angle Regression

While hard thresholding does not take the correlation of the set of predictors in account, forward selection does. Forward selection includes the (k+1)-th predictor in the set if it has the maximum correlation with the residual vector from the k-th step. However, forward selection is crude in the way that it leaves many predictors out if they are correlated with the ones that are included. A method to resolve this discomfort is to use forward stagewise regression, which takes smaller steps towards the final model. Efron et al. (2004) have shown that forward selection is a special case of Least Angle Regression (LARS). LARS, just as forward selection, adds 1 predictor at each step k. After k steps there are k active variables and so $N - k$ inactive variables. If one would continue until $k = N$ then there are N active variables in the set. For LARS each regressor is assumed to be standardized and the targeted variable is assumed to be centered.

Denote $\hat{\mu}_k$ as the estimate of the dependent variable with k number of predictors and $\hat{c} = X'(y - \hat{\mu}_k)$. Let K be the set of predictors with the largest absolute correlations, $\hat{\mu}$ as current estimate of y and $\hat{\mu}_0 = 0$ then;

$$\hat{C} = \max_{j \in K} |\hat{c}_j|, \; X_k = \text{sign}(\hat{c}_j) x_j \text{ for } j \in K$$

With X_k as matrix corresponding to K. Then define 1_k as vector of ones size K, $G_k = X_k' X_k$ and

$$A_k = (1_k' G_k^{-1} 1_k)^{-\frac{1}{2}}, \; \upsilon_K = A_K G_K^{-1} 1_K, \; \mu_K = X_K \upsilon_K, \; a_K = X' \mu_K$$

With the above information μ can be updated by the following relationship;

$$\hat{\mu}^* = \hat{\mu} + \hat{\tau}\mu_K, \text{ with } \hat{\tau} = \overset{+}{\min_{j \in K^c}} \left(\frac{\hat{C} - \hat{c}_j}{A_K - a_j}, \frac{\hat{C} - \hat{c}_j}{A_K + a_j} \right) \tag{13}$$

There are quite a few advantages of LARS. First of all, LARS gives a ranking of the predictors with the knowledge that there still are other regressors. This is not the case with hard thresholding and bagging. The second advantage is that LARS avoids strongly correlated predictors, since if one of the predictors is in the active set the new residuals will have a smaller correlation with the other correlated predictors that are still in the inactive set. Lastly the LARS algorithm is not as eager as forward regression to exploit a good direction.

Unlike LASSO and EN, LARS does not have a tuning parameter λ to determine the set of significant predictors. Instead the size of the active set can be chosen by a certain value of k. Because LARS ranks the predictors, k number of predictors can be used to create factors and construct a h-step ahead forecast and $N - k$ predictors are set to 0. To adjudicate the optimal amount of regressors, k^*, a information criterium can be used such as the BIC. Another way to determine the optimal k^* is the use of a predetermined amount of predictors in the active set. The reason behind this concept is that the BIC tends to malfunction when N and T grow large as shown by Bai and Ng (2002).

4 Data

This paper will use the same data as used in van Dongen et al. (2013). The data set originally contained a large number (132) of predictors of which 6 were removed due to incompleteness of the variables. This brings the data set down to a total of 126 predictors. After testing for unit roots a large part of the data was transformed. Also predictors with an exponential form were transformed. The regressors were transformed according to the transformation codes given by Stock and Watson (2002).

After the above modifications the data set ranged from 1960:3 until 2009:9 and therefore includes 595 monthly observations for each of the 126 variables. As done by Stock and Watson (2002) observations exceeding 10 times the interquartile range from he median were considered outliers and removed. The last made modification was standardizing the data. A complete listing of the series can be found in appendix A.2.

For direct h-step ahead forecasts the target variables were transformed to a h-th difference index. The three variables, personal income (PI), industrial production index (IP), non-agricultural employment (EMP), were transformed into first difference logarithms as done in (14).

$$y_{t+h} = \frac{1200}{h}\ln(\frac{IP_{t+h}}{IP_t}),\ y_t = 1200\ln(\frac{IP_t}{IP_{t-1}}) \tag{14}$$

5 Results

The in-sample period begins at 1960:3 and ends at 1989:12 so that the first estimation made is based on 358-h data points point. Instead of fixing the set of optimal predictors the best subset according to either soft or hard thresholding rules is reselected at every t and new factors are estimated.

Comparing a forecasting method to the benchmark method is done by evaluating the relative mean-squared prediction error (RMSE) which is calculated as in (15).

$$\text{RMSE(method)} = \frac{\text{MSFE(method)}}{\text{MSFE(benchmark)}} \tag{15}$$

In case the RMSE of a method is less then 1 it means that it is superior to the benchmark model.

The benchmark model is a factor model constructed by using all the predictors and the optimal amount of factors was selected by using the BIC, as in van Dongen et al. (2013). This means the benchmark model is a factor model with 3 factors as shown in equation (16).

$$\hat{y}_{t+h} = 0\hat{\alpha} + \hat{\beta}\hat{f}_t \tag{16}$$

with $\hat{f}_t$ containing the first 3 factors.

5.1 Predictor selection

As at each t the predictors are reevaluated, a record exists of relevant predictors per iteration. The fraction of predictors included averaged over t can be found in Table 1 and 2 in the column 'total'. The frequency that a predictor is included (when it is included) is also displayed in both tables. The frequency a predictors is included is found by averaging the total times a predictor is in the active set over t. The tables are divided in 4 equal sections that stand for the fraction of predictors that is selected in different frequencies. For example: for the one month ahead forecast for variable PI, 7% of the predictors was selected with a frequency of 0.25 to 0.5 times and on average 57% of the full set of predictors were selected.

