FORECASTING INFLATION: AN OVERVIEW

Jochem S. Kop

June 201

Abstract

This paper provides an overview of conventional and recent models used to forecast inflation. The evaluated models include autoregressive, leading indicator, heterogeneous autoregressive, Phillips Curve, factor and several threshold factor models. Forecasts of 1, 6, 12 and 24 month horizons by these models using up to 126 different macroeconomic predictors are evaluated over the period of 1990 until 2009. The forecasts by the different models are combined by constraint least squares regression to evaluate the possibility of a superior combined forecast. Conclusions include superior forecasting by factor models, especially by stage-wise regression using a least angle regression and by including an autoregressive term. Combining forecasts does not improve forecast quality because of individual model superiority.

1. Introduction

Recent advances in macro-economic data collection and storage technologies have opened new frontiers for forecasting. Traditional time series models incorporate only a few out of the hundreds of possible predictors in their forecasting models, such as autoregressive or vector autoregressive models. In recent research conducted by Stock and Watson (2002), another approach was sought which incorporated more variables in a concentrated manner using diffusion indexes. A model was formed with a certain number of factors computed by principle component analysis, on which an h-step ahead target variable was regressed. In this way, the most important information out of all these different time series could be used for forecasting. Convincing results were achieved, with factor model forecasts outperforming benchmark autoregressive (AR) models significantly for most macro-economic target variables.

There was only one exception: the pure factor models did not perform as convincingly when forecasting inflation (measured in the form of CPI). Stock and Watson (2002) acknowledged the forecasting power of the AR model for this highly auto-correlated variable, and added an autoregressive term to their factor model. The new model performed better than the benchmark models for inflation, but when comparing it to Phillips Curve derived forecasts, they were unable to find conclusive results. The potency of the Phillips Curve, which rests on real economic activity, was already re-acknowledged by Fuhrer (1995), and recent research by Stock & Watson (2002) has been unable to discredit it. Another benchmark model Stock & Watson (2002) use is a leading indicator model, which uses least squares regression of the target variable on a number of driving macroeconomic variables.

The Phillips Curve is the oldest of the models, being an inverse statistical relationship between unemployment and inflation, named after its reported discoverer A. W. Phillips (1958). The Phillips Curve was later attacked by several Nobel Prize winners, amongst whom Friedman (1968) and Phelps (1968). They contested the long term relationship between unemployment and inflation that the original Phillips Curve suggests, because this characteristic is inconsistent with the theory of rational expectations. They suggest an expectations-augmented Phillips curve which only depends on a short term relationship between unemployment and inflation. Throughout recent research, this new form of the Phillips Curve has remained empirically feasible. Moreover, it has been found to be a reliable method for forecasting inflation in the past (Gordon 1982; Fuhrer 1995; Gordon 1997; Staiger et al. 1997; Tootel 1994).

Stock and Watson (2002) provide a comparative analysis of AR, vector auto regression (VAR), leading indicator and factor models based on a large (>200) amount of macroeconomic time series. They find that factor models outperform the uni- and multivariate benchmarks for all their forecasted time series. In the case of inflation, it should be noted that adding an autoregressive term was found necessary to produce efficient results. They conclude that only a few of the factors of this model account for a large portion of the variance in the data. The question of how many factors were to be selected in the model was theoretically answered by Bai and Ng (2002), but their criterion is shown not to perform well in practice by Van Dongen et al. (2013). A simple solution that works well in practice for both Stock and Watson (2002) and Van Dongen et al. (2013) is using a Bayesian Information Criterion (BIC) that contains both the number of predictors and the number of observations in the penalty term.

Bai and Ng (2007) have used similar data and present a more sophisticated form of the factor models suggested by Stock and Watson. As a number of the variables used by Stock and Watson (2002) are highly correlated, some may very well be irrelevant in constructing a factor model. They use partial regressions on each of their predictor variables and compute t-statistics for each of them in their 'hard thresholding' approach. On the basis of a significance level, the variables can then be selected. In their 'soft thresholding approach', the correlation between predictor variables is taken into account, as this approach selects variables in groups to ensure complementary information and avoid multicollinearity. They use small partial samples and find that within these samples, they outperform the traditional factor models used by Stock and Watson (2002).

