Forecasting 1 to h steps ahead using partial least squares*

Philip Hans Franses†

Econometric Institute, Erasmus University Rotterdam

November 10, 2006

Econometric Institute Report 2006-47

*I thank Dick van Dijk and Richard Paap for helpful discussions. Rianne Legerstee provided very useful research assistance.

†Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands, `email: franses@few.eur.nl`

Abstract

This paper proposes a methodology to jointly generate optimal forecasts from an autoregression of order p for 1 to h steps ahead. The relevant model is a Partial Least Squares Autoregression, which is positioned in between a single AR(p) model for all forecast horizons and different AR models for different horizons. Representation, estimation and forecasting using the new model are discussed. An illustration for US industrial production shows the merits of the methodology.

Key words: Partial least Squares, Autoregression, Forecasting

1 Introduction and motivation

This paper deals with an autoregressive time series model of order p for a time series y_t, $t = 1, 2, \ldots, n$ that is to be used for forecasting 1 to h steps ahead, where the focus is on point forecasts.

A single model for all horizons

The autoregressive model of order p (AR(p), that is

$$y_t = \mu + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \ldots + \rho_p y_{t-p} + \varepsilon_t, \tag{1}$$

is often used for forecasting. The parameters can be estimated using ordinary least squares [OLS]. The 1-step ahead forecast at time n is generated as

$$\hat{y}_{n+1} = \hat{\mu} + \hat{\rho}_1 y_n + \hat{\rho}_2 y_{n-1} + \ldots + \hat{\rho}_p y_{n-(p-1)}. \tag{2}$$

The 2-step ahead forecast is created as

$$\hat{y}_{n+2} = \hat{\mu} + \hat{\rho}_1 \hat{y}_{n+1} + \hat{\rho}_2 y_n + \ldots + \hat{\rho}_p y_{n-(p-2)}, \tag{3}$$

and so on, while finally, when $h > p$, the h-step forecast follows from the recursion

$$\hat{y}_{n+h} = \hat{\mu} + \hat{\rho}_1 \hat{y}_{n+h-1} + \hat{\rho}_2 \hat{y}_{n+h-2} + \hat{\rho}_p \hat{y}_{n-(p-h)}. \tag{4}$$

In this case, a single time series model is used for all forecast horizons. This method of substituting earlier forecasts to forecast further ahead implies that for h large, the forecasted values converge to the unconditional mean, that is,

$$\hat{y}_{n+h} \rightarrow \frac{\mu}{1 - \rho_1 - \rho_2 - \ldots - \rho_p}. \tag{5}$$

For each horizon a different model

There are various reasons why an alternative strategy for forecasting 1 to h steps ahead is better. The first is that OLS aims at minimizing the sum of squared $\hat{\varepsilon}_t$, which amounts to the sum of squared 1-step ahead forecast errors. Indeed, abstaining from estimation errors, the difference between (1) and (2) is ε_{n+1}. There is however no guarantee that this minimization also implies a minimum of the sum of squared

$h-$step forecasts errors. The second reason is that one may want to have different models for different forecast horizons, that is, for each h there would be a different model. The third reason is that for stationary time series $h-$step forecasts quickly converge to the unconditional mean, which may be implausible in some practical situations.

An obvious alternative is then to have different models for different forecast horizons, and hence to replace (1) by

$$y_{t+h} = \mu_h + \rho_{1,h} y_{t-1} + \rho_{2,h} y_{t-2} + \ldots + \rho_{p,h} y_{t-p} + \varepsilon_{t,h}, \tag{6}$$

for $h = 0, 1, 2, ..$ where the variance of $\varepsilon_{t,h}$ is σ_h^2 and hence can also vary with the forecast horizon[1]. This line of thought is followed in Tiao and Xu (1993), Weiss (1991), Bhansali (1996), and Kang (2003) among others, where these studies derive properties of OLS-based estimators, propose model selection criteria and give illustrations of the merits of this approach. Note that the model orders may also differ across horizons, so one may also use the notation p_h instead of p.