The results are not shown for bagging because as it averages over 100 re-samples it would take a new dimension to keep track of all the information. Also the results of LARS with k = [10,20,30] are not shown as the active set depends on the predetermined k.

Table 1: *Contains the fraction of predictors that is included averaged over t and the frequency a predictor is included (when it is included) divided in 4 equal sections. It shows the results for the variables: personal income (PI), industrial production index (IP), non-agricultural employment (EMP). This table shows the results for hard thresholding with thresholds of 1%, 5%, and 10%*

variabele	**horizon**	**Hard thresholding 1%**					**Hard thresholding 5%**					**Hard thresholding 10%**				
		total	(0-0.25]	[0.25-0.5]	[0.5-0.75]	[0.75-1]	total	(0-0.25]	[0.25-0.5]	[0.5-0.75]	[0.75-1]	total	(0-0.25]	[0.25-0.5]	[0.5-0.75]	[0.75-1]
	1	0,57	0,17	0,07	0,05	0,28	0,73	0,16	0,02	0,06	0,49	0,83	0,10	0,05	0,06	0,62
PI	6	0,54	0,09	0,06	0,09	0,30	0,68	0,11	0,02	0,03	0,52	0,78	0,06	0,08	0,05	0,59
	12	0,45	0,10	0,07	0,04	0,25	0,57	0,07	0,08	0,05	0,37	0,69	0,10	0,05	0,06	0,48
	24	0,32	0,11	0,06	0,03	0,12	0,52	0,14	0,05	0,04	0,29	0,65	0,17	0,10	0,03	0,36
	1	0,42	0,10	0,03	0,03	0,26	0,66	0,14	0,06	0,06	0,40	0,75	0,12	0,06	0,07	0,50
IP	6	0,33	0,05	0,02	0,00	0,26	0,52	0,09	0,03	0,02	0,39	0,61	0,06	0,02	0,05	0,48
	12	0,33	0,06	0,04	0,00	0,22	0,54	0,06	0,04	0,03	0,41	0,62	0,06	0,03	0,06	0,47
	24	0,40	0,02	0,03	0,05	0,29	0,58	0,06	0,08	0,01	0,43	0,65	0,04	0,05	0,06	0,51
	1	0,56	0,07	0,09	0,02	0,39	0,68	0,06	0,04	0,08	0,51	0,80	0,10	0,06	0,05	0,60
EMP	6	0,57	0,01	0,02	0,02	0,53	0,69	0,06	0,02	0,02	0,59	0,82	0,13	0,06	0,03	0,60
	12	0,57	0,02	0,05	0,03	0,48	0,70	0,07	0,01	0,03	0,59	0,75	0,06	0,02	0,06	0,62
	24	0,52	0,09	0,06	0,04	0,34	0,69	0,09	0,05	0,04	0,52	0,75	0,08	0,03	0,02	0,61

Results in table 1 point out that as the threshold rule gets more strict the number of relevant predictors decreases as the total amount of predictors included is lower at $\alpha = 1\%$ than at $\alpha = 10\%$ which is what one would expect. Something else notable is that as h increases the total amount of predictors decreases for PI and IP which implies that there is a shortage of variables with predictive power for long periods. Another observation is that a lot predictors lay in the last quarter which means many variables are used at every iteration.

Table 2: *Contains the fraction of predictors that is included averaged over t and the frequency a predictor is included (when it is included) divided in 4 equal sections. It shows the results for the variables: personal income (PI), industrial production index (IP), non-agricultural employment (EMP). This table shows the results for soft thresholding where the active set of LARS was determined via the BIC and elastic net has a predetermined and unaltered $\lambda_2 = 0.25$.*

variabele	**horizon**	**LARS BIC**					**LASSO**					**Elastic Net** $\lambda = 0.25$				
		total	(0-0.25]	[0.25-0.5]	[0.5-0.75]	[0.75-1]	total	(0-0.25]	[0.25-0.5]	[0.5-0.75]	[0.75-1]	total	(0-0.25]	[0.25-0.5]	[0.5-0.75]	[0.75-1]
	1	0,26	0,19	0,02	0,02	0,03	0,22	0,16	0,02	0,03	0,02	0,09	0,06	0,01	0,01	0,01
PI	6	0,21	0,10	0,02	0,02	0,07	0,21	0,10	0,02	0,01	0,07	0,17	0,06	0,02	0,01	0,07
	12	0,29	0,19	0,02	0,01	0,06	0,27	0,17	0,02	0,02	0,05	0,20	0,08	0,05	0,00	0,07
	24	0,29	0,17	0,06	0,02	0,03	0,22	0,12	0,06	0,01	0,03	0,10	0,06	0,01	0,00	0,03
	1	0,29	0,11	0,04	0,02	0,13	0,25	0,09	0,02	0,02	0,12	0,23	0,04	0,02	0,02	0,14
IP	6	0,25	0,06	0,03	0,02	0,14	0,25	0,09	0,02	0,05	0,10	0,21	0,04	0,02	0,01	0,14
	12	0,25	0,12	0,05	0,02	0,06	0,21	0,10	0,03	0,02	0,06	0,13	0,02	0,02	0,02	0,07
	24	0,33	0,11	0,07	0,09	0,06	0,29	0,15	0,06	0,02	0,05	0,15	0,04	0,03	0,02	0,06
	1	0,22	0,11	0,03	0,02	0,06	0,22	0,11	0,03	0,02	0,06	0,14	0,06	0,02	0,02	0,04
EMP	6	0,52	0,21	0,05	0,06	0,20	0,45	0,16	0,06	0,06	0,17	0,30	0,08	0,06	0,01	0,16
	12	0,46	0,20	0,06	0,05	0,15	0,45	0,21	0,05	0,04	0,15	0,21	0,06	0,02	0,01	0,12
	24	0,52	0,25	0,06	0,06	0,13	0,45	0,23	0,06	0,06	0,10	0,12	0,02	0,04	0,01	0,06

