Forecasting in data-rich environments

by

Luís Aguiar-Conraria Yongmiao Hong*

Department of Economics, Cornell University, Ithaca

Abstract

Stock and Watson (1998 and 1999) developed a factor-model approach which allows for big data sets to be systematically reduced to a few explanatory factors. In this paper two other methods are proposed. The first one, Partial Least Squares is imported from the Chemometrics literature. The second one, which is based on the Combination of Forecasts literature is a modification of Stock and Watson's method. We will call this method Principal Components Combination. These methods are

Key Words and Phrases: Combination of forecasts; factor analysis; forecasting; inflation; partial least squares; principal components

*We are grateful to Nicholas Kiefer, Karl Shell, Yi Wen, Mathis Schroeder and to the participants of the Cornell/Penn State Macroeconomics Workshop (2003) for constructive discussions and comments.

compared in an empirical application to inflation. We conclude that the method with the best overall performance is the Principal Components Combination.

1 Introduction

With enormous amounts of new information on several economic indicators arriving in continuous time, applied Macroeconomists have the problem of dealing with huge data sets and with hundreds of explanatory variables that can be useful for forecasting purposes. Usually we have at most a few hundred observations, making the use of so many variables impossible. Even with financial data, where much longer time series may easily be found, it is of dubious interest to consider hundreds of regressors. On the other hand, it is inefficient not to use all available information. More information should be helpful, not a problem.

One popular method to deal with this problem of excessive explanatory variables is the Principal Components Regression (PCR), which was applied by Sargent and Sims (1977) and Geweke (1977). More recently, this method has been successfully applied to US Macroeconomic data (Stock and Watson (1998, 1999, and 2002)), Bernanke and Boivin (2003). Marcellino, Stock and Watson (2003) applied this method to European data, but in their paper the Principal Components Regression could not consistently improve upon a simple Auto Regression model.

This literature is growing, and some nice asymptotic results have already been

derived — see Stock and Watson (1998), Forni, Hallin, Lippi, and Reichlin (2000) and Bai and Ng (2002). Still, some criticisms to this approach remain:

1. the results are very sensitive to the scale measurement of the variables,

2. the principal components are constructed without taking into consideration any relationship between the regressors and the dependent variable, and

3. the results are usually very hard to interpret.

If the only objective is to produce forecasts the third criticism is not a serious problem. Since in this paper we are focusing on forecasting we will discuss the first two criticisms.

One method, which tries to overcome the second problem is the Partial Least Squares (PLS). This method, specially known in the Chemometrics literature, was proposed by Wold (1975). PLS became popular during the 80's and, a decade later, several papers appeared in the Statistics literature analyzing the properties of this method. Although popular among chemometricians, this method has never become popular among econometricians and economists. One recent application of this method to economic data can be found in Gibson and Pritsker (2000).

A different branch of literature is the Combination of Forecasts proposed by Bates and Granger (1969) — see also Granger (1989) and Deutsch, Granger and Terävirsta (1994). This literature deals with the problem of having multiple forecasts for the

same variable. These authors, and others, argue that combining the different forecasts in a suitable manner leads to better predictions than the individual ones. Bates and Granger (1969) argued that a simple way to combine the different forecasts is to run a simple regression (OLS) to find the best combination. Note that if one has a big number of forecasts then simple OLS will not be appropriate. Chan, Stock and Watson (1999) make the argument that a suitable way to combine a big number of different forecasts is by PCR.

As an alternative to the Principal Components Regression and to the Partial Least Squares approach, we will combine the PCR with the Forecast Combination approach. To be more precise, we will use each explanatory variable to obtain a forecast for the dependent variable, and then combine the several forecasts using the PCR method. The proposed method has two advantages: it is scale invariant, thereby dealing with the first criticism, and it takes into consideration the explanatory power of the independent variables on the dependent variable.

The rest of the paper is organized as follows: section 2 sets up the basic model, and describes and relates two well-known estimation methods: PCR and PLS. In section 3 another method is proposed and described: Principal Components Combination (PCC). In section 4 the different methods are applied to inflation forecasting and compared. Section 5 concludes.

2 The Model

Let the basic data be given by $X = (x_1, ..., x_N)$ (a matrix of T observations of N independent variables) and y (a vector with T observations of the dependent variable). To facilitate interpretation assume that all the variables are already given in deviations from their means.

Consider a factor model of the form:

$$\begin{cases} \underset{(T\times 1)}{x_n} = \underset{(1\times 1)}{\lambda_{n,1}}\underset{(T\times 1)}{F_1} + \cdots + \underset{(1\times 1)}{\lambda_{n,K}}\underset{(T\times 1)}{F_K} + \underset{T\times 1}{e_n} \qquad n = 1, ..., N \\ \\ \underset{(T\times 1)}{y} = \underset{(1\times 1)}{\beta_1}\underset{(T\times 1)}{F_1} + \cdots + \underset{(1\times 1)}{\beta_K}\underset{(T\times 1)}{F_K} + \underset{T\times 1}{\varepsilon} \end{cases}$$

or, stacking the vectors together:

$$\begin{cases} \underset{(T\times N)}{X} = \underset{(T\times K)(K\times N)}{F}\underset{}{\lambda} + \underset{T\times N}{e} \\ \\ \underset{(T\times 1)}{y} = \underset{(T\times K)(K\times 1)}{F}\underset{}{\beta} + \underset{T\times 1}{\varepsilon} \end{cases} \tag{1}$$

The crucial assumption of this model is that y depends on X by only a few unobserved factors F and not in any other way. A factor model of this type is useful when the number of predictor variables is large (possibly even larger than T) making more common forecasting techniques unattractive or not feasible. Since F may contain lagged values of the underlying factors, this model is also called a dynamic factor model.

A natural way to estimate the parameters of the second equation of the system 1 is to replace the unobservable factors by estimated factors, and then estimate β by Ordinary Least Squares (OLS).

In the next subsections of the paper we consider two different methods to estimate the unobserved factors:

— Principal Components Regression (PCR), and

— Partial Least Squares (PLS).