The techniques that are discussed in this section for forecasting claim different optimal methods. A plausible explanation for these conclusions is that the researchers forecasted in different samples. Inflation forecasting has been shown by D'Agostino (2006) and Stock and Watson (2007) to produce different results for different periods in time. In tumultuous periods the more refined models, such as the Phillips Curve and targeted factor models, are clearly superior to univariate benchmarks. From approximately 1980 onwards however, D'Agostino et al. (2006) identify a period they lable 'the Great Moderation', making comparative success against a benchmark harder than in earlier periods. Part of the reason for the start of the moderation period is the fiscal policy of governments to keep inflation around two percent.

This paper will provide an overview of the most successful techniques that have been used in the past to forecast inflation. Different uni- and multivariate benchmarks will be used to evaluate the performance of the more sophisticated methods, such as targeted principle components and the

Phillips Curve. Then, as Van Dongen et al. (2013) tried with different macroeconomic time series, a combination of these forecasts will be constructed by constraint least squares regression.

Besides providing an overview of the different methods, this research will also focus on using forecast averaging techniques. Stock and Watson (2004) have already found that averaging (or pooling) of factor model forecasts outperform single factor models. Timmermann (2005) provides a large amount of possible averaging techniques. Van Dongen et al. (2013) evaluate the most promising of these techniques and have found that using constraint least squares is most successful. This has as of yet been unsuccessful when forecasting inflation using different factor models. However, this research suggests combining forecasts that might have a smaller correlation with each other. The suggestion is to combine the different models that are described in the methods section using constraint least squares. This research is therefore different from Van Dongen et al. (2013) in the sense that it does not average different factor model forecasts, but altogether different model forecasts. This paper will attempt to answer the following main question; can forecast combination outperform individual model forecasts for inflation, or is there a superior single model?

Hypotheses include that the factor model with an autoregressive term will outperform the standard factor model, consistent with Van Dongen et al. (2013). Besides, the Phillips Curve's performance is expected to be close to the standard factor model, as Stock and Watson (2002) found. Moreover, least angle regression is also expected to outperform a standard factor model, as Bai and Ng (2002) found. Lastly, whether the combination of forecasts will be more efficient than any single model depends on how correlated they are.

2. Data

Similar data will be used as in Stock & Watson (2002). This provides a large (126) number of variables over a time span of 44 usable years. The period from 1960 to 1989 will be used to compute the original principle components and estimate the parameters in the model. Using approximately 60% of the data to calibrate the model follows other authors in the literature. Therefore, data from 1990 to 2009 will be used for forecasting. The data will be transformed according to expert opinion and statistical (unit root) tests. The transformations will include first and second (logarithmic) differences. Lastly, the data will be standardized and outliers exceeding 10 times the interquantile range will be removed. The target variable will be the consumer price index (CPI). This time series will be transformed differently; it will be annualized and h-step ahead second differences will be taken, as the variable contains a second unit root. This transformation allows for easy reverse transformation after forecasting, and can be seen in equation 1.

$$y_{t+h} = \frac{1200}{h}\ln\left(\frac{CPI_{t+h}}{CPI_t}\right) - 1200\ln\left(\frac{CPI_t}{CPI_{t-1}}\right) \quad (1)$$

3. Method

3.1 Benchmarks

The autoregressive models will be used with an amount of lags of y_t based on the Bayesian information criterion (BIC) as well as the simplest AR model containing only a single lag. Leading indicator models will be ordinary least square (OLS) models with lags of a few important macroeconomic variables as well as lags of y_t. These variables have been shown to perform well in the literature by Stock & Watson (2002). The form of the leading indicator model can be seen in equation 2.

$$\hat{y}^h_{T+h|T} = \beta_{h0} + \sum_{i=1}^{m} \beta'_{hi} W_T + \sum_{j=1}^{p} \gamma'_{hj} y_{T-j+1} \quad (2)$$

The selected variables included in W_T are the total unemployment rate, real manufacturing and trade sales, housing starts, the interest rate spread between 1-year U.S. treasury bonds and the federal funds rate, the nominal M1 money supply and the federal funds overnight interest rate. The selection of the number of lags of y_t will be based on the BIC. The index i counts up to m, the number of variables in W_T, and the index j counts up to p, the number of lags of y_T.