A big difference between soft and hard thresholding is shown in table 2 as the soft thresholding techniques include less predictors in the active set. A reason could be that there are many predictors that are 'similar'. As hard thresholding did not take other predictors in count these could have been included. Noteworthy is that LARS and LASSO use similar number of predictors and Elastic net is in all cases for every period more strict on including a variable. For the variable EMP both LASSO and LARS select a larger number of predictors compared to the other two target variables. The number of predictors is larger for every horizon but h=1, meaning that for EMP there are fewer predictors that hold predictive power for the short period.

In table 5 in Appendix A.1 a list can be found containing the top 10 ranked predictors used by LARS and LASSO. It is interesting to notice that for different periods different predictors are used for forecasting. Especially the 2 year step ahead forecasts makes use of other predictors then other period forecasts. Table 5 also shows that variables selected by LASSO are quite similar to the ones selected by LARS. Previous research done by Efron et al. (2004) indicated this phenomenon and it seems to also hold now.

5.2 Forecasting performance

Now we know that different thresholding rules select different subsets, but does it improve the forecast performance of a model? In table 3 the RSMEs of hard thresholding and bagging are displayed. Bagging was done in 2 different fashions. The results of bagging applied on the M largest factors with $M = 30$ can be found under 'bagging on factors' and the results of bagging on the predictors and applying PCA after variable selection on the bootstrap re-samples can be found under 'Bagging on X'.

A first glance on table 3 suggests that the 3 methods are superior to the benchmark method, as many RMSEs are below 1. If one takes a closer look, then the improvements of the methods are larger when the horizon is larger. In fact, when the horizon is equal to 1 month it is hard to outperform the benchmark model. The most improvement of all the three methods is gained on the variable EMP where bagging on factors method has the smallest RMSE. A keen thing to notice is that as $\alpha = 0.5$ the RMSE is at his best. Perhaps when $\alpha = 1\%$ the tresholding rule declines too many predictors and when α is set to 10% too many predictors are included which makes the forecasts errors more noisy.

As for the difference between the methods, there is not much difference between the RMSEs but it seems that bagging tends to do better then hard thresholding. This means that bagging does decrease the instability of the thresholding rule. Overall bagging on factors scores lower RMSEs then bagging on the predictors.

Table 3: *Contains the relative mean squared prediction error per technique for the variables: personal income (PI), industrial production index (IP), non-agricultural employment (EMP). This table shows the results for hard thresholding, bagging on factors and bagging on X with thresholds of 1%, 5%, and 10%. The 3 factor benchmark model is included for comparison.*

Variable	Horizon	**Hard thresholding**			**3 factor**	**Bagging on factors**			**Bagging on X**		
		$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$		$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
	1	1,02	1,02	1,01	**3095,67**	1	1	1,01	1,04	1,02	1,02
PI	6	0,97	0,98	0,99	**13322,51**	0,98	0,97	0,96	0,99	1	1,01
	12	1,02	0,99	0,94	**42812,48**	0,91	0,89	0,88	0,95	0,95	0,95
	24	1,02	0,98	0,94	**130227,90**	0,96	0,96	0,96	0,88	0,89	0,9
	1	0,97	1,02	0,99	**0,35**	1	1,01	1,02	0,95	0,95	0,95
IP	6	0,99	0,95	1,02	**4,57**	1	0,99	0,99	1,07	1,04	1,03
	12	1,02	0,97	0,97	**17,88**	0,93	0,91	0,9	0,94	0,95	0,96
	24	0,96	0,97	0,98	**45,48**	0,96	0,96	0,96	0,93	0,9	0,89
	1	0,98	1,00	0,99	**110910,47**	0,95	0,93	0,94	0,98	0,99	0,98
EMP	6	0,91	0,91	0,93	**1068959,58**	0,82	0,81	0,82	1,04	1,02	1,01
	12	0,93	0,91	0,96	**4488023,09**	0,77	0,75	0,73	0,96	0,92	0,9
	24	0,93	0,76	0,82	**14206806,89**	0,74	0,72	0,7	0,84	0,85	0,84

Table 4 shows the results of the soft thresholding techniques. As LARS ranks the predictors a subset can be extracted and used for factor analysis. The size of the subset depends on k = [10 20 30 k^*] where k^* is the number of predictors selected by the BIC. To set the second penalty of elastic net the predetermined $\lambda_2 = [1.5\ 0.5\ 0.25]$ is chosen. Besides a predetermined and unaltered λ_2, a random search was conducted at every t over the search space $\lambda_s = [0.01, 0.02, 0.03, ..., 0.3]$. Results of elastic net with the random search algorithm can be found under $\lambda_2 = \lambda_2^*$.

