Comparisons of Forecasting Methods with Many Predictors

Jin-Lung Lin [1]
Institute of Economics, Academia Sinica
Department of Economics, National Chengchi University
and
Ruey S. Tsay [2]
Graduate School of Business, University of Chicago

Abstract

In recent years, there is substantial interest in forecasting using many predictors. The methods used include principal component regression, partial least squares, ridge regression, combining forecasts, Bayesian model averaging, empirical Bayes methods, and vector autoregression. In this paper, we compare and discuss some of these forecasting methods using monthly U.S. economic variables consisting of a univariate dependent variable and 141 predictors with 430 observations. Multi-step ahead out-of-sample forecasts are used in the comparison. We then discuss practical implications of our empirical findings. The partial least squares with three components outperforms other methods in our empirical study of short-term forecasts, but the ridge regression fares better for longer-term forecasts. The combining forecast based on time-series regression models performs well. The principal component regression also provides accurate forecasts when the number of components used is between 3 and 9.

Keywords: combining forecast, diffusion index, macroeconomic forecasting, partial least squares, principal component regression, ridge regression.

[1] I would like to thank National Science Council, Taiwan, for providing financial support under the project NSC-92-2415-H-001-017 and NSC-94-2415-H-001-014.

[2] Research is supported in part by the Graduate School of Business, University of Chicago

1 Introduction

Macroeconomic variables of an economy are often inter-related, and they contain useful information in forecasting each other. The same situation also holds in other scientific fields. In fact, for many forecasting applications, lots of predictors are available to predict a dependent variable. For example, many economic variables are used to produce the forecasts of Gross Domestic Product (GDP) of a country. Stock and Watson (1989) apply the principal component regression (PCR) method to extract factors from a large number of macroeconomic variables and use the factors to predict some target variables, e.g., the industrial production index. Their empirical analysis on U.S. macroeconomic series confirms the usefulness of the PCR method. In chemometrics, partial least squares (PLS) of Wold (1975) has emerged as a leading forecasting method, especially when the number of predictors exceeds the number of data points; see Frank and Friedman (1993). In using many predictors, one often encounters the difficulty of multicollinearity. The ridge regression of Hoerl and Kennard (1970) is an effective method to overcome this difficulty.

The literature also contains many other forecasting methods, ranging from univariate time series model to adaptive forecasts to combining forecasts. Some of these methods are extremely simple to use, e.g., the arithmetic mean of forecasts produced by simple linear regressions. In some cases, vector time-series models or macroeconomic models are used. Furthermore, there are many Bayesian methods that take into account either parameter uncertainty or model uncertainty or both. Thus, forecasting methods are diverse and there is little consensus about the performance of various methods in a given application.

The goal of this paper is to compare various forecasting methods for the special case of a scalar dependent variable and many predictors (or explanatory variables). Specifically, we consider the problem of forecasting U.S. monthly industrial production index using 141 economic variables as predictors. We use multi-step ahead out-of-sample forecasts to evaluate the accuracy of various forecasting methods based on the criterion of root mean squares of forecast error (RMSFE). Our objective is to shed new lights in the performance of some commonly used forecasting methods, to compare their relative accuracy, and to discuss practical implications of the empirical results.

The forecasting methods employed in our study include (a) principal component regression, (b) partial least squares, (c) ridge regression (RR), (d) univariate autoregression (UAR), (e) combining forecast (COMB) based on simple linear regressions, (f) adaptive forecast (AF) using lagged values of the dependent variable as predictors, and (g) combining forecasts based on time-series regression models. We also extend the predictors to include some lagged values of the dependent variable for PCR, PLS, and RR. However, we do not cover all possible forecasting methods because of the computational constraints. For example, Bayesian model averaging is not used because it requires intensive computation due to the existence of too many possible models. Similarly, we do not use methods that require variable selection. For 141 predictors, there are 2^{141} all possible linear regressions. It is not a simple task even if one uses the stepwise procedure to select the variables.

In addition, we use 130 forecasting origins.

The paper is organized as follows. In Section 2, we briefly review the forecasting methods used. We pay more attention to the PLS method because it is less known in the economic literature. Section 3 describes the forecasting procedure used in the paper. Out-of-sample forecasts for one-step to twelve-step ahead are used in our evaluation of the forecasting performance. Section 4 gives the data used whereas Section 5 shows the empirical results. Section 6 concludes.

2 Forecasting Methods

In this section we briefly review the forecasting methods used. Our goal is to predict a dependent variable of interest using many predictors when the sample size is not large compared with the number of predictors. This is the situation in which the ordinary least squares method is often not applicable.

Let y_t be the dependent variable and $\boldsymbol{x}_t = (x_{1t}, \cdots, x_{kt})'$ be the vector of predictors, where k is the number of predictors. Let T be the sample size, i.e. the data are $\{y_t, \boldsymbol{x}_t\}_{t=1}^{T}$. For our empirical study, we use out-of-sample forecasts and let n be the first forecast origin. That is, the initial estimation subsample is $\{y_t, \boldsymbol{x}_t\}_{t=1}^{n}$ and the forecasting subsample is $\{y_t, \boldsymbol{x}_t\}_{t=n+1}^{T}$.

2.1 Principal component regression

PCR is arguably the best known statistical method commonly used to reduce the dimension in a linear framework. It is one of the effective methods for handling multicollinearity in regression analysis. The method is concerned with the variance-covariance structure of the predictors with the goal of using a few linear combinations of the predictors to explain the covariance structure. See, for instance, Johnson and Wichern (2002, chap. 8) for a detailed description of the method. In a series of articles, Stock and Watson (1989, 1991, 1999, 2002a, 2002b, 2004) apply the method with many predictors to obtain a few "factors" or "diffusion indexes" for forecasting. The dependent variable is then projected onto the space spanned by the factors to generate forecasts.