3.2 Heterogeneous Autoregressive Model

The heterogeneous autoregressive (HAR) model was introduced by Corsi (2009) to model the realized volatility of foreign exchange indexes. The model uses averages over different periods of time to have a rather large memory without losing the parsimonious property, which would have been the case if all the separate lags were included. For example, a HAR model could contain the average change in a variable over the last month, the last year and the last decade in three regressors. Since inflation has been shown to follow a clear trend in the forecast period (2% a year) including a model that can easily catch this trend and amend for recent changes could hold strong forecasting potential. A requirement of using a HAR model is a slowly decaying autocorrelation. This requirement is met since the autocorrelation of the CPI is still approximately 0.8 for 30 lags. It can capture a large time span using few variables, which is a quality not found in any of the other models. The model is described by equations 3 and 4.

$$y_T^k = \frac{1}{k}\sum_{i=1}^{k} y_{T-i-1} \quad (3)$$

$$\hat{y}_{T+h|T}^h = \beta_{h0} + \beta_{h1}\, y_T + \beta_{h2}\, y_T^6 + \beta_{h3}\, y_T^{30} \quad (4)$$

As can be seen in equation 4, the model that will be used here has an intercept, one standard lag, the average over the last 6 months and the average over the last 30 months.

3.3 Phillips Curve

The method of the Phillips curve that will be used in forecasting is based on lags of short-term unemployment and of y_t . The Phillips curve inflation forecasts considered here have the form that is described in equation 5.

$$\hat{y}_{T+h|T}^h = \beta_{h0} + \sum_{i=1}^{m} \beta_{hi}' U_{T-j+1} + \sum_{j=1}^{p} \gamma_{hj}' y_{T-j+1} + \sum_{k=1}^{q} \mu_{hk} P_{T-k+1} \quad (5)$$

U_t consists of the unemployment rate and the index i sums over m - 1 of its lags. P_T is the relative price of food and energy (current and one lagged value only). Lags m and p are chosen sequentially by the BIC with 1 ≤ m ≤ 6 and 0 ≤ p ≤ 6, as the correlogram is a declining function still measuring >0.8 at 24 month lags. When comparing the leading indicator model and the Phillips Curve, many similarities can be observed. Both include lags of y_t, unemployment and an intercept. The differences are only that the leading indicator model includes other variables besides lags of y_t and W_t whereas the Phillips Curve allows for more unemployment lags.

3.4 Factor Models

For factor models the method and forecasts structure proposed by Stock & Watson (2002) will be followed. Let y_{t+h} be the forecasted variable at time t+h and X_t the set of predictors at time t. The assumption is made that both y_{t+h} and X_t can be described by a limited amount of p factors. As in previous research by Stock & Watson (2002), two more assumptions are made. The first assumes a finite number of lag polynomials of factors when describing both y_{t+h} and X_t, which allows formulating equations 6 and 7 in their present form. F_t is a 126 by 126 square matrix of factors.

$$y_{t+h} = \alpha + \beta' F_t + \gamma(L) y_t + \varepsilon_{t+h} \quad (6)$$

$$X_t = \Lambda F_t + \epsilon_t \quad (7)$$

The factors in equations 6 and 7 can now be estimated using principal components, as the relationship is assumed linear. The costs of this limitation are possibly larger prediction errors, as limitations have been imposed on the model configuration. The second assumption is the use of direct forecasting. Another possibility would be iterative updates of the set F_t using a vector autoregressive approach. The upside of the indirect method is the possibility of a non-linear relationship. The downside is the large number of parameters that have to be estimated. Like Stock & Watson (2002), the direct approach is preferred here, which will be applied as shown in equation 8. The number of factors used in the factor model will be three, which is based on the BIC results in Van Dongen et al. (2013).

$$\hat{y}^h_{T+h|T} = \alpha + \beta_h' F_t + \gamma(L) y_t \quad (8)$$

3.5 Targeted Principle Components

3.5.1 Hard Thresholding

The use of targeted principle components is an extension on the aforementioned factor models. In this method the use of all variables in constructing the factors is questioned as some may contain irrelevant information for the target variable. Hard thresholding is one of the methods used by Bai and Ng (2007) to select the variables that do contain enough relevant information to warrant their use in the factor model. This method uses t-values from a regression of the target variable on each of the individual 126 predictors, since the assumption is made that principle components can be applied to both time series. The regression equation can be found in equation 9.