Table 4: *Contains the relative mean squared prediction error per technique for the variables: personal income (PI), industrial production index (IP), non-agricultural employment (EMP). This table shows the results for Lars with* $k = [10, 20, 30, k^*]$ *LASSO and Elastic net with* $\lambda_2 = [1.5, 0.5, 0.25, \lambda_2^*]$. *The 3 factor benchmark model is included for comparison*

Variable	Horizon	**LARS**				**3 factor**	**LASSO**	**Elastic net**			
		k = 10	k=20	k=30	k=k^*			$\lambda_2 = 1.5$	$\lambda_2 = 0.5$	$\lambda_2 = 0.25$	$\lambda_2 = \lambda_2^*$
	1	1,18	1,09	1,07	1,19	**3095,67**	1,18	1,01	1,02	1,04	1,13
PI	6	1,01	0,87	0,92	0,9	**13322,51**	0,90	1,11	1,06	1,03	0,93
	12	0,97	0,91	0,92	0,91	**42812,48**	0,92	1,17	1,02	0,97	0,98
	24	1,06	1,05	1,04	1,07	**130227,90**	1,06	1,12	1,06	1,06	1,04
	1	1,01	1,05	1	1	**0,35**	1,03	1,02	1,01	1,00	1,02
IP	6	1,20	0,97	0,98	0,97	**4,57**	1,04	1,30	1,09	0,98	1,02
	12	1,07	1,06	0,99	1,05	**17,88**	1,05	1,34	1,18	1,18	1,09
	24	1,17	1,16	1,15	1,19	**45,48**	1,16	1,22	1,25	1,21	1,17
	1	0,94	0,98	1,04	0,94	**110910,47**	0,95	0,98	0,96	0,93	0,95
EMP	6	1,01	0,91	0,83	0,81	**1068959,58**	0,79	1,18	1,00	0,86	0,81
	12	0,89	0,76	0,77	0,77	**4488023,09**	0,77	1,04	0,91	0,87	0,79
	24	0,92	0,72	0,73	0,73	**14206806,89**	0,72	1,12	1,04	0,98	0.83

Just as hard thresholding and bagging the soft thresholding methods struggle to outperform the benchmark model at the shortest horizon but perform better when the horizon is equal to 6 or 12 months. At the largest horizon the methods are again outperformed by the benchmark model except for the variable EMP. There hardly is any improvement at the variable IP.

For LARS the forecasts get better as the predetermined number of predictors for PCA grows larger. The number of predictors as calculated by the BIC keeps up with this trend as it fluctuates between 15-25 predictors at every iteration.

Noteworthy is that the RMSEs of LARS with $k = k^*$ and LASSO are alike. This comes due the fact that both methods select similar subsets as shown in appendix A.1 and by Efron et al. (2004).

The results of elastic net tend to do better as the tuning parameter for the second penalty shrinks. As $\lambda_2 = 1.5$ the method outperforms the benchmark model only once and this number grows to 5 when $\lambda_2 = 0.25$ so perhaps an even smaller λ_2 could enhance the forecasts even more. That this is true follows from the random search algorithm as the mean of λ is under 0.1 for every forecast period and variable. The forecasts made by EN in combination with random search most of the times perform better than with a predetermined and unaltered λ_2. With a set λ_2, elastic net is out performed by the other two soft thresholding methods but combined with the random search algorithm, EN seems to perform equally well as LARS and LASSO.

6 Conclusion

Given the results, selecting relevant predictors before applying principal component analysis yields improved forecasts. Both soft and hard thresholding show signs of improvement. The degree of improvement depends on the target variable and the target variable with the most improvement compared to the benchmark model is non-agricultural employment.

Bagging tends to soften the crudeness of the pre-test rule and results show that this, most of the times, yields to better results than hard thresholding. If one has the option of applying bagging on the full set of predictors or on the first M factors one should use the latter as it is inclined to give better results and saves computing time.

Even though bagging outperforms hard thresholding, on the horizons 6 and 12 soft thresholding provides better results for Personal Income and non-agricultural employment. Hardly any improvement is made in forecasting the industrial production index for any of the techniques. Of the three soft thresholding procedures LARS and LASSO operate the best. That there is little difference in the relative mean-squared errors of LARS and LASSO comes as no surprise as both select close to identical subsets of predictors.

Techniques demonstrated in this paper came to good results for two target variables: personal income and non-agricultural employment. As shown by Bai and Ng (2007) 4 of the 5 methods used also improve inflation forecasts. Further research could use the techniques on other macro-economic variables, or even different fields of expertise, to enhance forecasts.

Further research into bagging could provide more insight in the technique when used on large data sets. Also a change in the number of factors M used in the bootstrap re-samples could alter the results. Noteworthy is that when M increases the computing time also increases such that when M = N creating one forecast could take up to 15 minutes, depending on the available computer power. For both hard thresholding and the bootstrap re-samples the pre-test was based on a t-statistic. It was never stated that this was the best pre-test but it tends to work well. Further research could alter pre-test rules and this might resolve in improved forecasts.