To avoid possible scaling effects in applying PCR, one often standardizes the predictors in the estimation subsample. That is, the predictors satisfy $\sum_{t=1}^{n} x_{it} = 0$ and $\sum_{i=1}^{n} x_{it}^2 = 1$ for $i = 1, \cdots, k$. Let $\boldsymbol{S}_n = \sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'$ be the sample variance-covariance matrix of the predictors in the estimation subsample. Since $\boldsymbol{S}_n$ is positive-definite almost surely, there exists an orthonormal matrix $\boldsymbol{P}$ such that

$$\boldsymbol{S}_n = \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}',$$

where $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \cdots, \lambda_k\}$ is a diagonal matrix of eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$, and the orthonormal matrix $\boldsymbol{P} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k]$ with $\boldsymbol{e}_i$ being the eigenvector associated with the eigenvalue λ_i, i.e. $\boldsymbol{S}_n \boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i$. Let $z_{it} = \boldsymbol{e}_i' \boldsymbol{x}_t$ be the ith principal component of $\boldsymbol{x}_t$. PCR is concerned with using the multiple linear regression

$$y_t = \beta_0 + \sum_{i=1}^{p} \boldsymbol{\beta}_i z_{it} + \epsilon_t,$$

to predict y for some selected number of components p. In practice, one can also adjust the mean of y_t in the estimation subsample so that β_0 can be omitted from the prior regression.

While PCR is popular and appealing, it has a drawback in the forecasting application. The factors are formed by spectral decomposition of the covariance matrix of predictors or, equivalently, the factors are obtained by maximizing the variation among linear combinations of predictors. Eigenvectors corresponding to large eigenvalues are kept whereas those associated with small eigenvalues are discarded. No information of the target (dependent) variable is used in the decomposition. Thus, the retained factors might not have any predictive power of the dependent variable whereas the discarded factors might be useful. On the other hand, discarding components with $\lambda_i \approx 0$ seems appropriate because the component is close to being a constant. Obviously, PCR would be useful when the number of non-zero eigenvalues is small, i.e., $p << k$.

2.2 Partial least squares

PLS is originally proposed by Herman Wold (1966, 1975) as an econometric technique but its most avid proponents are chemical engineers and chemometricians. For example, chemometricians often use spectrographs to analyze the composition of a chemical sample. Hundreds of highly correlated factors are used for measurement of a few objects. The number of predictors is often much greater than the sample size and the ordinary least squares (OLS) method breaks down. Unlike OLS or PCR, PLS does not utilize the covariance among predictors and, hence, is immune from the problem of curse of dimensionality. PLS first projects all predictors into a few components, using the correlation information between the dependent variable and predictors, and then uses those components for forecasting.

PLS constructs components sequentially as follows. The first PLS component z_{1t} is defined as

$$z_{1t} \propto \sum_{i=1}^{k} \mathrm{Cov}(y_t, x_{it}) x_{it},$$

where, again, the predictors x_{it} are standardized in the estimation subsample. Next, perform simple linear regressions of y_t and $\{x_{it}\}$ on z_{1t}, and let the residuals be $y_{1,t}$ and $x_{1,it}$, respectively. The second PLS component z_{2t} is defined as

$$z_{2t} \propto \sum_{i=1}^{k} \mathrm{Cov}(y_{1,t}, x_{1,it}) x_{1,it}.$$

The above procedure is iterated to construct the necessary PLS components. Let p be the number of components used. The PLS forecasting method uses the linear regression

$$y_t = \sum_{i=1}^{p} \beta_i z_{it} + e_t,$$

to perform forecasts, where y_t is also mean-adjusted in the estimation sample. Since z_{it}'s are linear functions of x_{jt}, y_t is a linear function of the original predictors. This linear function is used to produce forecasts. In other words, there is no need to form the factor z_{it} in prediction.

The success of PLS in practical applications has led some statisticians to study its statistical properties and its relationship with other dimension reduction methods. Relevant literature is voluminous and has been accumulating at a fast rate. For example, *Chemometrics and Intelligent Laboratory Systems*, a journal published by Elsevier, has devoted a special issue (vol 58, 2001) to PLS. A comprehensive survey of PLS is beyond the scope of this paper and only a brief review is presented here.

Helland (1990) proposes the concept of relevant components in regression and examines the conditions when PLS and PCR are equivalent. Naes and Helland (1993) generalize the concept of relevancy and propose a new restricted principal component regression. Helland and Almoy (1994) compare PLS with PCR when only a few components are relevant. Almoy (1996) conducts an extensive simulation analysis of the predictive power of various methods when only a few components are relevant. Garthwaite (1994) gives a simple and excellent interpretation of partial least squares from the view point of combining forecasts produced by univariate regressions. Stone and Brook (1990) provide a unified continuum regression framework on which PLS and PCR can be characterized by an index taking different values. Frank and Friedman (1993) give an insightful account of the relationship between PCR, PLS, ridge regression, and variable selection methods. They analyze how these methods balance the bias and variance and why PLS tends to work well in more symmetric design but performs similarly to PCR in empirical applications.

Similar to multivariate time series analysis, PLS has been generalized to forecasting multiple response variables. Instead of forecasting one at a time, statisticians have developed several algorithms to forecast all variables at once, which is commonly referred to as the PLS2 method; see Wold, Eriksson, Trygg, and Kettaneh (2004) for some discussions. Nonlinear PLS has also been suggested. Naik and Tsai (2000) propose a single-index model where PLS is first used to combine many variables into one single index which is then fed into a nonparametric framework. Simulation results confirm the superiority of their proposed method over the sliced inverse regression.

Factor selection has always been a problematic issue in applying PLS or PCR. Cross-validation is commonly adopted to select the number of components though Hwang and Nettleton (2003) propose a data-chosen component method to minimize the MSE of the regression coefficients. Similar to model selection in time series analysis, many information criteria are available for component selection, but no single method dominates the others. Instead of selecting some particular components, we compare the forecasting performance of PLS and PCR with different numbers of components. We further divide the whole sample into estimation, training, and evaluation subsamples. By comparing the selected factors in the training and evaluation samples, we examine the usefulness of out-of-sample forecast in choosing the number of components.