Towards something in between

To have a different model for different forecast horizons has an important shortcoming and that is that in these models the correlation between adjacent time series observations is not exploited. Indeed, in an autoregression the variables y_t and y_{t-j} are correlated, and so must be the forecasts y_{n+h} and y_{n+h-j} for any j. This suggests that yet an alternative view on forecasting from an AR(p) model would be that one aims to jointly predict $(y_{t+h}, y_{t+h-1}, y_{t+h-2}, \ldots, y_{t+1})$ from $(y_t, y_{t-1}, \ldots, y_{t-p})$. The regression technique that enables one to do this is called Partial Least Squares [PLS], and it is this method that shall be introduced in this paper for jointly forecasting 1 to h steps ahead from an AR(p) model. It will be shown that the resulting PLS Autoregression [PLSAR] amounts to a case in between (1) and (6).

The outline of this paper is as follows. In Section 2 the PLS Autoregressive model for order (h,p) is introduced. Parameter estimation and forecasting will be addressed

[1]This notion of having a different forecasting model for different forecasting horizons comes close to the idea behind periodic models, which allow for different models for different seasons, see Franses and Paap (2004) for a recent review.

as well. Section 3 illustrates this new model for quarterly seasonally adjusted US industrial production. It is shown that forecasts for 2 to 5 quarters ahead are indeed better when the PLSAR model is used. Section 4 deals with an outline of further research topics.

2 PLS Autoregression

The interest is in jointly predicting $(y_{t+h}, y_{t+h-1}, y_{t+h-2}, \ldots, y_{t+1})$ to be collected in Y, using the available information on the time series until and including t, that is for $(y_t, y_{t-1}, \ldots, y_{t-p})$, which will be collected in X. Note that the focus is on prediction and not on correlation. The latter would amount to computing the so-called canonical correlations (CCA), see Esposito Vinzi et al. (2007) for details of the PLS technique and its comparison with related methods[2].

Representation

The data matrices that are input to the regression problem are Y and X, and they are of size $n \times h$ and $n \times (p + 1)$, respectively[3]. The idea behind PLS is that it seeks for components of X which are also relevant for Y. PLS regression aims at finding a set of latent variables which together simultaneously decompose Y and X, given that these latent variables explain most of the covariance between Y and X. Note that this last feature makes PLS different from Principal Components Analysis (PCA).

Suppose there are k such latent variables, where k can take values from 0 to $p+1$. The first step is now to decompose X as

$$X = KW, \tag{7}$$

where K collects these k latent variables, with K is of size $n \times k$, and where the loadings are collected in a $k \times (p + 1)$ matrix W.

[2]Further readings on PLS and related techniques include Helland (1990) and Wold (1966). The last author is said to be the first to have put forward the PLS technique.

[3]It is assumed that there are n effective observations, and hence that the full sample contains $n + p$ observations where the first p are needed for start-up.

Next, it is assumed that the fit of Y, given the PLS regression model, is

$$\hat{Y} = KBC, \tag{8}$$

where the regression weights are collected in the $k \times k$ matrix B and where C is again a loading matrix, where now these loadings concern K on $\hat{Y}$. This C matrix is of size $k \times h$. These two equations underlie the PLS regression model, which in the present application is called a Partial Least Squares Autoregression of order (h, p), or PLSAR(h, p).

Estimation

The method of estimation of the components in K, the parameters in B and the loadings in W and C bears similarities with related techniques such as PCA and CCA, see also Esposito Vinzi et al. (2007) and for example Abdi (2003). A popular approach is an iterative least squares method, see Abdi (2003) for the sequential steps, and which is also used below. Another approach would follow eigenvalue and singular value decompositions, and in practice small numerical differences can be expected. Note that, as is usual, prior to estimation, the variables in Y and X are scaled towards $z-$scores (that is the mean is subtracted and the variables are scaled with their standard deviation) in order to facilitate computations.