The first one is becoming increasingly popular among econometricians, while the latter one is most popular in the Chemometrics literature. After that we will propose a modification of the PCR based on the Forecast Combination literature. This modification follows the spirit of PLS (by taking into consideration the effect of each predictor on the dependent variable) but essentially uses the analytical tools of PCR, with the advantage of being scale invariant.

2.1 Principal Components Regression

If the model described above is correct, then a possible procedure is to use the principal components of X as an estimate of the factors, and then use these to estimate de second equation of 1.

As Stone and Brooks (1990) showed, the idea of this method is to find the linear combinations of the X variables, such that a vector of weights, p_1, maximizes $p'X'Xp$, then p_2 is chosen to maximize $p'X'Xp$ such that $p'p_1 = 0$, with the vectors of weights being normalized to have unit distance. Thus p_1 is the normalized eigenvector of $X'X$ associated with the highest eigenvalue, p_2 is the normalized eigenvector associated

with the second highest eigenvalue, and so on.

By choosing the components associated with the highest eigenvalues one obtains the linear combinations of X that are orthogonal to each other and simultaneously have the highest variance. Intuitively, by choosing linear combinations with the highest variance possible, one is, in some sense, maximizing the information contained in those linear combinations. The number of estimated factors to include is a problem to which we will return later, when carrying out the empirical application.

Stock and Watson (1998), Forni *et al.* (2000) and Bai and Ng (2002) provide consistency results for this method. The asymptotic theory of this method has not only $T \to \infty$ but also $N \to \infty$. E.g. Bai and Ng assume that $E\left\|F_t\right\|^4 < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} F_t' F_t \to \Sigma_F$ as $T \to \infty$, with Σ_F being some positive definite matrix. They also assume that each factor has a nontrivial contribution to the variance of X: $\left\|\frac{\lambda'\lambda}{N} - D\right\| \to 0$ as $N \to \infty$, with D being some positive definite matrix, and $\|\lambda_n\| \leqslant \bar{\lambda} < \infty$. They also impose some conditions on the error terms of the X variables, allowing for heteroskedasticity in both time and cross section dimensions and some dependence between factors and the errors. Bai and Ng — and also Stock and Watson (1998) with a different set of assumptions — show that, asymptotically, the estimated factors and the true factors span the same space.

2.2 Partial Least Squares

In the previous section only the information contained in the $X-$data was used to estimate the factors. Obviously not all the information is used, as the relationship to the dependent variable is not considered.

PLS first appeared in the form of an algorithm (which is described bellow). Stone and Brooks (1990) showed that with PLS a vector of weights p1 is chosen to maximize $p'X'yy'Xp$. p_2 is chosen to maximize $p'X'yy'Xp$ such that $p'(X'X)p_1 = 0$. So one is finding the linear combination of the X variables which maximizes the squared sample covariance. Although PLS deals with the second criticism to PCR, it fails to address the first, as it is scale dependent as well. The usual procedure is to normalize all the variables to have unit variance. By doing this, maximizing the squared sample covariance amounts to maximizing the squared sample correlation.

There are at least two algorithms (one proposed by Wold and another proposed by Martens (1985)). Helland (1988) proved the equivalence between both and also provided a third method, which is computationally more convenient. Next we will describe the algorithmWold proposed and, after that, the alternative basis Helland proposed. For a description of both algorithms and the proof of their equivalence and also the equivalence of the alternative basis the reader is referred to Helland (1988). For some consistency results of PLS the reader can consult Naik and Tsai (2000)[1].

[1]Assuming that the explanatory variables are i.i.d.,these authors prove consistency of the PLS

2.2.1 The original PLS algorithm

Define $E_0 = X$ and $f_0 = y$. Define E_a and f_a recursively as:

$$\begin{aligned} E_a &= E_{a-1} - \hat{F}_a \hat{\lambda}_a' \\ f_a &= f_{a-1} - \hat{F}_a \hat{\beta}_a \end{aligned} \tag{2}$$

where $\hat{F}$ stands for the factor estimate.

We will need to determine $\hat{F}_a$, $\hat{\lambda}_a$ and $\hat{\beta}_a$ to fit into these equations. As with the Principal Components approach, each estimated factor $\hat{F}_a$ will be a linear combination of the X variables. E.g. for $a = 1$ we want:

$$\underset{T\times 1}{\hat{F}_1} = \sum_{n=1}^{N} \underset{T\times 1}{x_n}\, \underset{1\times 1}{p_{n1}} = \underset{T\times N}{X}\, \underset{N\times 1}{p_1} \tag{3}$$

Since we want to use the information contained in y to estimate the factors the weights will be chosen as:

$$p_1 = X'y \tag{4}$$

With this method, explanatory variables with a higher covariance with Y will receive a higher weight.

In general we have:

$$\hat{F}_a = E_{a-1} p_a \tag{5}$$

$$p_a = E_{a-1}' f_{a-1} \tag{6}$$

for $T \to \infty$. Extension to stationary variables is immediate.

We still need to determine $\hat{\lambda}_a$ and $\hat{\beta}_a$. To have the best fit in equations 2 we use the regression coefficients. For $a = 1$ we have $y = \hat{F}_1\hat{\beta}_1 + f_1$ and $X = \hat{F}_1\hat{\lambda}_1' + E_1$, so the regression coefficients are given by $\hat{\beta}_1 = \left(\hat{F}_1'\hat{F}_1\right)^{-1} y_1'\hat{F}_1$, and $\hat{\lambda}_1' = \left(\hat{F}_1'\hat{F}_1\right)^{-1} \hat{F}_1'X$. In general we have:

$$\hat{\lambda}_a = \left(\hat{F}_a'\hat{F}_a\right)^{-1} E_{a-1}'\hat{F}_a \tag{7}$$

$$\hat{\beta}_a = \left(\hat{F}_a'\hat{F}_a\right)^{-1} f_{a-1}'\hat{F}_a \tag{8}$$

Note that since the $\hat{F}_a$'s are orthogonal to each other (again see Helland (1988)), instead of formulas 7 and 8 we can use:

$$\hat{\lambda}_a = \left(\hat{F}_a'\hat{F}_a\right)^{-1} X'\hat{F}_a$$

$$\hat{\beta}_a = \left(\hat{F}_a'\hat{F}_a\right)^{-1} y'\hat{F}_a$$

With this method, the first factor to be estimated is $\hat{F}_1 = \underset{T\times N}{(X)}\underset{N\times 1}{(X'y)}$. So instead of finding the linear combination of the X variables that maximizes the variance, one is using the covariance between each predictor and the dependent variable as the weight of that variable. Then the second factor will be estimated using the covariance between $\left(X - \hat{F}_1\hat{\lambda}_1'\right)$ and $\left(y - \hat{F}_1 q_1'\right)$, and so on.