2.3 Ridge regression

RR was introduced as a method to stabilize regression estimates in the presence of multicollinearity, that is, when the covariance matrix $\boldsymbol{S}_n$ of the predictors is singular or nearly so; see Hoerl and Kennard (1970). Consider the multiple linear regression

$$y_t = \boldsymbol{\beta}'\boldsymbol{x}_t + \epsilon_t,$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)'$ is the coefficient vector. The RR estimates are defined as

$$\hat{\boldsymbol{\beta}}_\lambda = \text{argmin}_{\boldsymbol{\beta}} \left[\sum_{t=1}^{n} (y_t - \boldsymbol{\beta}'\boldsymbol{x}_t)^2 + \lambda \boldsymbol{\beta}'\boldsymbol{\beta} \right].$$

Thus, the RR estimates are the solution of a penalized least squares criterion with the penalty being proportional to the squared norm of the coefficient vector $\boldsymbol{\beta}$. It turns out that the solution is

$$\hat{\boldsymbol{\beta}}_\lambda = \left[\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' + \lambda \boldsymbol{I}_k \right]^{-1} \left(\sum_{t=1}^{n} \boldsymbol{x}_t y_t \right),$$

where $\boldsymbol{I}_k$ is the $k \times k$ identity matrix and $\lambda > 0$, whose value can be selected by many criteria, e.g., cross-validation. The *ridge* parameter λ regulates the degree of stabilization. A value of $\lambda = 0$ gives rise to the OLS estimates whereas a value of $\lambda = \infty$ results in using the sample mean of y_t in the estimation sample.

2.4 Univariate time series

In our empirical study, the dependent variable y_t is a univariate time series. One can simply use univariate time series models to produce forecasts. In this paper, we use the autoregressive (AR) model

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + a_t, \quad a_t \sim_{iid} N(0, \sigma_a^2).$$

The order p is selected using the AIC criterion in the estimation sample. The traditional forecasting procedure is used to produce multi-step ahead forecasts of the fitted AR model. We use an AR(5) model in the empirical study.

2.5 Adaptive forecasts

Adaptive forecast (AF) is less known in the literature; see Tiao and Tsay (1994) for some discussion. It is concerned with model uncertainty in multi-step ahead forecasts. For a given model, the basic idea of AF is to estimate the model parameters by minimizing the sum of squares of multi-step forecast errors. We introduce the method based on our empirical study. Suppose that we are interested in 1-step to h-step ahead forecasts at the forecast origin n. To this end, we construct a new dependent variable

$$y_{h,t} = y_{t+1} + \cdots + y_{t+h}, \quad t = 1, \cdots, T-h; \quad h \geq 1. \tag{1}$$

Denote the vector of predictors by $\boldsymbol{z}_t = (z_{1t}, \cdots, z_{mt})'$. An adaptive forecast approach employs the model

$$y_{h,t} = \boldsymbol{\beta}'\boldsymbol{z}_t + e_t, \quad t = 1, \cdots, n-h,$$

and performs OLS estimation of $\boldsymbol{\beta}$. The fitted model is then used to predict $y_{h,n-h+1}$ using the predictor $\boldsymbol{z}_{n-h+1}$. Note that from the definition in Eq. (1), $y_{h,n-h+1}$ contains the quantity y_{n+1}, which is the first observation of the forecasting sample.

In the empirical study, we use $\boldsymbol{z}_t = (y_t, \cdots, y_{t-4})'$. Obviously, for $h > 1$, this approach is different from the univariate AR(5) model for which the parameters are estimated by minimizing the sum of squares of 1-step ahead forecast errors.

2.6 Combining forecast

Combining forecast represents a special approach to forecasting. Consider the 1-step ahead forecast. For each predictor x_{it}, one employs the simple linear regression

$$y_{t+1} = \beta_{i,0} + \beta_{i,1}x_{it} + e_{i,t+1}, \quad t = 1, \cdots, n-1.$$

Note that the dependent variable is shifted by one to avoid the simultaneity in producing forecasts. At the forecast origin n, the prediction based on the ith predictor is $\hat{y}_{i,n+1} = \hat{\beta}_{i,0} + \hat{\beta}_{i,1}x_{i,n}$, where $\hat{\beta}$ denotes the OLS estimate of β. The overall forecast of y_{n+1} is the arithmetic mean of $\hat{y}_{i,n+1}$, i.e.

$$\hat{y}_{n+1} = \sum_{i=1}^{k} \hat{y}_{i,n+1}/k.$$

Multi-step ahead forecasts can be handled in the same manner using the dependent $y_{h,t}$ defined in Eq. (1).

2.7 Dynamic models

To better use of the available information, one can treat lagged values y_{t-i} as additional predictors. This results in using dynamic models. In our empirical study, we augment $\boldsymbol{x}_t$ by y_t and y_{t-1} in PCR, PLS, and RR to obtain dynamic models. For the combining forecasts, the multiple linear regression

$$y_{t+1} = \beta_{i,0} + \beta_{i,1}x_{it} + \beta_{i,2}y_t + \beta_{i,3}y_{t-1} + e_{i,t+1},$$

is used for each predictor x_{it}. Thus, combining forecasts under dynamic models are based on time-series regression models.

3 Forecasting Procedure

We are interested in 1-step to 12-step ahead forecasts. For h-step ahead forecasts, we follow Stock and Watson (2004) and construct the dependent variable

$$Y_{h,t} = \frac{1200[\ln(Y_{t+h}) - \ln(Y_t)]}{h} = \frac{1200(y_{t+h} + \cdots + y_{t+1})}{h}, \tag{2}$$

where y_t is the log return of the U.S. monthly industrial production (IP) index, that is, $y_t = \ln(Y_t) - \ln(Y_{t-1})$ with Y_t being the monthly industrial production index. From the definition, the $Y_{h,t}$ series denotes the annualized growth rate of the IP index in percentage.