$$y^h_{t+h} = \delta W_t + \Gamma X_t + \varepsilon_{t+h} \quad (9)$$

In equation 9, W_t contains a constant and lags of y_t. δ and Γ are parameters that are estimated using OLS. From this regression we obtain the t-statistic of Γ. This allows forming a ranking of the predictive power of X_{it}. The variable X_{it} is included in the set of targeted predictors if $|t_i|$ exceeds a threshold significance level of alpha, which is pre-set. When the relevant variables are isolated from the complete data set, principle components can be used to construct a model as was described in equation 6 and 7. Forecasts will be produced using equation 10, which is similar to equation 8 with the exception that different factors were computed using a subset of the variables used. As a results, the parameters in the regression are different as well.

$$\hat{y}^{h}_{T+h|T} = \bar{\alpha} + \overline{\beta_h}'\overline{F_t} + \bar{\gamma}(L)\overline{y_t} \tag{10}$$

3.5.1 Soft Thresholding

Hard thresholding has several drawbacks. Because of the rigidness of the decision rule the method can be sensitive to small changes in the data. Besides, hard thresholding can select overly correlated predictors that do well in the partial regression, but are too similar to complement each other in applying principle components. The idea of soft thresholding is that only the top predictors are kept. In order to achieve such an optimal set of predictors, the methods of soft thresholding select subsets of data and performs shrinkage simultaneously. There are multiple selection methods that are possible. Bai and Ng (2007) use a least absolute shrinkage and selection operator (LASSO), an elastic net (EN) and general least angle regression (LARS). Efron et al. (2004) have shown that LASSO and EN are special cases of least angle regression (LARS), and can therefore be seen as extensions with extra restrictions. Bai and Ng (2007) find that LARS in combination with principle components almost universally performs best. Since the results of Bai and Ng (2007) have already shown that LARS is superior over both LASSO and EN, only the former will be included in this paper's overview. It must be noted that the elastic net extension of LARS can still be improved over Bai and Ng's (2007) application by a better setting of the shrinkage parameter lambda, but this is beyond the scope of this paper. LARS is a stage wise selection where predictors are selected by the highest correlation with the residual vector of the model with the previously selected predictors. As a result, the variable that has the most unknown information always has the algorithm's preference. This way of selection has the advantage of being less aggressive than forward regressions that tend to be too aggressive in eliminating correlated predictors. Let û be the estimate of y with k predictors and let ĉ = X'(y - $\hat{u}_{\boldsymbol{k}}$). Define K then as the set of indices corresponding to the variables with the highest absolute correlation.

$$\hat{C} = \max_j(\hat{c}_j) \text{ and } K = \{ j \, \varepsilon \, |\hat{c}_j| = |\hat{C}| \} \tag{11}$$

Let $s_j = sign(\hat{c}_j)$ and define the active matrix corresponding to K as

$$X_k = (s_j x_j)_{j \, \varepsilon \, K} \tag{12}$$

Let $G_k = X_k{}'X_k$ and $A_k = 1_k{}'G_K^{-1}1_k$ where 1_k is a vector of K ones. A unit equiangular vector with columns of the active set matrix X_k can be defined as

$$u_k = X_k v_k \quad \text{with } u \quad = 0 \text{ and } v_k = A_k G_K^{-1} 1_k \text{ and } a_k = X' u_k \tag{13}$$

So that $X'u_k = A_k 1_k$. LARS then updates û as

$$\hat{u}^{new} = \hat{u} + \gamma u_k \tag{14}$$

$$\text{Where } \gamma = \min{}^{+}_{j \, \varepsilon \, K} \left(\frac{\hat{C} - \hat{c}_j}{A_k - a_j}, \frac{\hat{C} + \hat{c}_j}{A_k + a_j} \right) \tag{15}$$

LARS has several advantages over hard thresholding. It gives a ranking of the predictors taking into account the influence of other predictors. Moreover, the algorithm implicitly avoids selecting (overly) correlated predictors, as selecting a single predictor from a group of correlated predictors will ensure that the residue will have a low correlation with the rest of the variables out of that group. Although LARS is computationally as fast as OLS, it will still require quite some computing power to evaluate all variables, especially when the number of predictors is set high. Following Bai and Ng (2007), the number of predictors is set at 30. Subsequently a factor model will be constructed following equations 6,7 and 10, after which the amount of factors will be selected using BIC.