There are several other soft thresholding methods and this paper only covers 3. Other procedures such as adaptive LASSO by Zou (2006) might increase the amount of improvement on the original model. But perhaps small modifications yield better results. For example the use of different methods to estimate the second tuning variable in elastic net such as the Adaptive Step Size Random Search algorithm of Schumer and Steiglitz (1968) or Optimized Relative Step Size Random Search algorithm of Schrack and Choit (1976).

A Appendix

A.1 Predictor selection

Table 5: *The top 10 selected predictors per technique and variable for the horizons: 1, 6, 12 and 24*

		LARS				Lasso			
		h = 1	h= 6	h = 12	h = 24	h = 1	h = 6	h = 12	h = 24
	1	'Emp: mfg'	'Emp: gds prod'	'Help wanted indx'	'M3'	'Emp: mfg'	'Emp: gds prod'	'Help wanted indx'	'M3'
	2	'Aaa bond'	'10 yr T-bond'	'M3'	'1 yr T-bond'	'PI less transfers'	'Aaa bond'	'M3'	'1 yr T-bond'
	3	'PI less transfers'	'Aaa bond'	'1 yr T-bond'	'10 yr T-bond'	'10 yr T-bond'	'10 yr T-bond'	'1 yr T-bond'	'S&P PE ratio'
	4	'10 yr T-bond'	'Help wanted/emp'	'10 yr T-bond'	'S&P PE ratio'	'Aaa bond'	'Help wanted/emp'	'10 yr T-bond'	'10 yr T-bond'
PI	5	'Aaa-FF spread'	'M3'	'Aaa bond'	'5 yr T-bond'	'Aaa-FF spread'	'M3'	'Aaa bond'	'5 yr T-bond'
	6	'PI'	'Help wanted indx'	'Emp: nondbles'	'5 yr-FF spread'	'PI'	'Help wanted indx'	'Emp: nondbles'	'5 yr-FF spread'
	7	'IP: cons dble'	'Emp: nondbles'	'Emp: wholesale'	'Unf orders: dble'	'IP: cons dble'	'Emp: nondbles'	'Emp: mfg'	'6 mo T-bill'
	8	'Emp: nondbles'	'Emp: wholesale'	'Emp: mfg'	'NAPM vendor del'	'Emp: nondbles'	'Emp: mfg'	'Emp: wholesale'	'IP: nondble matls'
	9	'U: mean duration'	'Emp: mfg'	'IP: total'	'6 mo T-bill'	'U: mean duration'	'Emp: wholesale'	'IP: total'	'Emp: wholesale'
	10	'IP: matls'	'Emp: Govt'	'Emp: retail'	'3 mo T-bill'	'1 yr T-bond'	'Emp: Govt'	'Emp: retail'	'NAPM vendor del'
	1	'Help wanted/emp'	'IP: nondble matls'	'Help wanted/emp'	'Unf orders: dble'	'Help wanted/emp'	'IP: nondble matls'	'Help wanted/emp'	'Unf orders: dble'
	2	'U 41760 wks'	'Help wanted indx'	'M3'	'M3'	'UI claims'	'Help wanted indx'	'M3'	'M3'
	3	'UI claims'	'Help wanted/emp'	'C&I loans'	'6 mo T-bill'	'Emp: nondbles'	'Help wanted/emp'	'C&I loans'	'6 mo T-bill'
	4	'Emp: mfg'	'UI claims'	'10 yr T-bond'	'1 yr T-bond'	'Unf orders: dble'	'UI claims'	'10 yr T-bond'	'1 yr T-bond'
IP	5	'Emp: nondbles'	'Emp: nondbles'	'Aaa bond'	'10 yr T-bond'	'CPI-U: ex med'	'Emp: nondbles'	'Aaa bond'	'Baa bond'
	6	'Unf orders: dble'	'Unf orders: dble'	'IP: nondble matls'	'Baa bond'	'3 mo T-bill'	'Unf orders: dble'	'IP: nondble matls'	'10 yr T-bond'
	7	'3 mo T-bill'	'M3'	'CP-FF spread'	'U 27+ wks'	'U 41760 wks'	'M3'	'CP-FF spread'	'Emp: dble gds'
	8	'5 yr T-bond'	'5 yr T-bond'	'Emp: nondbles'	'Emp: dble gds'	'IP: matls'	'5 yr T-bond'	'Help wanted indx'	'Emp: wholesale'
	9	'10 yr T-bond'	'10 yr T-bond'	'Overtime: mfg'	'Emp: wholesale'	'10 yr T-bond'	'PCE defl: dlbes'	'Emp: nondbles'	'5 yr T-bond'
	10	'Aaa bond'	'Aaa bond'	'Help wanted indx'	'S&P: indust'	'PCE defl: dlbes'	'CP-FF spread'	'Overtime: mfg'	'U 27+ wks'
	1	'Help wanted indx'	'Help wanted indx'	'Help wanted indx'	'Unf orders: dble'	'Help wanted indx'	'Help wanted indx'	'Help wanted indx'	'M3'
	2	'Emp: mfg'	'Emp CPS total'	'Help wanted/emp'	'M3'	'Emp: mfg'	'Emp CPS total'	'Help wanted/emp'	'Baa bond'
	3	'Emp: nondbles'	'U 15+ wks'	'U 41760 wks'	'3 mo T-bill'	'Emp: nondbles'	'U 15+ wks'	'Emp: mfg'	'10 yr T-bond'
	4	'Emp: wholesale'	'UI claims'	'Emp: mfg'	'6 mo T-bill'	'Emp: wholesale'	'UI claims'	'Emp: nondbles'	'Unf orders: dble'
EMP	5	'Emp CPS total'	'Emp: mfg'	'Emp: nondbles'	'10 yr T-bond'	'U 41760 wks'	'Emp: mfg'	'Emp: retail'	'3 mo T-bill'
	6	'U 41760 wks'	'Emp: nondbles'	'Emp: retail'	'Baa bond'	'Emp CPS total'	'Emp: nondbles'	'M3'	'6 mo T-bill'
	7	'Consumption'	'Emp: TTU'	'M3'	'Emp: retail'	'Consumption'	'Emp: TTU'	'6 mo T-bill'	'Emp: retail'
	8	'Emp: TTU'	'Emp: wholesale'	'Inst cred/PI'	'IP: fuels'	'Emp: TTU'	'Emp: wholesale'	'10 yr T-bond'	'Help wanted indx'
	9	'Emp: gds prod'	'Emp: FIRE'	'6 mo T-bill'	'M2'	'Emp: gds prod'	'Emp: FIRE'	'U 41760 wks'	'IP: fuels'
	10	'IP: fuels'	'Inst cred/PI'	'10 yr T-bond'	'Help wanted indx'	'IP: fuels'	'Inst cred/PI'	'Inst cred/PI'	'M2'