For a given h ($1 \leq h \leq 12$), we construct $Y_{h,t}$ for $t = t_0, \cdots, T-h$, where $t_0 = 1$ if no lagged values of y_t are used as predictors and $t_0 = p+1$ otherwise, where p is the maximum AR lag used in the predictors. For PCR, PLS and RR, we standardize the predictors $\boldsymbol{x}_t$ in the estimation subsample. For example, at the starting forecast origin n, the predictors are standardized as

$$x_{it} = \frac{X_{it} - \bar{X}_{i,n}}{s_{i,n}}, \quad t = t_0, \cdots, T-h$$

where X_{it} is the ith observed predictor, and

$$\bar{X}_{i,n} = \frac{1}{n-h-t_0+1} \sum_{t=t_0}^{n-h} X_{it}, \quad s_{i,n} = \sqrt{\sum_{t=t_0}^{n-h} (X_{it} - \bar{X}_i^n)^2/(n-h-t_0)}.$$

Note that, from the definition in Eq. (2), the effective sample size is $n-h-t_0+1$ in the estimation. For PCR and PLS, we do not fit a constant term in the model but remove the mean value of $Y_{h,t}$ in the estimation subsample. This can easily be done in Eq. (2) by replacing y_{t+j} with $y^*_{t+j} = y_{t+j} - \bar{y}_n$, where $\bar{y}_n = \sum_{t=t_0}^{n} y_t/(n-t_0+1)$. On the other hand, for the RR method, we add a constant term to the model.

For each forecasting method, we estimate the model

$$Y_{h,t} = \boldsymbol{\beta}'\boldsymbol{z}_t + e_t, \quad t = t_0, \cdots, n-h,$$

where $\boldsymbol{z}_t$ is the predictor vector constructed from $\boldsymbol{x}_t$. The fitted model is used to produce the forecast $\hat{Y}_{h,n-h+1}$. The estimation subsample is then augmented by the new data point $(Y_{h,n-h+1}, \boldsymbol{z}_{n-h+1})$. We repeat the estimation and forecasting procedure throughout the forecasting subsample. The criterion used to measure forecasting accuracy is

$$\text{RMSFE} = \sqrt{\sum_{t=n-h+1}^{T-h} (Y_{h,t} - \hat{Y}_{h,t})^2/(T-n)}.$$

Thus, our forecasting procedure is an iterated one consisting of estimation-forecasting throughout the forecasting subsample.

The two authors have developed computer programs separately for each forecasting method and compared the results among different programs. The results obtained are close to each other. This step is taken to ensure that the results reported in the paper are reliable.

4 Data

The data used are taken from Macrellino, Stock and Watson (2004) that in turn are compiled from DRI Basic Economic datasets. The original dataset consists of 171 major monthly U.S. macroeconomic time series with time span 1959:1-2002:12. A few modifications are made. As we do not handle imbalanced panel in this paper, we discard those series which cannot be updated to 2002:12. Series computed by Macrellino, Stock and Watson (2004) are excluded[3]. The M2 money stock in 1996 price is updated from 2002:2 to 2002:12. Finally, since several series are unavailable before 1967:1, we set this date as the starting date for the whole sample series. As a result, we have 142 series starting from 1967:1 to 2002:12 with effective sample size being 430 (after 2nd differencing of some variables). We use 141 variables to forecast the industrial production index. The dataset consists of five categories of time series:

1. Income, output, sales, and capacity utilization (24 series);
2. Employment and unemployment (27 series);
3. Construction, inventories, and orders (23 series);
4. Interest rates and asset prices (33 series);
5. Nominal prices, wages and money (35 series).

We adopt the approach of Macrellino, Stock and Watson (2004) by taking log transformation of quantities, indexes and prices and using differencing to achieve stationarity when necessary. For details, see the data appendix of Macrellino, Stock and Watson (2004).

It is worthy mentioning that this paper aims at forecasting the transformed series rather than the original series for two reasons. First, it is well known that the optimal forecast of a nonlinear function of a series is generally not equivalent to nonlinear function of the optimal forecast of the original series. See Lin and Granger (1994) for details. Second, the existing theoretical analysis about PCR and PLS all assume stationarity and it is unclear if these methods are applicable to nonstationary series.

Figure 1 shows time plots of the dependent variable $Y_{h,t}$ for $h = 1$, 6, and 12. The original series has 430 observations. From the plots, the dependent $Y_{h,t}$ becomes smoother as h increases. This is understandable because of the averaging effect. Figure 2 gives time plots of the first three principal component series. Here all 430 observations are used in the principal component analysis. Figure 3 shows time plots of the first three partial least squares series. The difference between principal components and partial least squares components is evident.

[3]Excluded series are, in their notations, msmq, ipd, ipn, ipmin, iput, wtq, wtdq, msdq, msmtq, msnq, wtnq, rtdrq, rtnrq, ivmtq, ivmfgq, ivmfdq, ivmfnq, ivwrq, ivrrq,ivsrq, ivsrmq, ivsrwq, ivsrrq, moq, mdoq, muq, mduq.

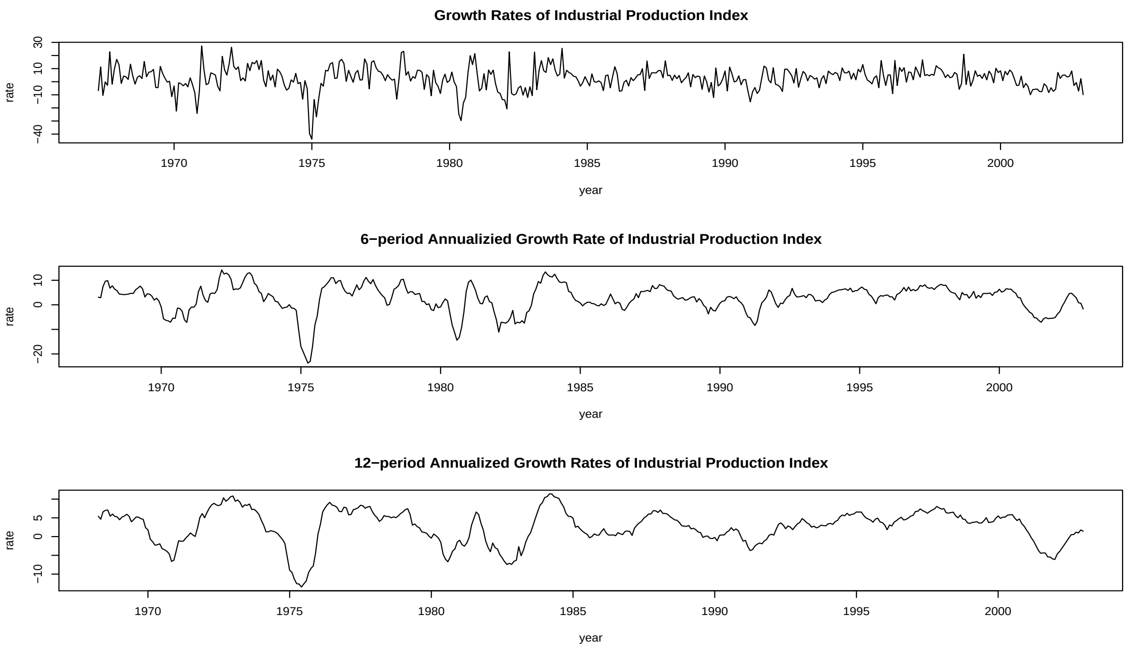

Figure 1: Time plots of dependent variables: (a) monthly growth rate, (b) average of 6 month growth rate, (c) average of 12 month growth rate. All series are annualized.