2.2.2 An alternative basis

The next proposition allows us to use a computationally more convenient

method.

Proposition 1 *Let $\mathcal{S}_A$ be the space spanned by $p_1, \ldots p_A$. As long as p_A is nonzero, an alternative basis for $\mathcal{S}_A$ is given by the vectors $(X'y)$, $(X'X)(X'y)$, $\ldots$, $(X'X)^{A-1}(X'y)$.*

Proof. *See Helland (1988) or Stone and Brooks (1990).* ■

2.3 Prediction, spectral representation and relation between PLS and PCR

For a moment consider a population version of the model described in system 1, so that there is no noise.

Consider the spectral decomposition of $S = X'X = \sum_{k=1}^{K} \varphi_k p_k p_k'$, where p_k is the eigenvector associated with the strictly positive eigenvalue φ_k ($X'X$ has rank K).

Note that, using the principal components regression, the predicted value for y is

$$\begin{aligned} \hat{y} &= F\left(F'F\right)^{-1} F'y \\ &= \sum_{k=1}^{K} Xp_k \left(p_k' X'X p_k\right)^{-1} p_k' \left(X'y\right) \end{aligned}$$

For prediction purposes all the non-relevant eigenvectors of $X'X$ can be deleted. Also note that if an eigenvalue has multiple eigenvectors associated with it, the corresponding terms can be substituted by only one term by rotating in eigenspaceswith equal eigenvalue, such that we get only one eigenvector. E.g. suppose that $\lambda_1 = \lambda_2$, then we can replace p_1 and p_2 by $p_1^* = \left(\frac{p_1 p_1' + p_2 p_2'}{\left(\left(p_1' s\right)^2 + \left(p_2' s\right)^2\right)^{\frac{1}{2}}}\right)(X'y)$. Note that $p_1^{*\prime} p_1 = 1$, and that $p_1 p_2' \left(X'y\right) + p_2 p_2' \left(X'y\right) = p_1^* p_1^{*\prime} \left(X'y\right)$.

Definition 2 *The relevant eigenvectors of $X'X$ to predict y are the ones associated with different eigenvalues which satisfy $p_k'(X'y) \neq 0$. The corresponding factors $F_k = Xp_k$ are the relevant factors in X for prediction of y. Let A be the total number of relevant eigenvectors.*

Proposition 3 *The population PLS space has dimension A and when this minimal number of terms is used, the population PLS regression vector and the population PCR regression vector are equivalent.*

Proof. See Helland 1990. ■

This proposition tells us that the PLS and PCR regression vectors are equivalent when the appropriate basis is chosen. Some stopping rule must be defined when applying the algorithm and hence the previous results will only be approximate: with real and noisy data it is highly unlikely that we find exact repeated values for the eigenvalues or that $p_k'(X'y) = 0$ (the sample relevant components will be very close to $\min(N, T-1)$).

Maybe the biggest advantage of PLS over PCR is that the possible nonsense of giving a large weight to an irrelevant explanatory variable is avoided. E.g. suppose that the variable Xg is completely uncorrelated with y ($cov(X_i, y) \approx 0$). Using the PCR algorithm there is nothing to prevent this variable from receiving a large weight, while with the PLS approach this variable receives approximately zero weight.

3 Combination of Forecasts and Principal Components

Bates and Granger (1969) — see also Granger (1989) and Deutsch, Granger, and Terävirsta. (1994) — suggest that when there are several forecasts for the same variable one sensible thing to do is to combine these several forecasts. Several combination methods have already been by proposed (again the reader is invited to check the references already mentioned). Chan, Stock and Watson (1999) argue that a suitable way to combine the different forecasts is by modeling them as an approximate factor model.

If one has N explanatory variables, then, using univariate regressions it is possible to produce N forecasts that can be combined using the PCR approach. We will call this procedure Principal Components Combination (PCC).

Let us see in detail how to implement the PCC method:

1. project y onto the space spanned by each of the N explanatory variables: $z_n = x_n \left(x_n' x_n\right)^{-1} x_n' y$, for $n = 1, 2, ..., N$,

2. create a new matrix of explanatory variables: $Z = (z_1, ..., z_N)$,

3. find the eigenvectors u_i of $Z'Z$ associated with positive eigenvalues. Let u_1 be the eigenvector associated with the highest eigenvalue, u_2 with the second highest, and so on,

4. use as new regressors the variables Zu_A associated with the A highest eigenvalues.

By choosing the principal components one is choosing a linear combination of the explanatory variables (Z in this case) that maximizes the variance. In this case the variance of each individual predictor has a natural interpretation: it is the explained variance of y by the corresponding original explanatory variable. One is no longer finding the principal components without taking into consideration the information contained in y. The weight that each variable receives is not independent of the relationship between the regressors and the dependent variable. Variables with higher explanatory power are also the variables with the highest variance, and hence they will tend to receive a higher weight. On the other extreme, if some variable x_n has no explanatory power over y, then the estimated y's will be constant (since all variables are in deviations from the mean, z_n will be a column of zeros), and this variable will receive zero weight when constructing the principal components.