A.2 Data

Table 6 contains the short name of each series, the transformation code and a short description of the series. The meaning of the transformation codes are: 1 = no transformation, 2 = first difference, 4 = logarithm, 5 = first difference of logarithms, 6 = second difference of logarithms.

Table 6: *Series list*

Name	T-code	description
'PI'	5	'Personal Income (AR, Bil. Chain 2000 $) (TCB)'
'PI less transfers'	5	'Personal Income Less Transfer Payments (AR, Bil. Chain 2000 $) (TCB)'
'Consumption'	5	'Real Consumption (AC) a0m224/gmdc (a0m224 is from TCB)'
'M&T sales'	5	'Manufacturing And Trade Sales (Mil. Chain 1996 $) (TCB)'
'Retail sales'	5	'Sales Of Retail Stores (Mil. Chain 2000 $) (TCB)'
'IP: total'	5	'Industrial Production Index - Total Index'
'IP: products'	5	'Industrial Production Index - Products, Total'
'IP: final prod'	5	'Industrial Production Index - Final Products'
'IP: cons gds'	5	'Industrial Production Index - Consumer Goods'
'IP: cons dble'	5	'Industrial Production Index - Durable Consumer Goods'
'IP: cons nondble'	5	'Industrial Production Index - Nondurable Consumer Goods'
'IP: bus eqpt'	5	'Industrial Production Index - Business Equipment'
'IP: matls'	5	'Industrial Production Index - Materials'
'IP: dble matls'	5	'Industrial Production Index - Durable Goods Materials'
'IP: nondble matls'	5	'Industrial Production Index - Nondurable Goods Materials'
'IP: mfg'	5	'Industrial Production Index - Manufacturing (Sic)'
'IP: res util'	5	'Industrial Production Index - Residential Utilities'
'IP: fuels'	5	'Industrial Production Index - Fuels'
'NAPM prodn'	1	'Napm Production Index (Percent)'
'Cap util'	2	'Capacity Utilization (Mfg) (TCB)'
'Help wanted indx'	2	'Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)'
'Help wanted/emp'	2	'Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf'
'Emp CPS total'	5	'Civilian Labor Force: Employed, Total (Thous.,Sa)'
'Emp CPS nonag'	5	'Civilian Labor Force: Employed, Nonagric.Industries (Thous.,Sa)'
'U: all'	2	'Unemployment Rate: All Workers, 16 Years & Over (%,Sa)'
'U: mean duration'	2	'Unemploy.By Duration: Average(Mean)Duration In Weeks (Sa)'
'U ¡ 5 wks'	5	'Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)'
'U 41760 wks'	5	'Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)'
'U 15+ wks'	5	'Unemploy.By Duration: Persons Unempl.15 Wks + (Thous.,Sa)'
'U 15-26 wks'	5	'Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)'
'U 27+ wks'	5	'Unemploy.By Duration: Persons Unempl.27 Wks + (Thous,Sa)'
'UI claims'	5	'Average Weekly Initial Claims, Unemploy. Insurance (Thous.) (TCB)'
'Emp: total'	5	'Employees On Nonfarm Payrolls: Total Private'
'Emp: gds prod'	5	'Employees On Nonfarm Payrolls - Goods-Producing'
'Emp: mining'	5	'Employees On Nonfarm Payrolls - Mining'
'Emp: const'	5	'Employees On Nonfarm Payrolls - Construction'
'Emp: mfg'	5	'Employees On Nonfarm Payrolls - Manufacturing'
'Emp: dble gds'	5	'Employees On Nonfarm Payrolls - Durable Goods'
'Emp: nondbles'	5	'Employees On Nonfarm Payrolls - Nondurable Goods'
'Emp: services'	5	'Employees On Nonfarm Payrolls - Service-Providing'
'Emp: TTU'	5	'Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities'
'Emp: wholesale'	5	'Employees On Nonfarm Payrolls - Wholesale Trade'
'Emp: retail'	5	'Employees On Nonfarm Payrolls - Retail Trade'
'Emp: FIRE'	5	'Employees On Nonfarm Payrolls - Financial Activities'
'Emp: Govt'	5	'Employees On Nonfarm Payrolls - Government'
'Emp-hrs nonag'	5	'Employee Hours In Nonag. Establishments (AR, Bil. Hours) (TCB)'
'Avg hrs'	1	'Avg Weekly Hrs of Prod or Nonsup Workers On Private Nonfarm Payrolls - Goods-Producing'
'Overtime: mfg'	2	'Avg Weekly Hrs of Prod or Nonsup Workers On Private Nonfarm Payrolls - Mfg Overtime Hours'
'Avg hrs: mfg'	1	'Average Weekly Hours, Mfg. (Hours) (TCB)'
'NAPM empl'	1	'Napm Employment Index (Percent)'
'Starts: nonfarm'	4	'Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-)(Thous.,Saar)'
'Starts: NE'	4	'Housing Starts:Northeast (Thous.U.)S.A.'
'Starts: MW'	4	'Housing Starts:Midwest(Thous.U.)S.A.'
'Starts: South'	4	'Housing Starts:South (Thous.U.)S.A.'
'Starts: West'	4	'Housing Starts:West (Thous.U.)S.A.'
'BP: total'	4	'Housing Authorized: Total New Priv Housing Units (Thous.,Saar)'
'BP: NE'	4	'Houses Authorized By Build. Permits:Northeast(Thou.U.)S.A'
'BP: MW'	4	'Houses Authorized By Build. Permits:Midwest(Thou.U.)S.A.'
'BP: South'	4	'Houses Authorized By Build. Permits:South(Thou.U.)S.A.'
'BP: West'	4	'Houses Authorized By Build. Permits:West(Thou.U.)S.A.'
'PMI'	1	'Purchasing Managers" Index (Sa)'
'NAPM new ordrs'	1	'Napm New Orders Index (Percent)'
'NAPM vendor del'	1	'Napm Vendor Deliveries Index (Percent)'