5 Empirical results

Tables 1 summarizes the root mean squares of forecast errors of the forecasting methods employed in the paper. The methods include adaptive forecast, univariate autoregression, combining forecast both with and without lagged values of the industrial production index, partial least squares with 1 to 6 components, principal component regression with 1 to 20 components, and ridge regression with ridge parameters $\{0,10,\cdots,90\}$. The table contains results for 1-step to 12-step ahead out-of-sample forecasts, using the forecasting procedure of Section 3. For the univariate autoregression, we use an AR(5) model. For the adaptive forecast, the predictors used are $(y_t, \cdots, y_{t-4})$. For combining forecasts of time-series regression models, we use y_t and y_{t-1} and x_{it} as the predictors for y_{t+1}. Table 2 gives the ratio of RMSFE of various forecasting methods with respect to the adaptive forecasting method. The ranking of the forecasting methods in term of RMSFE is reported in Table 3.

From the tables, we make the following observations:

1. The multiple linear regression using all predictors, shown under ridge regression with zero ridge parameter, gives unstable numerical results as the 141 predictors are highly correlated and the covariance matrix is nearly singular. Simple OLS regression using matrix inverse algorithm is terminated for singularity while regression using singular value decomposition

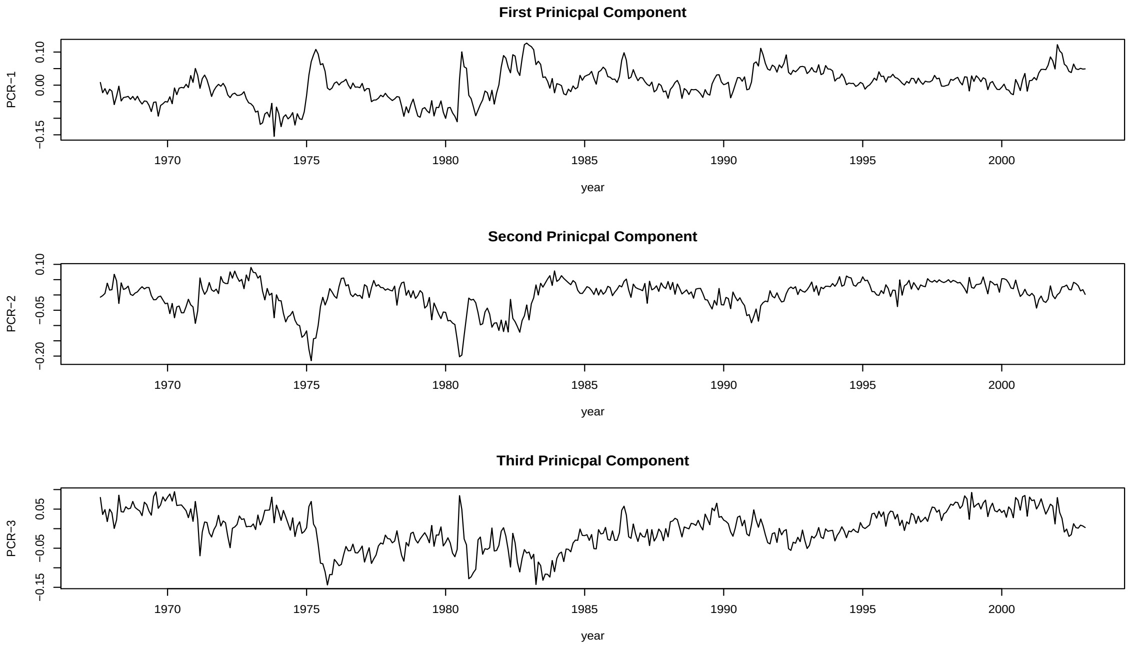

Figure 2: Time plots of selected principal component series: (a) the first component, (b) the second component, (c) the third component. The analysis is based on all 430 observations.

method does produce some RMSFEs. As is shown later, adding two lagged dependent variables as predictors inflates the RMSFE by more than 10 times. For completeness, we report those figures in the tables but exclude those from the ranking comparison. This simple exercise confirms the importance of using parsimonious models in out-of-sample prediction.

2. The univariate autoregression performs well. It produces forecasts slightly inferior to those of adaptive forecast. The forecasting inefficiency starts with 1.1% for the 2-step ahead forecast and gradually increases to 6.8% for 12-step ahead forecast. This might indicate the possibility of minor model misspecification in using an AR(5) model for the industrial production index.

3. The combining forecast of simple linear regressions does a decent job, but seems to be dominated by the adaptive forecast.

4. The combining forecast of time-series regression models perform very well. It produces forecasts comparable with those of adaptive forecast.

5. For partial least squares, the model with 3 components performs best. In fact, it outperforms not only the PCR but also all other methods for 1-step to 6-step ahead forecasts. For 7-step to 12-step ahead forecasts, the PLS with three components is either close or better than the adaptive forecast. However, PLS fares poorly when the number of components used increases.