3.6 Constraint Least Squares Averaging

To combine the forecasts discussed in this section the method of constraint least squares (CLS) regression will be utilized. The CLS-weights are based on the suggestions by Timmermann (2004) and contain the following restrictions. A convexity constraint is included to ensure that the combined forecast does not leave the range of individual ones. An intercept is included to allow the regression to deal with possible bias in the individual forecasts. The weights will be computed using both moving and expanding windows and will include (initial) window sizes of 10,30 and 60 months. The weights will naturally be updated after every step in time. The regression equation can be found in equation 16.

$$\hat{y}^h_{T+h|T} = \hat{a}^h_t + \sum_{l=1}^{r} w^h_{r,t} \hat{y}^{h,r}_{T+h|T} \quad \text{subject to} \quad \sum_{l=1}^{r} w^h_{l,t} \leq 1 \text{ and } w^h_{l,t} \geq 0 \tag{16}$$

Here $w^h_{r,t}$ is the weight of forecast r on time T. Index l sums over the amount of forecasts r.

4. Results

4.1 Model Configuration

Using the Bayesian information criterion, the amount of lags of CPI and unemployment that are included in the AR, leading indicator and Phillips curve models are shown in table 1. The associated BIC scores can be found in the Appendix. The imposed maximum of six lags follows Stock & Watson (2002). The results for unemployment lag selection support the monetarist claim of Friedman (1968); the inverse relationship between inflation and unemployment is indeed short term, as the amount of lags is low for every forecast horizon. The CPI lag selection shows more lags for smaller forecasts horizons and less for larger horizons. These results were to be expected, as the correlation of a sixth lag with a 1 month ahead forecast will most probably be higher than for a 24 month ahead forecast.

BIC outcomes	h = 1	h = 6	h = 12	h = 24
CPI lags	6	6	2	3
Unemployment lags	1	2	1	1

T*able 1 contains the number of lags of CPI and Unemployment that should be included based on the BIC.*

4.2 Uni- and Multivariate Results

Table 2 displays the results for all uni- and multivariate models discussed in the method section. These include two AR models, one based on the BIC and one with just a single lag, as simple models have been shown to produce superior forecasts over larger ones (Stock and Watson 2002). The LI model and HAR model are also included. Lastly, there are two Phillips Curve (PC) models, one based on the BIC and one simple (1,1) version. The values in the table are relative forecast mean squared errors with the AR(BIC) model as a benchmark. All results in the following section are computed by equation 17. The actual AR(BIC) FMSE's are included in the Appendix for reproducibility.

$$RFMSE_{model,h} = \frac{FMSE_{model,h}}{FMSE_{AR(BIC),h}} \quad (17)$$

RFMSE	Horizon	AR (BIC)	AR(1)	LI	HAR	PC(BIC)	PC(1,1)
	h = 1	1.00	1.08	1.05	1.10	1.10	1.03
CPI-U	h = 6	1.00	0.98	0.98	0.97	0.95	0.94
	h = 12	1.00	0.99	1.02	1.06	1.01	0.99
	h = 24	1.00	1.01	1.20	1.00	1.00	1.02

Table 2 contains the relative forecast means squared errors for all uni- and multivariate models. The results are relative to a benchmark of an autoregressive model with a BIC suggested amount of lags.

The difference between the AR model with a single lag or the BIC version is small. The BIC values for the different lags – which can be found in the Appendix – did not vary much. These criterion values explain that the amount of lags does not really affect the forecasting performance. The same argument can be made for the negligible difference between the forecasting performances of the two Phillips Curves.