If we choose A components the estimated value for y is

$$\hat{y} = Z\left(u_1, ..., u_A\right)\left[\left(Z\left(u_1, ..., u_A\right)\right)' Z\left(u_1, ..., u_A\right)\right]^{-1}\left(Z\left(u_1, ..., u_A\right)\right)' y$$

The final forecasts will be independent of the scale of the original variables X, because the matrix Z will not be changed with the scale of the original variables.

Proposition 4 *Let K be the number of eigenvectors (p_k) of $X'X$ associated with*

nonzero eigenvalues and assume that $cov(y, x_n) \neq 0$, $n = 1, ..., N$. *Then* $(Zu_1, ..., Zu_K)$ *and* $(Xp_1, ..., Xp_K)$ *span the same space.*

Proof. Note that $a_n = (x_n' x_n)^{-1} X_n' y$ is a scalar different from zero as long as $cov(y, x_n) \neq 0$. So $z_n = a_n x_n$ and hence X and Z span the same space and the number of eigenvectors associated with nonzero eigenvalues of $X'X$ and $Z'Z$ are the same (i.e. K). Since $(Xp_1, ..., Xp_K)$ span the same space as X, and $(Zu_1, ..., Zu_K)$ span the same space as Z, we must have that $(Xp_1, ..., Xp_K)$ and $(Zu_1, ..., Zu_K)$ span the same space. ■

This proposition tells us that when considering the population version of the model PLS and PCC are equivalent, as long as all the components associated with strictly positive eigenvalues are used. In a sample regression this result will have some noise because the number of positive eigenvalues will be $\min(N, T-1)$, and obviously it is unfeasible to use so many components. In small samples, one would expect that when only a few components are considered then the components estimated by PCC will produce better forecasts (we will be able to confirm this later) but asymptotically, with N and T approaching infinity, the results should converge.

4 Empirical Application

In this section of the paper we will apply the previous methods to forecast inflation using monthly data.The data was taken from the DRI-Mcgraw Hill Basic Economics

database spanning a time horizon from October 1968 to March 2003. This amounts to 413 observations of 140 variables.

All these variables are economic indicators measuring different aspects of the economy activity, such as real output and income, employment, sales, consumption, housing starts inventories, stock prices, exchange rates, interest rates, monetary aggregates, wages and, obviously, inflation.

Most variables were logarithmized (namely all the strictly positive variables that were not in the form of rates or ratios). We individually tested (using the ADF and Phillips Perron tests) each series to check if it was stationary or not. In the cases in which the series were not stationary we took first differences.

We will produce h month ahead inflation forecasts using different specifications. We will estimate the model using T observations and use the estimated model to produce an out of sample inflation forecast and compare this forecast with the realized inflation rate. This will be done recursively for the complete sample Then the Mean Square Prediction Error (MSE) and the Mean Absolute Prediction Error (MAE) of the out of sample forecasts are computed and used to compare the accuracy of the different methods proposed. E.g., if we consider a sample size of 100 observations, we use the first 100 observations to predict the inflation of period 101. Then we will reestimate the model using observations 2-101 to produce a forecast of the inflation in period 102, and so on

As in Stock and Watson (1999) we will consider two different measures of inflation. One of them is the Consumer Price Index (with the mnemonic PUNEW) — a Laspeyres index — and the other is the Personal Consumer Expenditure deflator (with the mnemonic GMDC) — a chain weighting.

In the more general form, the equation to be estimated is:

$$\pi_{t+h}^{h} = \alpha + \boldsymbol{\beta}(L)\,\mathbf{x}_t + \gamma(L)\,\pi_t + e_{t+h}^{h} \tag{9}$$

The dependent variable is π_{t+h}^{h} is given by $\pi_{t+h}^{h} = \left(\left[\prod_{i=1}^{h} (1 + \pi_{t+i}) \right]^{\frac{1}{h}} - 1 \right)$. This specification can be thought of as predicting inflation over the next h months.

The regressor(s) $\mathbf{x}_t$ is (are) some explanatory variable(s) available at time t. $\boldsymbol{\beta}(L)$ is a polynomial vector in the lag operator L, and $\gamma(L)$ is a polynomial in the lag operator L.

We will consider several competing methods for the choice of $\mathbf{x}_t$:

- the Phillips curve: $\mathbf{x}_t$ is just the unemployment rate between all workers of 16 years or older of period t,
- the pure AR model: $\mathbf{x}_t$ is omitted,
- three other models: $\mathbf{x}_t$ is recursively chosen in each regression according to the methods described below.

The last three competing methods mentioned above are:

1. Principal Components Regression,

2. Principal Components Combination,

3. Partial Least Squares.

In the first case we compute the principal components, using the procedure described in section 3, and choose the one associated with the highest eigenvalue. Then to determine if we should include the component associated with the second highest eigenvalue we use a modified version of the Bayes Information Criterion (BIC), proposed by Bai and Ng (2002)[2]. If the inclusion of the second component is rejected the process stops, if not then the same criterion is used again to evaluate the score associated with the third eigenvalue, and so on. A maximum of 10 components is allowed. With the PCC the procedure is the same as with the PCR method. The only difference is that instead of considering the original variables, these are pre-transformed (as described in section 3).

For example, if the original variable is a vector x_i, we will work with $z_i = X_i(X_i'X_i)^{-1\prime}y$ (where y is the dependent variable, the h-period ahead inflation rate). Finally to estimate the components using the PLS method we use the alternative basis described in proposition 2. The first component to be included is $X(X'y)$. Then one

[2]Bai and Ng showed that the standard BIC can only consistently estimate the correct number of factors if the factors are known. If one has to estimate the factors then the BIC may not consistently estimate the correct number of factors. The same criterion was used by Marcellino et al. (2003).

checks if $X\left[\left(X'X\right)X'y\right]$ should be included. If the inclusion is rejected the process stops, if not we check if $X\left[\left(X'X\right)^2X'y\right]$ should also be included, and so on and so forth. Again a maximum of 10 components is allowed.