Like PCA, the estimation method also delivers the degree of the variance of Y and and of X that is explained by each of the latent variables in K. Comparing the percentages of explained variance allows one to fix the value of k, if one indeed wants to make a choice.

Forecasting

Taking (7) and (8) together shows that the forecasting scheme based on the PLSAR is

$$\hat{Y} = X\hat{B}_{pls}, \tag{9}$$

where $\hat{B}_{pls}$ is a $(p+1) \times h$ matrix computed as $\hat{B}_{pls} = \hat{W}^{-1}\hat{B}\hat{C}$, where the inverse potentially is the Moore-Penrose inverse. In the present paper the focus is on the quality of the forecasts $\hat{Y}$, which of course are $\hat{y}_{n+h}, \hat{y}_{n+h-1}, \hat{y}_{n+h-2}, \ldots, \hat{y}_{n+1}$.

Interestingly, when K is a full rank matrix, that is k is equal to $p+1$, then the number of latent variables is equal to the number of explanatory variables. In that case the $\hat{B}_{pls}$ matrix has full rank too, and hence (9) implies a different model for each of the columns of Y, and hence a model like (6). In the exceptional case that $\hat{B}_{pls}$ has rank 1, the AR(p) model in (1) appears.

3 Illustration

To illustrate the PLSAR model, consider the quarterly index of US industrial production, for the period 1945.1 to 2000.4. The data have been seasonally adjusted. For the full sample, the estimated (partial) autocorrelation functions suggest that the value of p could well 4, which means an AR(5) in (1). Next, it is assumed that there is an interest in forecasting 1 to 5 quarters ahead.

The models

To compare models on their forecasting performance a recursive procedure is followed. First, the sample 1945.1 to 1990.4 is used, and forecasts are made for 1991.1 to 1992.1. Then the sample moves one quarter, that is, it becomes 1945.1 to 1991.1 and again 1- to 5-step ahead forecasts are made. This results in 40 1-step ahead forecasts, 39 2-step ahead forecasts and finally 36 5-step ahead forecasts. Each time, the model parameters are re-estimated. Before estimation, the variable is standardized to have mean zero and scaled by the standard deviation for the full sample. A second exercise is that we keep the estimation sample fixed at 120 observations, that is, a moving window sample. Each time we move on in time with 1 step, the first observation is dropped.

The forecasting experiment concerns a comparison of the same AR(5) model for all forecast horizons with five versions of the PLSAR(5,4) model. Each of these five versions assumes a different amount of latent variables. The case with 5 such latent variables is the same as assuming 5 different AR(5) models for the 5 different forecast horizons.

For the full sample, the full PLSAR model gives latent explanatory variables which contribute the fractions 0.376, 0.349, 0.139, 0.097 and 0.039 of the variance of Y. Hence, based on these values one would tentatively set k equal to 2.

Results

The forecasting results are displayed in Table 1. The first panel gives the results for the case of recursive samples, and the second panel deals with the 120-quarters moving window. With PLS_k we denote a Partial Least Squares Autoregression with k latent variables, for k is 1 to 4. The last column is the case with $k = 5$, which amounts to a different AR(5) model for each h. The first column is the AR(5) model with the same parameters for each forecast horizon h.

For the recursive samples, the results are quite conclusive. Of course, for 1-step ahead forecasts the AR(5) model is best. However, for horizons 2 to 5, the PLSAR models quickly become better relative to both competitors, notably the PLS_1 model with a single latent explanatory variable. Note that this would not be the selected model based on the explained fractions of the variance though.

For the moving samples, the results suggest that for horizons 1, 4 and 5 one would be better off using a single AR(5) model. For horizons 2 and 3, the PLSAR model gives the best forecasts. Different models for different horizons are never the best. This seems to echo the findings in Kang (2003) who finds that models like (6) are not very successful in practice.