The results for the leading indicator model are inferior to those of the AR benchmark for most forecasting horizons. The extra variables included in this model apparently do not help, but rather hurt the model's forecasting performance. In this case the parsimony of the AR model is superior to the better fit created by the LI model. The HAR model that possesses a long memory does not outperform the AR model either. Although the CPI clearly follows a trend, the average over 30 months does not help forecasting performance. Moreover, the 6 month average information in the HAR model is already included in a more detailed manner in the AR(BIC) model where necessary, as for the 1 and 6 month horizons six lags are included. The Phillips Curves perform slightly worse for the 1 month forecast, slightly better on the 6 month forecasts and similar to the AR model on the 12 and 24 month forecasts. The classical economist's argument that there cannot be a long-term variant of the short-term relationship between inflation and unemployment could explain these results. A six month horizon could convincingly be categorized as short term, whereas the further horizons cannot. As such, an improvement on the 6 month horizon due to the unemployment lag(s) can be rationalized. On the longer horizons the advantage over the AR model is no longer there, and the Phillips Curve is similar to an AR model.

4.3 Factor Model Results

Table 3 contains the results of the different factor models that were discussed in the methods section. These include a 3 factor model and a 3 factor model with an AR term. Also, two different hard thresholding models are included, one with a significance level of 0.05 and one with 0.10. Lastly, a least angle regression model with 30 predictors was evaluated.

RFMSE	Horizon	3 Factors	3 Factors AR	Hard(0.05)	Hard(0.10)	LARS(30)	LARS(30) AR
	h = 1	1.16	2.36	1.19	1.19	1.16	2.44
Cpi-U	h = 6	0.86	0.79	0.87	0.88	0.82	0.75
	h = 12	0.81	0.83	0.79	0.81	0.82	0.79
	h = 24	0.78	0.74	0.84	0.90	0.76	0.76

Table 3 contains the relative forecast means squared errors for a 3 factor model, a 3 factor model with an AR term, two hard thresholding models with different alphas, a LARS model and a LARS model with an AR term. The results are relative to a benchmark of an autoregressive model with a BIC suggested amount of lags.

Although none of the factor models perform better than the AR(BIC) model on the 1 month forecasting horizon, all factor models outperform the autoregressive models on the other horizons. As the 1 month horizon is the least hard to forecast, conclusively, factor models are superior for this forecasting period over the uni- and multivariate models.

The 3 factor model's forecast performance improves when an AR term is added. CPI is highly autocorrelated - which partly explains its relatively strong results in the previous section – and this stylized fact is the reason why the 3 factor AR model outperforms the 3 factor model on most of the relevant horizons. The hard thresholding does overall not improve on the standard 3 factor model. This may be because it does not take correlation between selected variables into account, which shows in the selection of variables. It selects all partial CPI's, which hardly complement each other's set of information. Setting the selection level alpha to 0.10 does only hurt the forecast performance, because too many predictors are included with harm the forecast performance with its extra noise. It should be noted that the hard thresholding model with an alpha of 0.05 performs best on the 12 months horizon. The LARS model performs quite strongly, overall beating the regular 3 factor model. The LARS model does not outperform the 3 factor AR model, but comes quite close.

The initial decision of not including an AR term in the original LARS model gave a ceteris paribus comparison to the original factor model. However, it can be concluded that both LARS and adding an AR term improve forecasting performance. Therefore, it is interesting to see whether a LARS model with an AR term does indeed provide a superior forecast. In the last column of table 3 the RFMSE by the LARS AR model shows that it does indeed outperform both the original LARS model as the 3 factor AR model.

4.4 Averaging Results

Table 4 contains expanding and moving window combinations of all models in this section computed using constraint-least squares. FaEXP and FaMov models are included, which are similarly computed combinations using only the factor models, which were most successful in forecasting.