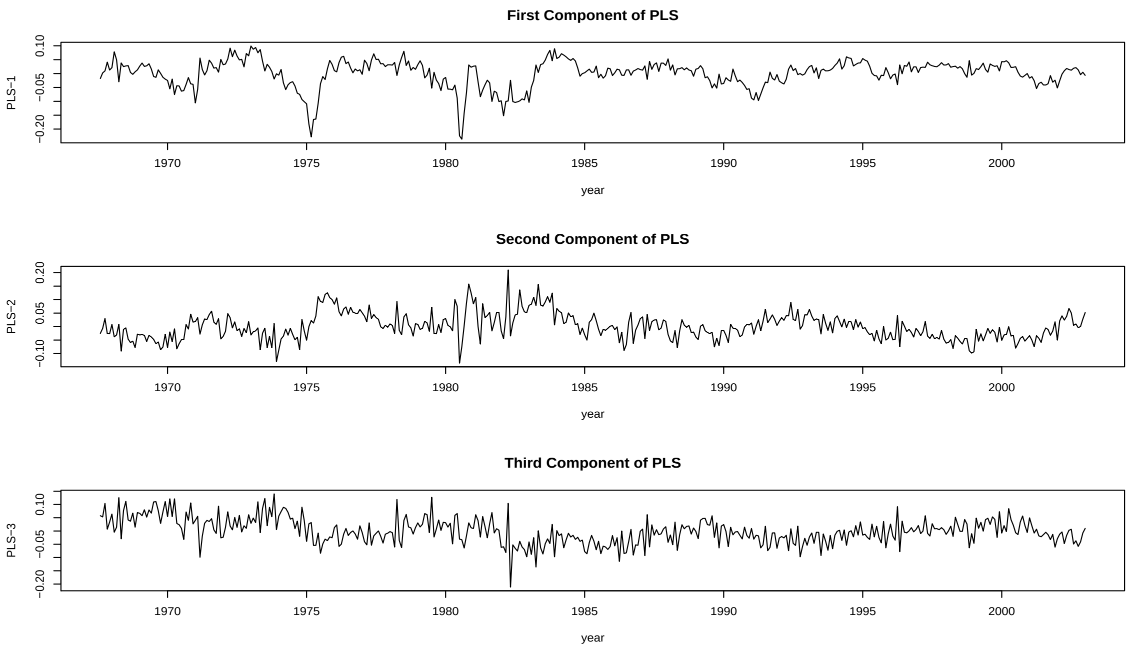

Figure 3: Time plots of selected partial least squares components: (a) the first component, (b) the second component, (c) the third component. The analysis is based on all 430 observations.

6. The principal component regression does not provide good forecasts when the number of components is small. The method performs well when the number of components is between 3 and 5, especially for longer forecasting horizon. The performance of PCR seems relatively stable when the number of components is between 6 and 20.

7. The ridge regression does not fare well for 1-step ahead forecasts, but performs well when the ridge parameter is sufficiently large. This is particularly so for the case of longer forecasting horizon. In fact, ridge regression gives the most accurate 9- to 12-step ahead forecasts. This is true for a wide range of ridge parameters. Note that the maximum ridge parameter used in 90, which seems too small for the 1-step ahead forecasts.

Next, we augmented y_t and y_{t-1} as two additional predictors and repeated the out-of-sample forecast procedure for PCR, PLS and RR. The results are shown in Tables 4 and 5. From the tables, the two added predictors do not significantly change the forecast results.

Finally, we examine the usefulness of out-of-sample forecasts in determining the number of factors for PLS and PCR. We divide the 430 observations into several subsamples with the first 200 observations as the estimation subsample, observations between 201 and 300 as the training subsample, and observations between 300 to 430 as the evaluation subsample. The 1-step to 12-step RMSFEs with various numbers of factor for the training and evaluation subsamples are computed and compared. The results are shown in Figures 4 and 5. Ideally, the two graphs should have a common minimum for the RMSFE to select a stable number of factors. The selected number of factors for two sub-samples are given in Table 6.

From the figures and the table, we observe that for both PLS and PCR, the training and validation subsamples show a similar pattern of RMSFE but the two subsamples do not generally give the same optimal number of factors. The discrepancy is at most one for the PLS method but could go as high as 30 for PCR. Further, RMSFE of PLS increases rapidly as number of factors becomes large whereas there exists some flat region for PCR. To summarize, RMSFE is helpful in selecting the number of factor in using PLS and PCR methods, but there are rooms for improvement. Also, it is more difficult to select the number of factors for PCR than for PLS.

6 Conclusion

In this paper, we used out-of-sample multi-step ahead forecasts to compare the performance of several forecasting methods when there is a scalar dependent variable and many predictors. Our empirical study shows that the partial least squares method with 3 components outperforms other forecasting methods in 1-step to 6-step ahead forecasts. The method also performs adequately for longer horizon forecasts. Thus, our study confirms the usefulness of forming the factor using the cross covariance between the target variable and predictors rather than using the covariance of predictors. However, the performance of PLS depends critically on the number of factors, and it deteriorates quickly when the number of components increases. The principal component regression is less sensitive to the number of components used, but it requires more components than the PLS method to produce accurate forecasts. When the number of components is sufficient, the performance of PCR appears to be stable. However, we expect the performance of the PCR method to deteriorate when the number of components used is sufficiently large. The simple idea of ridge regression works nicely when the forecast horizon is long. However, ridge regression fares poorly for 1-step ahead forecasts. The combining forecast of time-series regression models performs well. Finally, the adaptive forecast outperforms univariate forecasts for multi-step ahead forecasts even though the margin of improvement is not large.

Table 1: Root Mean Squares of Forecast Errors of Selected Forecasting Methods, where h denotes the number of steps.