Name	T-code	description
'NAPM Invent'	1	'Napm Inventories Index (Percent)'
'Orders: cons gds'	5	'Mfrs" New Orders, Consumer Goods And Materials (Bil. Chain 1982 $) (TCB)'
'Orders: dble gds'	5	'Mfrs" New Orders, Durable Goods Industries (Bil. Chain 2000 $) (TCB)'
'Orders: cap gds'	5	'Mfrs" New Orders, Nondefense Capital Goods (Mil. Chain 1982 $) (TCB)'
'Unf orders: dble'	5	'Mfrs" Unfilled Orders, Durable Goods Indus. (Bil. Chain 2000 $) (TCB)'
'M&T invent'	5	'Manufacturing And Trade Inventories (Bil. Chain 2000 $) (TCB)'
'M&T invent/sales'	2	'Ratio, Mfg. And Trade Inventories To Sales (Based On Chain 2000 $) (TCB)'
'M1'	6	'Money Stock: M1(Curr,Trav.Cks,Dem Dep,Other Ck"able Dep)(Bil$,Sa)'
'M2'	6	'Money Stock:M2(M1+O"nite Rps,Euro$,G/P&B/D Mmmfs&Sav&Sm Time Dep(Bil$,Sa)'
'M3'	6	'Money Stock: M3(M2+Lg Time Dep,Term Rp"s&Inst Only Mmmfs)(Bil$,Sa)'
'M2 (real)'	5	'Money Supply - M2 In 1996 Dollars (Bci)'
'MB'	6	'Monetary Base, Adj For Reserve Requirement Changes(Mil$,Sa)'
'Reserves tot'	6	'Depository Inst Reserves:Total, Adj For Reserve Req Chgs(Mil$,Sa)'
'Reserves nonbor'	6	'Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil$,Sa)'
'C&I loans'	6	'Commercial & Industrial Loans Oustanding In 1996 Dollars (Bci)'
'C&I loans'	1	'Wkly Rp Lg Com"l Banks:Net Change Com"l & Indus Loans(Bil$,Saar)'
'Cons credit'	6	'Consumer Credit Outstanding - Nonrevolving(G19)'
'Inst cred/PI'	2	'Ratio, Consumer Installment Credit To Personal Income (Pct.) (TCB)'
'S&P 500'	5	'S&P"s Common Stock Price Index: Composite (1941-43=10)'
'S&P: indust'	5	'S&P"s Common Stock Price Index: Industrials (1941-43=10)'
'S&P div yield'	2	'S&P"s Composite Common Stock: Dividend Yield (% Per Annum)'
'S&P PE ratio'	5	'S&P"s Composite Common Stock: Price-Earnings Ratio (%,Nsa)'
'Fed Funds'	2	'Interest Rate: Federal Funds (Effective) (% Per Annum,Nsa)'
'Comm paper'	2	'Cmmercial Paper Rate (AC)'
'3 mo T-bill'	2	'Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.(% Per Ann,Nsa)'
'6 mo T-bill'	2	'Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.(% Per Ann,Nsa)'
'1 yr T-bond'	2	'Interest Rate: U.S.Treasury Const Maturities,1-Yr.(% Per Ann,Nsa)'
'5 yr T-bond'	2	'Interest Rate: U.S.Treasury Const Maturities,5-Yr.(% Per Ann,Nsa)'
'10 yr T-bond'	2	'Interest Rate: U.S.Treasury Const Maturities,10-Yr.(% Per Ann,Nsa)'
'Aaa bond'	2	'Bond Yield: Moody"s Aaa Corporate (% Per Annum)'
'Baa bond'	2	'Bond Yield: Moody"s Baa Corporate (% Per Annum)'
'CP-FF spread'	1	'cp90-fyff (AC)'
'3 mo-FF spread'	1	'fygm3-fyff (AC)'
'6 mo-FF spread'	1	'fygm6-fyff (AC)'
'1 yr-FF spread'	1	'fygt1-fyff (AC)'
'5 yr-FF spread'	1	'fygt5-fyff (AC)'
'10 yr-FF spread'	1	'fygt10-fyff (AC)'
'Aaa-FF spread'	1	'fyaaac-fyff (AC)'
'Baa-FF spread'	1	'fybaac-fyff (AC)'
'Ex rate: avg'	5	'United States;Effective Exchange Rate(Merm)(Index No.)'
'Ex rate: Switz'	5	'Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)'
'Ex rate: Japan'	5	'Foreign Exchange Rate: Japan (Yen Per U.S.$)'
'Ex rate: UK'	5	'Foreign Exchange Rate: United Kingdom (Cents Per Pound)'
'EX rate: Canada'	5	'Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)'
'PPI: fin gds'	6	'Producer Price Index: Finished Goods (82=100,Sa)'
'PPI: cons gds'	6	'Producer Price Index: Finished Consumer Goods (82=100,Sa)'
'PPI: int matls'	6	'Producer Price Index:I ntermed Mat.Supplies & Components(82=100,Sa)'
'PPI: crude matls'	6	'Producer Price Index: Crude Materials (82=100,Sa)'
'Spot market price'	6	'Spot market price index: bls & crb: all commodities(1967=100)'
'Sens matls price'	6	'Index Of Sensitive Materials Prices (1990=100)(Bci-99a)'
'NAPM com price'	1	'Napm Commodity Prices Index (Percent)'
'CPI-U: all'	6	'Cpi-U: All Items (82-84=100,Sa)'
'CPI-U: apparel'	6	'Cpi-U: Apparel & Upkeep (82-84=100,Sa)'
'CPI-U: transp'	6	'Cpi-U: Transportation (82-84=100,Sa)'
'CPI-U: medical'	6	'Cpi-U: Medical Care (82-84=100,Sa)'
'CPI-U: comm.'	6	'Cpi-U: Commodities (82-84=100,Sa)'
'CPI-U: dbles'	6	'Cpi-U: Durables (82-84=100,Sa)'
'CPI-U: services'	6	'Cpi-U: Services (82-84=100,Sa)'
'CPI-U: ex food'	6	'Cpi-U: All Items Less Food (82-84=100,Sa)'
'CPI-U: ex shelter'	6	'Cpi-U: All Items Less Shelter (82-84=100,Sa)'
'CPI-U: ex med'	6	'Cpi-U: All Items Less Medical Care (82-84=100,Sa)'
'PCE defl'	6	'Pce, Impl Pr Defl:Pce (1987=100)'
'PCE defl: dlbes'	6	'Pce, Impl Pr Defl:Pce; Durables (1987=100)'