Two more things should be mentioned. First since the PLS and PCR are scale sensitive we followed the suggestion in the literature and, in each regression, we normalized all the variables to have unit variance. Although not reported, we also considered the case with no normalization. The performance of these two methods is severely worse without the normalization. we should also note that since we have 140 explanatory variables and when constructing the X matrix we include two more lags of each explanatory variable, the matrix of explanatory variables has 420 columns.

To choose the order of the polynomials of $\boldsymbol{\beta}\left(L\right)$ and $\gamma\left(L\right)$ we use the typical BIC.

4.1 Results

In tables 1 to 5 we can informally check the performance of the various methods. On the top part of each table we have the relative (to PCR) mean square forecast errors and in the lower part the relative mean absolute forecast error. we considered several sample sizes, so that one can evaluate the performance on small and on bigger samples. Naturally the bigger is the sample size the lesser is the number of feasible estimations.

By a simple counting procedure it is apparent that the PCC method is the method

giving the most accurate forecasts: in 76 times, out of 120, the PCC had the smallest out of sample relative forecast errors. PLS also performed reasonably well being able to produce the smallest mean forecast errors 32 times, followed by the PCR (8 times) and the AR model (4 times).

Taking the PCR model as the benchmark, we conclude that PCC was able to beat PCR 101 times (out of 120), while PLS produced more accurate forecasts than PCR (according to the two different criteria) 70 times. Comparing the PCC method with PLS we can see the PCC produces more accurate forecasts 84 times (out of 120).

To compare the performance of these methods in a more formal way we consider two tests. One is a sign test (see Diebold and Mariano (1995) for details), the other is the Diebold and Mariano Statistic (again see Diebold and Mariano (1995) for details) to test if the MSE and MAE of two different methods are statistically significantly different (the null being that the forecast performances are similar) — negative values of the test statistics mean that PCC performed better according to the criterion of the test. In tables 6 to 10 we have the results of the tests comparing PCR with PCC (bellow the value of each statistic we have the one sided $p-$value).

Of all the tests applied to each series of forecasts, only once it was concluded that the PCR had a significantly better performance (considering 10% significance level) than PCC – namely when predicting the 6 months inflation, using the GMDC price index, and the MAE criterion to evaluate the performance.

On the other hand we can see that PCC performs significantly better than PCR several times and according to the several tests. For example, when predicting the two years inflation, the PCC performance is always significantly better than PCR, according to the three different statistics (except when we have the sample size of 300). For shorter horizons, like one month or three month inflation forecasts although PCC systematically performs better, only sporadically the better performance is statistically significant. Looking at intermediate horizon forecasts (6 and 12 months), we conclude that about half of the times the difference between the performance of the two methods is statistically significant.

In tables 11 to 15, we can see the results of the same tests comparing PCC with PLS — as before, negative values for the test statistics mean that PCC performed better. PCC was significantly more accurate (considering a significance level of 10%) 81 times while PLS was significantly more precise 19 times. Given these results, it is fair to consider PCC as being the method with the overall best performance.

5 Conclusions

Stock and Watson (1999) considered several forecasting models to predict inflation in the US. Of the several models they considered, PCR was the one with the best performance. In this paper we took this model as a benchmark and proposed two other methods, which can be applied in similar situations. The main results of Stock

andWatson was reproduced in this paper: PCR leads to significant improvements over the typical AR model, or over the traditional Phillips curve.

To overcome some of the criticisms to the PCR method, two other methods were proposed:

- the Partial Least Squares, which is very well-known in the Chemometrics literature, and its relation with PCR has already been widely studied, and
- the Principal Components Combination, which tries to overcome the shortcomings of the PCR method by combining this method with the literature on combinations of forecasts. This method is scale invariant with respect to the original explanatory variables, and takes into consideration the explanatory power of each of the explanatory variables when choosing the weights to give to each variable.

PLS seems to produce better forecasts than the PCR method for longer horizons (one or two years inflation forecasts), but these results are not confirmed when considering smaller horizons.

PCC performs systematically better than PCR, and, more formally, using some tests, we concluded that performs significantly better several times. Comparing PLS with PCC, we can see that PCC performs better again.

References

Bai, J. and Ng, S. (2002), Determining the Number of Factors in Approximate Factor Models, Econometrica, 70, 191—221.

Bates, J. and C.W. J. Granger (1969), The Combination of Forecasts, Operational Research Quarterly, 20, 451—468.

Bernanke, B and Boivin, J. (2003), Monetary Policy in a Data-Rich Environment, Journal of Monetary Economics, 50, 525—546.

Chan, Y., Stock, J. and Watson, M. (1999), A Dynamic Factor Model Framework for Forecast Combination, Spanish Economic Review, 1, 91—121.

Deutsch, M., C. W. J. Granger, and T. Terävirsta (1994), The Combination of Forecasts Using Changing Weights, International Journal of Forecasting, 10, 47—57.

Diebold, F. and Mariano, R. (1995), Comparing Predictive Accuracy, Journal of Business & Economic Statistics, 13, 134 — 144.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), The Generalized Dynamic Factor Model: Identification and Estimation, Review of Economics and Statistics, 82, 540—554.

Geweke, J. (1977), The Dynamic Factor Analysis of Economic Time Series, in Aigner and Goldberger, Latent Variables in Socio-Economic Models, Amsterdam: North Holland.

Granger, C. (1989), Combining Forecasts — Twenty Years Later, Journal of Forecasting, 8, 167—173.

Helland, I. (1988), On the Structure of Partial Least Squares, Communications in Statistics – Simulation and Computation, 17, 581—607.

Helland, I. (1990), Partial Least Squares Regression and Statistical Models, Scandinavian Journal of Statistics, 17, 97-114.

Marcellino, M., Stock, J. andWatson, M. (2003), Macroeconomic forecasting in the Euro area: Country specific versus area-wide information, European Economic Review, 47, 1—18.

Martens, H. (1985), Multivariate Calibration, Doctoral Thesis, Technical University of Norway, Trondheim.