RFMSE	Horizon	EXP(10)	EXP(30)	EXP(60)	MOV(10)	MOV(30)	MOV(60)	FaEXP(10)	FaMOV(10)
	h = 1	1.07	1.16	1.33	1.11	1.22	1.35	1.20	1.28
CPI-U	h = 6	0.83	0.91	1.06	0.94	0.91	1.06	0.83	0.90
	h = 12	0.85	0.92	1.08	0.99	0.94	1.10	0.85	0.99
	h = 24	0.82	0.90	1.07	0.87	0.95	1.17	0.82	0.86

Table 4 contains the relative forecast means squared errors of constraint least squares combinations of forecasts produced by all the models in the method section except the LARS AR model. The table contains moving (MOV) and expanding (EXP) windows with an (initial) size of (…). In the case of a moving window the weights are estimated using $\hat{y}^{h}_{T+h|T} = \hat{a}^{h}_{t} + \sum_{l=1}^{r} w^{h}_{r,t}\hat{y}^{h,r}_{T+h|T}$ *where* $\hat{Y}^{h}_{T+h|T}$ *and* $\hat{Y}^{h,r}_{T+h|T}$ *are column vectors of length (...). In the case of an expanding window* $\hat{Y}^{h}_{T+h|T}$ *and* $\hat{Y}^{h,r}_{T+h|T}$ *are column vectors initially of length (…) which increase with t.The FaEXP() models are combinations with only factor models included, for which the same equation and vector lengths apply. The results are relative to a benchmark of an autoregressive model with a BIC suggested amount of lags.*

From the results in table 4 it becomes apparent that the expanding window outperforms the moving windows for every forecasting horizon and for every (initial) window size. To a large extent this is because the exponential window is more successful in selecting the superior factor models every time, which is almost always the better choice. There is some variance in the weights, but in the more successful methods the weights of LARS and the 3 factor model with an AR term almost without exception sum up to a value near 1. On first glance, it would seem odd that the performance of the expanding window deteriorates rapidly as the initial window is larger. The results of D’Agostino et al. (2006) and Bai and Ng (2007) give a plausible explanation. As the initial window size increases, the amount of out-of-sample observations that are used to evaluate forecasting decreases. The difference between the 60 and 10 initial windows is that the years 1990 until begin 1995 are not included in the forecast evaluation for the larger window model. Bai and Ng (2007) estimate the parameters of their models separately for different periods, and the relevant period is one of those. This first period is not at all volatile, which is therefore easy to forecast for all models. If this period is therefore taken into account when calculating the RFMSE, this value will be inclined to be lower than when the first years are excluded. Another argument for small (initial) windows is that in the HAR model results it already became clear that a long memory is unnecessary for forecasting CPI. A large window to estimate forecast weights will similarly not be expected to function optimally.

The FaEXP(10) model performs better than the FaMOV(10) model, which is consistent with the previously discussed results. The FaEXP(10) model does not perform better than the EXP(10) model. As becomes apparent from the weights in the EXP(10) model, the factor models are already selected almost exclusively in the distribution of forecast weights by the constraint least squares regression. More interesting however, is that the combination of forecasts is not better than the 3 factor model with an AR term or LARS. The intercept is quite low, but it might still deteriorate forecast quality. More importantly, the selection of other models than the 3 factor AR or LARS model is hardly ever efficient because these models are almost exclusively superior, as can be deduced from the individual RFMSE's of these models. Allowing for that possibility therefore can only hurt forecast performance. It can be concluded that if a superior model can be found, averaging forecast is not a viable idea.

5. Conclusion

In forecasting CPI using uni- and multivariate models it is of little consequence whether a large or a small number of autoregressive lags are selected. Adding leading indicator variables has not been shown to improve forecasting performance in the evaluated sample. Adding a large memory by adding 6 and 30 month averages does not improve forecast performance either. The Phillips Curve, which rests on the relationship between inflation and unemployment, does improve forecasts marginally in the short run (6 months ahead point forecast), but not for any of the other horizons. In comparison to the factor models, the Phillips Curve performed considerably worse, which rejects the initial hypothesis based on Stock and Watson (2002). A possible reason for the disappointing results of the Phillips Curve may be the absence of Gordon's (1982) variable that controls for the imposition and removal of the Nixon wage and price controls, which is beyond the scope of this paper.

Except on the 1 month horizon, factor models conclusively outperform all the uni- and multivariate models which is in accordance with the hypotheses. Using hard thresholding to select only the variables that have a statistically significant influence on the target variable does not improve forecasting results over the complete factor model. A more sophisticated approach, using LARS and thus taking into account the correlation between predictors, does provide a significant improvement over the complete 3 factor model. Besides the LARS succes, adding an AR term to the complete 3 factor model led to better forecasts performance as well.