h	Adap	UAR	Comb	CoAR	PLS : 1 to 6 components					
1	5.673	5.673	5.762	5.330	5.848	5.498	5.320	5.577	5.813	6.132
2	3.979	4.023	4.396	3.970	4.498	3.964	3.681	3.977	4.295	4.609
3	3.562	3.618	4.023	3.681	4.094	3.755	3.407	3.683	4.105	4.457
4	3.430	3.527	3.816	3.504	3.862	3.657	3.279	3.550	3.969	4.208
5	3.285	3.437	3.650	3.337	3.674	3.557	3.197	3.526	3.927	4.015
6	3.116	3.296	3.512	3.223	3.521	3.427	3.095	3.448	3.840	3.918
7	3.092	3.300	3.437	3.148	3.434	3.368	3.094	3.455	3.739	3.762
8	3.056	3.287	3.364	3.082	3.380	3.279	3.053	3.431	3.656	3.645
9	3.017	3.252	3.301	3.048	3.421	3.232	3.031	3.388	3.539	3.494
10	3.025	3.254	3.254	3.029	3.596	3.194	2.957	3.284	3.382	3.371
11	3.034	3.251	3.210	3.020	3.535	3.152	2.942	3.256	3.312	3.331
12	3.031	3.236	3.163	3.020	3.575	3.115	2.926	3.230	3.266	3.332
PCR: 1 to 10 components										
1	5.758	5.626	5.471	5.506	5.536	5.626	5.483	5.481	5.440	5.570
2	4.439	4.157	3.829	3.842	3.882	4.136	3.884	3.870	3.914	4.009
3	4.140	3.949	3.590	3.504	3.546	3.883	3.620	3.587	3.678	3.801
4	3.930	3.856	3.496	3.371	3.405	3.672	3.489	3.433	3.525	3.665
5	3.725	3.730	3.388	3.252	3.298	3.593	3.438	3.376	3.496	3.652
6	3.592	3.590	3.257	3.118	3.149	3.459	3.350	3.269	3.392	3.576
7	3.527	3.512	3.192	3.083	3.102	3.422	3.350	3.267	3.396	3.558
8	3.465	3.421	3.099	3.007	3.020	3.336	3.272	3.184	3.333	3.492
9	3.419	3.373	3.040	2.970	2.983	3.300	3.244	3.158	3.284	3.435
10	3.386	3.354	3.010	2.943	2.960	3.269	3.187	3.086	3.203	3.361
11	3.349	3.311	2.969	2.915	2.931	3.224	3.145	3.054	3.172	3.325
12	3.317	3.272	2.931	2.884	2.903	3.223	3.141	3.047	3.165	3.315
PCR: 11 to 20 components										
1	5.569	5.467	5.349	5.408	5.433	5.470	5.497	5.479	5.443	5.380
2	4.010	3.998	3.932	3.953	3.969	4.031	3.925	3.868	3.892	3.841
3	3.807	3.795	3.721	3.736	3.777	3.824	3.730	3.630	3.702	3.638
4	3.682	3.673	3.616	3.633	3.665	3.684	3.600	3.517	3.549	3.515
5	3.670	3.658	3.581	3.589	3.609	3.630	3.556	3.489	3.539	3.504
6	3.605	3.593	3.521	3.527	3.548	3.551	3.480	3.417	3.463	3.400
7	3.584	3.571	3.502	3.513	3.527	3.526	3.483	3.427	3.463	3.426
8	3.517	3.490	3.431	3.448	3.477	3.468	3.435	3.400	3.429	3.418
9	3.461	3.449	3.391	3.407	3.418	3.401	3.388	3.365	3.389	3.393
10	3.385	3.385	3.329	3.335	3.339	3.307	3.309	3.282	3.286	3.277
11	3.344	3.345	3.290	3.300	3.311	3.275	3.281	3.256	3.249	3.242
12	3.324	3.316	3.270	3.276	3.293	3.248	3.257	3.229	3.222	3.214
Ridge regression: weight (0-90)										
1	7.308	6.246	6.131	6.046	5.979	5.923	5.874	5.833	5.796	5.764
2	5.261	4.379	4.266	4.192	4.140	4.100	4.070	4.045	4.024	4.007
3	5.027	4.288	4.161	4.072	4.007	3.957	3.916	3.882	3.854	3.830
4	4.668	3.937	3.843	3.781	3.735	3.700	3.672	3.649	3.630	3.613
5	4.220	3.725	3.646	3.593	3.553	3.523	3.500	3.481	3.465	3.452
6	3.973	3.552	3.480	3.434	3.402	3.379	3.361	3.346	3.335	3.326
7	3.473	3.286	3.241	3.216	3.201	3.192	3.185	3.181	3.178	3.177
8	3.047	3.158	3.127	3.110	3.101	3.095	3.092	3.091	3.090	3.090
9	2.724	2.951	2.944	2.946	2.950	2.956	2.962	2.968	2.974	2.979
10	2.562	2.825	2.818	2.824	2.832	2.841	2.850	2.858	2.866	2.873
11	2.450	2.753	2.757	2.769	2.782	2.793	2.805	2.815	2.824	2.832
12	2.169	2.685	2.714	2.737	2.756	2.771	2.785	2.796	2.806	2.815

Table 2: Ratios of RMSFE of Selected Forecasting Methods with that of Adaptive Forecast