Gibson. M. and Pritsker, M. (2000), Improving grid-based methods for estimating value at risk of fixed-income portfolios, Finance and Economics Discussion Series 2000-25, Board of Governors of the Federal Reserve System.

Naik, P. and Tsai, C. (2000), Partial Least Squares for Single-Index Models, Journal of the Royal Statistics Association B, 62, 163—771

Sargent, T. and Sims, C. (1977), Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory, in C. Sims, New Methods in Business Cycle Research, Minneapolis: Federal Reserve Bank of Minneapolis, 45—109.

Stock, J., and Watson, M. (1998), Diffusion Indexes, NBER working paper.

Stock, J., and Watson, M. (1999), Forecasting Inflation, Journal of Monetary Economics, 44, 293—335.

Stock, J., and Watson, M. (2002), Macroeconomic Forecasting Using Diffusion Indexes, Journal of Business & Economic Statistics, 20, 147 — 162.

Stone, M. and Brooks, R. (1990), Continuum Regression: Cross-validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partials Least Squares and Principal Components Regression, Journal of the Royal Statistics Association B, 52, 237—269

Wold, H. (1975), Soft Modelling by Latent Variables: the Nonlinear Partial Least Squares Approach, in Perspectives in Probability and Statistics, Papers in Honour of M. S. Bartlett, ed. J. Gani, London: Academic Press.

6 Tables

Table 1: one month inflation

	GMDC					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.007	1.054	1.246	1.242	1.308	1.396
PLS	1.129	1.153	1.155	1.052	1.025	1.026
AR	0.947	1.018	1.195	1.244	1.337	1.396
PCRC	1.003	0.975	0.946	0.886	0.913	0.909
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.021	1.078	1.157	1.123	1.137	1.137
PLS	1.075	1.113	1.094	1.031	1.035	1.042
AR	0.989	1.048	1.128	1.127	1.148	1.130
PCRC	0.987	0.985	0.965	0.941	0.924	0.918
	PUNEW					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.071	1.034	1.154	1.181	1.252	1.191
PLS	0.946	1.013	1.137	1.370	1.342	1.179
AR	0.883	0.996	1.081	1.117	1.229	1.134
PCRC	0.948	1.085	0.931	0.929	0.936	0.882
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.029	1.063	1.137	1.100	1.129	1.106
PLS	0.996	1.008	1.089	1.173	1.171	1.127
AR	0.964	1.043	1.092	1.062	1.109	1.059
PCRC	0.989	1.018	0.960	0.963	0.956	0.965

Table 2: three months inflation

	GMDC					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.311	1.227	1.572	1.544	1.827	1.614
PLS	0.896	0.936	1.011	0.916	1.207	0.800
AR	1.165	1.187	1.433	1.512	1.896	1.569
PCRC	0.993	0.887	0.900	0.899	0.894	0.830
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.061	1.224	1.364	1.243	1.393	1.298
PLS	0.911	1.043	1.046	0.936	1.113	0.952
AR	1.078	1.204	1.299	1.232	1.402	1.249
PCRC	0.963	0.977	0.975	0.957	0.926	0.926
	PUNEW					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.299	1.256	1.521	1.314	1.718	1.770
PLS	0.849	0.910	1.017	1.076	1.484	1.445
AR	1.097	1.211	1.335	1.137	1.633	1.560
PCRC	0.786	0.951	0.992	0.929	0.931	0.797
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.103	1.203	1.304	1.161	1.323	1.390
PLS	0.935	1.012	1.054	1.043	1.193	1.180
AR	1.058	1.144	1.186	1.087	1.280	1.262
PCRC	0.889	0.987	1.000	0.984	0.944	0.892

Table 3: six months inflation

	GMDC					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.376	1.475	1.729	1.889	2.160	1.892
PLS	0.808	1.114	0.896	0.931	1.284	0.873
AR	1.350	1.483	1.537	1.782	2.224	1.760
PCRC	0.775	0.882	0.851	0.912	0.914	1.052
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.123	1.322	1.366	1.367	1.608	1.436
PLS	0.889	1.086	0.916	0.931	1.171	0.942
AR	1.193	1.304	1.291	1.318	1.607	1.358
PCRC	0.827	0.927	0.927	0.962	0.918	1.006
	PUNEW					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.531	1.541	2.159	1.828	2.051	2.180
PLS	0.814	0.910	1.071	1.009	1.235	1.121
AR	1.425	1.467	1.778	1.487	1.915	1.827
PCRC	0.674	0.816	1.120	1.001	0.973	1.003
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.184	1.327	1.530	1.426	1.518	1.592
PLS	0.868	0.936	1.025	1.040	1.144	1.088
AR	1.156	1.192	1.315	1.263	1.442	1.398
PCRC	0.813	0.906	1.089	1.085	0.972	1.080

Table 4: twelve months inflation

	GMDC					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	2.020	1.583	1.942	2.234	2.369	2.365
PLS	0.814	0.895	0.746	0.763	0.848	0.704
AR	1.712	1.614	1.660	1.878	2.353	2.009
PCRC	0.711	0.569	0.795	0.772	0.685	1.063
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.293	1.375	1.465	1.512	1.610	1.552
PLS	0.921	1.000	0.823	0.863	0.898	0.772
AR	1.412	1.366	1.340	1.363	1.583	1.385
PCRC	0.821	0.771	0.914	0.902	0.746	1.005
	PUNEW					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.936	1.792	2.373	2.189	2.404	2.234
PLS	0.788	0.811	0.962	0.879	0.935	0.639
AR	1.982	1.689	1.829	1.580	2.191	1.719
PCRC	0.583	0.652	0.936	0.880	0.903	0.805
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.285	1.349	1.588	1.556	1.707	1.654
PLS	0.873	0.951	0.985	0.951	1.020	0.793
AR	1.413	1.280	1.327	1.291	1.599	1.398
PCRC	0.753	0.838	1.032	1.016	0.906	0.900