Averaging or combining forecasts made by different models is only a viable idea if the forecast contain complementary information. Another implicit assumption must be that there is not a single superior model. If there is, any combination containing inferior forecasts can only compromise forecast quality.

The conclusions of this paper concerning thresholding support those found by Bai and Ng (2007) in a general sense. Both papers conclude that LARS forecasting results are superior over those of factor models based on all available predictors. A reason for any discrepancies can be the fact that Bai and Ng (2007) divided their forecast sample into 5 subsamples and estimated new parameters for each period, as they suspect structural breaks. This paper builds on the more general period deductions by D'Agostino (2006), which classifies the period after 1984 as a single period of moderation.

Finally, concerning forecasting CPI, the overview this paper presents provides a clear-cut conclusion. Considering that both using least angle regression as well as adding an autoregressive term improve forecast performance, the most potent model is a LARS 3 factor model with a single autoregressive term.

6. References

A. D'Agostino, D. Giannone and P. Surico, '(Un)Predictability and Macroeconomic Stability', *ECB Working Paper 605*, 2006.

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.

T.J. Van Dongen, P. Klaassens, J.S. Kop and L.S. Tijssen, ‘Averaging Forecasts Across Number of Factors’, 2013.

B. Effron, T. Hastie, I. Johnstone, and R. Tibshirani, ‘Least Angle Regression’, *Annals of Statistics,* 32:2, 407–499, 2004.

M. Friedman, ‘The role of monetary policy’, *American Economic Review*, 58(1) (March), 1-17, 1968.

J.C. Fuhrer, “The Phillips Curve is Alive and Well,” *New England Economic Review of the Federal Reserve Bank of Boston*, March/April 1995, 41-56, 1995.

R. J. Gordon, ‘Price Inertia and Ineffectiveness in the United States,’ *Journal of Political Economy*, 90, 1087-1117. 1982.

R. J. Gordon, ‘The Time-Varying NAIRU and its Implications for Economic Policy,’ *Journal of Economic Perspectives*, 11-32. 1997.

E. S. Phelps, ‘Money-Wage Dynamics and Labor-Market Equilibrium’, *Journal of Political Economy (Chicago University Press)*, 76: 678–711, 1968.

A. W. Phillips, ‘The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957,” *Economica*, New Series, Vol. 25, No. 100, 283-299, 1958.

D. Staiger, J. H. Stock and M. W. Watson, ‘The NAIRU, Unemployment, and Monetary Policy,’ *Journal of Economic Perspectives*, 11, 33-51. 1997.

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, “Combination forecasts of output growth in a seven-country data set,” *Journal of Forecasting*, 23:405-430, 2004.

J.H. Stock and M. Watson, ‘Why Has Inflation Become Harder to Forecast?’, *Journal of Money, Credit and Banking*, 39:3-34, 2007.

A. G. Timmermann, ‘Forecast combinations,’ *CEPR Discussion Paper*, (5361), 2005.

G. M. B. Tootell, ‘Restructuring the NAIRU, and the Phillips Curve,’ *New England Economic Review of the Federal Reserve Bank of Boston*, Sept./Oct., 31-44. 1994.

7. Appendix

The values of the BIC for different amounts of lags in an AR(X) are shown in the first table. The values of the BIC for different amount of unemployment lags in the Phillips Curve given the amount of lags of CPI are given in the second table. The optimal values are printed in bold.

BIC AR(X)	h = 1	h = 6	h = 12	h = 24
X = 1	1713	1600	1530	1444
X = 2	1714	1597	**1529**	1444
X = 3	1714	1586	1530	**1441**
X = 4	1710	1587	1532	1443
X = 5	1709	1586	1534	1443
X = 6	**1709**	**1583**	1535	1445

BIC Phillips	h = 1	h = 6	h = 12	h = 24
m = 1	**750**	698	**687**	**708**
m = 2	755	**690**	690	712
m = 3	758	694	695	717
m = 4	758	699	699	719
m = 5	762	701	703	723
m = 6	763	705	707	728

FMSE	Horizon	AR (BIC)
	h = 1	0.30
CPI-U	h = 6	13.31
	h = 12	34.72
	h = 24	107.80