References

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

J. Bai and S. Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2007.

E. Bair and R. Tibshirani. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biology*, 2:511–522, 2004.

E. Bair, T. Hastie, D. Paul, and R. Tibshirani. Prediction by supervised principal components. *Journal of the American Statistical Association*, 101(473):119–137, 2006.

J. Boivin and S. Ng. Undertanding and comparing factor based forecasts. *International Journal of Central Banking*, 1(3):117–152, 2005.

J. Boivin and S. Ng. Are more data always better for factor analysis. *Journal of Econometrics*, 132:169–194, 2006.

L. Breiman. Bagging predictors. *Machine Learning*, 36:105–139, 1996.

P. Buhlmann and B. Yu. Analyzing bagging. *Annals of Statistics*, 30:927–961, 2002.

D. Donoho and I Johnstone. Regularization and variable selection via te elastic net. *Ideal Spatial Adaptiationby Wavelet Shrinkage, Biometrika*, 81:425–455, 1994.

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

A. E. Hoerl and R. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970.

A. Inoue and L. Kilian. How useful is bagging in forecasting economic time series? A case study of U.S. CPI inflatoin. *manuscript, University of Michigan*, 2005.

W. Newey and K. West. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econnometrica*, 55:703–708, 1987.

L.A. Rastrigin. The convergence of the random search method in the extremal control of a many parameter system. *Automation and Remote Control*, 24(10):1337–1342, 1963.

G. Schrack and M. Choit. Optimized relative step size random searches. *Mathematical Programming*, 10(1):230–244, 1976.

M.A. Schumer and K. Steiglitz. Adaptive step size random search. *IEEE Transactions on Automatic Control*, 13(3):270–276, 1968.

J.H. Stock and M. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162, 2002.

J.H. Stock and M. Watson. An empirical comparison of methods for forecasting using many predictors,. *mimeo, Princeton University*, 2004a.

R Tibshirani. Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society Series*, 58(1):267–288, 1996.

T.J van Dongen, P Klaassens, J.S Kop, and L.S Tijssen. Averaging forecasts across number of factors. 2013.

H. Zou. The adaptive lasso and its oracle properties. *The adaptive lasso and its oracle properties.*, 101(476):1418–1429, 2006.

H. Zou and T. Hastie. Regularization and variable selection via te elastic net. *Journal of Royal Statistical Society*, 67(2):301–320, 2005.