Table 5: two years inflation

	GMDC					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.619	1.828	1.979	2.670	2.802	3.076
PLS	0.714	0.693	0.747	0.429	0.504	0.602
AR	2.080	1.872	1.719	1.976	2.722	2.500
PCRC	0.553	0.585	0.711	0.665	0.498	1.024
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.331	1.353	1.489	1.712	1.687	1.793
PLS	0.888	0.851	0.770	0.634	0.681	0.788
AR	1.576	1.376	1.353	1.458	1.643	1.597
PCRC	0.758	0.773	0.798	0.822	0.668	1.030
	PUNEW					
	Relative Mean Square Error					
sample size	50	100	150	200	250	300
PCR	1	1	1	1	1	1
Phillips	1.615	1.967	1.984	3.079	3.146	4.159
PLS	0.703	0.711	0.802	0.707	0.683	0.823
AR	2.345	1.928	1.557	1.993	2.893	3.182
PCRC	0.519	0.580	0.831	0.766	0.680	1.018
	Relative Mean Absolute Error					
PCR	1	1	1	1	1	1
Phillips	1.343	1.377	1.479	1.828	1.898	2.133
PLS	0.861	0.876	0.801	0.799	0.816	0.878
AR	1.603	1.386	1.241	1.429	1.800	1.840
PCRC	0.709	0.742	0.837	0.894	0.821	0.994

Table 6: tests for the one month inflation forecasts — PCC vs PCR

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	0.641	-1.270	-0.506	-1.556	-1.796	-0.600
p-value	0.261	0.102	0.306	0.060	0.036	0.274
DM statistic (MSE)	0.049	-0.342	-0.640	-1.268	-0.864	-0.727
p-value	0.481	0.366	0.261	0.102	0.194	0.234
DM statistic (MAE)	-0.420	-0.392	-1.065	-1.616	-1.794	-1.391
p-value	0.337	0.347	0.144	0.053	0.036	0.082
	PUNEW					
Sign statistic	-0.855	-0.115	-0.759	-1.697	-1.306	-1.200
p-value	0.196	0.454	0.224	0.045	0.096	0.115
DM statistic (MSE)	-0.674	1.069	-1.086	-0.904	-0.816	-1.524
p-value	0.250	0.142	0.139	0.183	0.207	0.064
DM statistic (MAE)	-0.284	0.472	-1.142	-0.953	-1.053	-0.739
p-value	0.388	0.319	0.127	0.170	0.146	0.230

Table 7: tests for the three months inflation forecasts — PCC vs PCR

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-1.608	0.000	-0.127	-1.137	-1.644	-0.202
p-value	0.054	0.500	0.449	0.128	0.050	0.420
DM statistic (MSE)	-0.077	-0.904	-1.098	-1.372	-1.163	-1.061
p-value	0.469	0.183	0.136	0.085	0.122	0.144
DM statistic (MAE)	-0.943	-0.448	-0.510	-1.048	-1.387	-0.888
p-value	0.173	0.327	0.305	0.147	0.083	0.187
	PUNEW					
Sign statistic	-1.823	0.116	0.000	-0.711	-1.151	-1.818
p-value	0.034	0.454	0.500	0.239	0.125	0.035
DM statistic (MSE)	-2.006	-0.598	-0.081	-0.751	-0.609	-1.404
p-value	0.022	0.275	0.468	0.226	0.271	0.080
DM statistic (MAE)	-2.258	-0.306	0.009	-0.290	-0.884	-1.381
p-value	0.012	0.380	0.497	0.386	0.188	0.084

Table 8: tests for the six months inflation forecasts — PCC vs PCR

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-4.899	-2.154	-0.447	-0.645	-1.080	0.103
p-value	0.000	0.016	0.327	0.260	0.140	0.459
DM statistic (MSE)	-2.027	-0.631	-1.526	-0.941	-0.714	0.394
p-value	0.021	0.264	0.063	0.173	0.238	0.347
DM statistic (MAE)	-3.116	-1.049	-1.227	-0.614	-0.913	0.060
p-value	0.001	0.147	0.110	0.270	0.181	0.476
	PUNEW					
Sign statistic	-3.499	-2.154	0.831	0.931	-0.415	0.718
p-value	0.000	0.016	0.203	0.176	0.339	0.236
DM statistic (MSE)	-2.539	-1.544	1.185	0.010	-0.231	0.021
p-value	0.006	0.061	0.118	0.496	0.409	0.492
DM statistic (MAE)	-3.048	-1.412	1.297	1.133	-0.398	0.851
p-value	0.001	0.079	0.097	0.129	0.345	0.197

Table 9: tests for the twelve months inflation forecasts — PCC vs PCR

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-2.879	-4.294	-1.100	-2.255	-3.986	0.742
p-value	0.002	0.000	0.136	0.012	0.000	0.229
DM statistic (MSE)	-2.625	-1.491	-1.622	-1.481	-2.371	0.321
p-value	0.004	0.068	0.052	0.069	0.009	0.374
DM statistic (MSE)	-3.694	-2.100	-1.534	-1.403	-2.548	0.040
p-value	0.000	0.018	0.063	0.080	0.005	0.484
	PUNEW					
Sign statistic	-4.182	-3.000	1.100	0.800	-1.272	-1.378
p-value	0.000	0.001	0.136	0.212	0.102	0.084
DM statistic (MSE)	-3.975	-1.520	-0.520	-0.643	-0.596	-0.914
p-value	0.000	0.064	0.301	0.260	0.276	0.180
DM statistic (MSE)	-4.669	-1.766	0.476	0.141	-0.763	-1.053
p-value	0.000	0.039	0.317	0.444	0.223	0.146