4 Conclusion

This paper has put forward a simple autoregressive time series model that can be used to jointly predict 1- to h-steps ahead. The models nests two specific cases, one is an AR model with the same parameters for all horizons and the other is an AR model with different parameters for different forecast horizons. The illustrative results showed that the resultant Partial Least Squares Autoregression can deliver more accurate forecasts than these two specific cases. Hence, this new model deserves further analysis and application.

There are quite a number of further issues that need to be studied, additional to extensive applications to other time series. The first is that this paper only looked at point forecasts and not at forecast densities. Perhaps bootstrapping-based techniques can be used to retrieve this forecast distribution based on the errors defined by $\hat{Y} - X\hat{B}_{pls}$.

Another interesting area concerns the derivation of a formal test for the rank of $\hat{B}_{pls}$, which would allow for a choice between the various PLSAR models. Also, if the rank of that matrix is equal to 1, one ends up with the standard fixed-parameters AR model.

Finally, this paper dealt with a single time series, but one can easily extend the PLSAR model to the case of 2 or more variables. Even more interesting would be to put forward a method to analyse 2 or more such series when they are non-stationary and have a unit root. At present, the most popular test for cointegration is based on canonical correlation, see Johansen (1995), and perhaps a PLS-based method would give more power to the test procedure.

Table 1: Root Mean Squared Prediction Errors for an AR(5) model for all forecast horizons, PLSAR(5,4) models with the number of latent variables being 1, 2, 3, or 4, and finally, the AR(5) model with different parameters for different forecast horizons, denoted as $AR_h(5)$. Boldface numbers are smallest in the row.

Horizon h	$AR(5)$	PLS_1	PLS_2	PLS_3	PLS_4	$AR_h(5)$
			Recursive samples			
1	**0.893**	0.992	0.963	0.973	0.929	**0.893**
2	0.785	**0.779**	0.785	0.785	**0.779**	**0.779**
3	0.784	**0.770**	0.783	0.778	0.779	0.779
4	0.792	**0.771**	0.787	0.801	0.810	0.815
5	0.770	**0.766**	0.773	0.798	0.791	0.791
			Moving window samples			
1	**0.872**	0.979	0.923	0.902	0.900	**0.872**
2	0.776	**0.763**	0.770	0.775	0.776	0.778
3	0.775	0.772	0.772	**0.768**	0.769	0.771
4	**0.777**	0.796	0.803	0.810	0.819	0.816
5	**0.774**	0.788	0.809	0.814	0.812	0.813

References

Abdi, H. (2003), Partial Least Squares regression, in Lewis-Beck M., A. Bryman and T. Futing (eds.), *Encyclopedia of Social Sciences Research Methods*, Thousand Oaks (CA): Sage.

Bhansali, R.J. (1996), Asymptotically efficient autoregressive model selection for multistep prediction, *Annals of the Institute of Statistical Mathematics*, 48, 577-602.

Esposito Vinzi, V., W.W. Chin, J. Henseler and H. Wang (2007), *Handbook of Partial Least Squares*, Berlin: Springer.

Franses, P.H. and R. Paap (2004), *Periodic Time Series Models*, Oxford: Oxford University Press.

Helland, I.S. (1990), PLS regression and statistical models, *Scandinavian Journal of Statistics*, 17, 97-114.

Johansen, S. (1995), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.

Kang, I.-B. (2003), Multi-period forecasting using different models for different horizons: An application to US economic time series data, *International Journal of Forecasting*, 19, 387-400.

Shibata, R. (1980), Asymptotically efficient selection of the order of the model for estimating the parameters of a linear process, *Annals of Statistics*, 8, 147-164.

Tiao, G.C., and D. Xu (1993), Robustness of maximum likelihood estimates for multi-step predictors: the exponential case, *Biometrika*, 80, 623-641.

Weiss, A.A. (1991), Multi-step estimation and forecasting in dynamic models, *Journal of Econometrics*, 48, 135-149.

Wold, H. (1966), Estimation of principal components and related models by iterative least squares, in P.R. Krisnaiah 9ed.) *Multivariate Analysis*, New York: Academic Press.