Table 10: tests for the two years inflation forecasts — PCC vs PCR

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-5.807	-4.867	-4.580	-2.330	-4.348	0.798
p-value	0.000	0.000	0.000	0.010	0.000	0.213
DM statistic (MSE)	-5.095	-2.217	-2.198	-2.477	-5.801	0.272
p-value	0.000	0.013	0.014	0.007	0.000	0.393
DM statistic (MSE)	-6.123	-3.071	-3.286	-3.445	-4.309	0.675
p-value	0.000	0.001	0.001	0.000	0.000	0.250
	PUNEW					
Sign statistic	-6.470	-5.468	-3.252	-2.931	-1.863	-0.114
p-value	0.000	0.000	0.001	0.002	0.031	0.455
DM statistic (MSE)	-2.817	-2.240	-2.607	-1.683	-2.005	0.598
p-value	0.002	0.013	0.005	0.046	0.022	0.275
DM statistic (MSE)	-4.482	-10.55	-2.507	-1.457	-1.506	-0.099
p-value	0.000	0.000	0.006	0.073	0.066	0.461

Table 11: tests for the one month inflation forecasts — PCC vs PLS

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-3.100	-3.233	-2.403	-1.414	-1.633	-1.000
p-value	0.001	0.001	0.008	0.079	0.051	0.159
DM statistic (MSE)	-2.161	-2.406	-2.550	-1.807	-1.023	-1.110
p-value	0.015	0.008	0.005	0.035	0.153	0.134
DM statistic (MAE)	-2.863	-3.444	-3.020	-1.798	-1.906	-1.829
p-value	0.002	0.000	0.001	0.036	0.028	0.034
	PUNEW					
Sign statistic	-1.604	-1.386	-2.403	-4.101	-3.266	-1.400
p-value	0.054	0.083	0.008	0.000	0.001	0.081
DM statistic (MSE)	0.504	0.926	-2.542	-3.940	-3.496	-2.311
p-value	0.307	0.177	0.006	0.000	0.000	0.010
DM statistic (MAE)	-0.010	0.065	-2.758	-3.652	-3.513	-2.380
p-value	0.496	0.474	0.003	0.000	0.000	0.009

Table 12: tests for the three months inflation forecasts — PCC vs PLS

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	0.322	-1.738	0.000	0.711	-2.466	0.000
p-value	0.374	0.041	0.500	0.239	0.007	0.500
DM statistic (MSE)	1.055	-0.621	-1.080	-0.163	-1.760	0.229
p-value	0.146	0.267	0.140	0.435	0.039	0.409
DM statistic (MAE)	1.216	-1.524	-1.232	0.352	-2.205	-0.275
p-value	0.112	0.064	0.109	0.362	0.014	0.392
	PUNEW					
Sign statistic	-1.501	-3.244	-0.381	-0.284	-3.124	-2.424
p-value	0.067	0.001	0.352	0.388	0.001	0.008
DM statistic (MSE)	-0.841	0.627	-0.331	-1.476	-3.345	-3.067
p-value	0.200	0.265	0.370	0.070	0.000	0.001
DM statistic (MAE)	-1.130	-0.646	-1.014	-0.928	-2.984	-2.798
p-value	0.129	0.259	0.155	0.177	0.001	0.003

Table 13: tests for the six months inflation forecasts — PCC vs PLS

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-1.669	-3.086	0.958	0.501	-3.239	1.949
p-value	0.048	0.001	0.169	0.308	0.001	0.026
DM statistic (MSE)	-0.562	-1.490	-0.307	-0.127	-2.047	0.670
p-value	0.287	0.068	0.379	0.449	0.020	0.251
DM statistic (MAE)	-1.565	-2.408	0.155	0.404	-2.388	0.361
p-value	0.059	0.008	0.438	0.343	0.008	0.359
	PUNEW					
Sign statistic	-0.700	-1.572	1.086	1.217	-1.246	1.129
p-value	0.242	0.058	0.139	0.112	0.106	0.130
DM statistic (MSE)	-1.573	-0.948	0.285	-0.049	-1.731	-0.451
p-value	0.058	0.172	0.388	0.480	0.042	0.326
DM statistic (MAE)	-0.953	-0.540	0.723	0.454	-1.860	-0.065
p-value	0.170	0.295	0.235	0.325	0.031	0.474

Table 14: tests for the twelve months inflation forecasts — PCC vs PLS

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-1.684	-5.588	2.652	0.800	-2.120	3.922
p-value	0.046	0.000	0.004	0.212	0.017	0.000
DM statistic (MSE)	-1.570	-2.536	0.577	0.053	-1.784	1.462
p-value	0.058	0.006	0.282	0.479	0.037	0.072
DM statistic (MAE)	-1.765	-3.332	1.312	0.380	-2.142	1.430
p-value	0.039	0.000	0.095	0.352	0.016	0.076
	PUNEW					
Sign statistic	-3.313	-2.529	1.488	0.364	-1.612	1.378
p-value	0.000	0.006	0.068	0.358	0.054	0.084
DM statistic (MSE)	-2.840	-1.450	-0.305	0.007	-0.184	0.745
p-value	0.002	0.074	0.380	0.497	0.427	0.228
DM statistic (MAE)	-2.219	-2.155	0.774	1.445	-0.844	0.598
p-value	0.013	0.016	0.220	0.074	0.199	0.275

Table 15: tests for the two years inflation forecasts — PCC vs PLS

	GMDC					
sample size	50	100	150	200	250	300
Sign statistic	-4.479	-1.742	1.792	3.533	-0.266	4.900
p-value	0.000	0.041	0.037	0.000	0.395	0.000
DM statistic (MSE)	-1.674	-1.127	-0.361	2.387	-0.046	2.220
p-value	0.047	0.130	0.359	0.008	0.481	0.013
DM statistic (MAE)	-1.729	-1.082	0.383	2.387	-0.152	2.247
p-value	0.042	0.140	0.351	0.009	0.439	0.012
	PUNEW					
Sign statistic	-4.258	-3.305	1.261	2.029	0.089	2.165
p-value	0.000	0.000	0.104	0.021	0.465	0.015
DM statistic (MSE)	-1.903	-1.190	0.361	0.858	-0.034	0.814
p-value	0.029	0.117	0.359	0.195	0.486	0.208
DM statistic (MAE)	-2.308	-2.04	0.489	1.520	0.070	0.780
p-value	0.011	0.021	0.312	0.064	0.472	0.218