h	Adap	UAR	Comb	CoAR	PLS : 1 to 6 components					
1	5.673	1.000	1.016	0.940	1.031	0.969	0.938	0.983	1.025	1.081
2	3.979	1.011	1.105	0.998	1.130	0.996	0.925	1.000	1.079	1.159
3	3.562	1.016	1.130	1.034	1.150	1.054	0.957	1.034	1.153	1.251
4	3.430	1.028	1.113	1.021	1.126	1.066	0.956	1.035	1.157	1.227
5	3.285	1.046	1.111	1.016	1.118	1.083	0.973	1.073	1.195	1.222
6	3.116	1.058	1.127	1.034	1.130	1.100	0.993	1.106	1.232	1.258
7	3.092	1.067	1.111	1.018	1.110	1.089	1.001	1.117	1.209	1.217
8	3.056	1.076	1.101	1.009	1.106	1.073	0.999	1.123	1.196	1.193
9	3.017	1.078	1.094	1.011	1.134	1.071	1.005	1.123	1.173	1.158
10	3.025	1.076	1.076	1.001	1.189	1.056	0.978	1.085	1.118	1.114
11	3.034	1.072	1.058	0.996	1.165	1.039	0.970	1.073	1.092	1.098
12	3.031	1.068	1.044	0.996	1.180	1.028	0.965	1.066	1.078	1.099
PCR: 1 to 10 components										
1	1.015	0.992	0.964	0.970	0.976	0.992	0.966	0.966	0.959	0.982
2	1.116	1.045	0.962	0.966	0.976	1.040	0.976	0.973	0.984	1.008
3	1.162	1.109	1.008	0.984	0.996	1.090	1.016	1.007	1.033	1.067
4	1.146	1.124	1.019	0.983	0.993	1.071	1.017	1.001	1.027	1.068
5	1.134	1.135	1.031	0.990	1.004	1.094	1.046	1.027	1.064	1.112
6	1.153	1.152	1.045	1.001	1.010	1.110	1.075	1.049	1.089	1.148
7	1.141	1.136	1.032	0.997	1.003	1.107	1.083	1.057	1.098	1.151
8	1.134	1.120	1.014	0.984	0.988	1.092	1.071	1.042	1.091	1.143
9	1.133	1.118	1.008	0.985	0.989	1.094	1.075	1.047	1.089	1.139
10	1.119	1.108	0.995	0.973	0.978	1.081	1.054	1.020	1.059	1.111
11	1.104	1.091	0.979	0.961	0.966	1.063	1.037	1.007	1.046	1.096
12	1.094	1.080	0.967	0.951	0.958	1.063	1.037	1.005	1.044	1.094
PCR: 11 to 20 components										
1	0.982	0.964	0.943	0.953	0.958	0.964	0.969	0.966	0.959	0.948
2	1.008	1.005	0.988	0.994	0.998	1.013	0.986	0.972	0.978	0.965
3	1.069	1.066	1.045	1.049	1.061	1.074	1.047	1.019	1.039	1.022
4	1.074	1.071	1.054	1.059	1.068	1.074	1.050	1.025	1.035	1.025
5	1.117	1.114	1.090	1.092	1.098	1.105	1.082	1.062	1.077	1.066
6	1.157	1.153	1.130	1.132	1.139	1.140	1.117	1.097	1.112	1.091
7	1.159	1.155	1.132	1.136	1.141	1.140	1.126	1.108	1.120	1.108
8	1.151	1.142	1.123	1.128	1.138	1.135	1.124	1.113	1.122	1.119
9	1.147	1.143	1.124	1.129	1.133	1.127	1.123	1.116	1.124	1.125
10	1.119	1.119	1.100	1.102	1.104	1.093	1.094	1.085	1.086	1.083
11	1.102	1.103	1.085	1.088	1.091	1.079	1.081	1.073	1.071	1.069
12	1.097	1.094	1.079	1.081	1.087	1.072	1.075	1.065	1.063	1.061
Ridge regression: weight (0-90)										
1	1.288	1.101	1.081	1.066	1.054	1.044	1.035	1.028	1.022	1.016
2	1.322	1.101	1.072	1.054	1.041	1.031	1.023	1.017	1.011	1.007
3	1.411	1.204	1.168	1.143	1.125	1.111	1.099	1.090	1.082	1.075
4	1.361	1.148	1.120	1.102	1.089	1.079	1.071	1.064	1.058	1.053
5	1.285	1.134	1.110	1.094	1.082	1.072	1.065	1.059	1.055	1.051
6	1.275	1.140	1.117	1.102	1.092	1.084	1.079	1.074	1.070	1.067
7	1.123	1.063	1.048	1.040	1.035	1.032	1.030	1.029	1.028	1.027
8	0.997	1.034	1.023	1.018	1.015	1.013	1.012	1.011	1.011	1.011
9	0.903	0.978	0.976	0.977	0.978	0.980	0.982	0.984	0.986	0.988
10	0.847	0.934	0.932	0.933	0.936	0.939	0.942	0.945	0.947	0.950
11	0.808	0.907	0.909	0.913	0.917	0.921	0.924	0.928	0.931	0.934
12	0.716	0.886	0.895	0.903	0.909	0.914	0.919	0.923	0.926	0.929

Table 3: Ranking Forecast Errors of Selected Forecasting Methods.

h	Adap	UAR	Comb	CoAR	PLS : 1 to 6 components					
1	24	25	27	2	32	16	1	21	30	38
2	18	23	36	16	38	14	1	17	34	39
3	4	7	32	12	34	18	1	13	35	39
4	4	12	32	8	35	21	1	14	38	39
5	3	8	30	5	34	22	1	18	38	39
6	2	8	26	5	27	19	1	21	38	39
7	2	17	25	5	24	19	3	26	38	39
8	4	19	22	5	23	18	3	28	39	38
9	12	19	22	15	34	17	13	25	39	38
10	14	20	21	15	39	18	11	25	35	34
11	15	24	20	14	39	18	12	25	33	35
12	15	26	19	14	39	17	12	25	29	38
PCR: 1 to 10 components										
1	26	22	11	17	18	23	14	13	7	20
2	37	31	2	4	7	29	8	6	10	21
3	36	29	6	2	3	27	8	5	11	21
4	36	34	7	2	3	24	6	5	11	22
5	35	37	7	2	4	25	9	6	14	31
6	35	34	6	3	4	22	12	7	15	33
7	33	30	10	1	4	21	18	15	20	35
8	32	26	11	1	2	21	17	16	20	36
9	33	24	14	8	11	21	18	16	20	35
10	38	32	13	10	12	22	17	16	19	33
11	38	31	13	10	11	21	17	16	19	34
12	36	31	13	10	11	23	18	16	20	34
PCR: 11 to 20 components										
1	19	9	3	5	6	10	15	12	8	4
2	22	19	12	13	15	25	11	5	9	3
3	22	20	15	17	19	23	16	9	14	10
4	27	26	17	19	23	28	15	10	13	9
5	33	32	23	24	27	28	21	13	19	16
6	37	36	28	29	30	31	24	18	23	16
7	37	36	29	31	34	32	28	23	27	22
8	37	35	29	31	34	33	30	24	27	25
9	37	36	28	31	32	30	26	23	27	29
10	36	37	29	30	31	27	28	24	26	23
11	36	37	29	30	32	27	28	26	23	22
12	37	35	30	32	33	27	28	24	22	21
Ridge regression: wgt (0-90)										
1		39	37	36	35	34	33	31	29	28
2		35	33	32	30	28	27	26	24	20
3		38	37	33	31	30	28	26	25	24
4		37	33	31	30	29	25	20	18	16
5		36	29	26	20	17	15	12	11	10
6		32	25	20	17	14	13	11	10	9
7		16	14	13	12	11	9	8	7	6
8		15	14	13	12	10	9	8	6	7
9		4	1	2	3	5	6	7	9	10
10		3	1	2	4	5	6	7	8	9
11		1	2	3	4	5	6	7	8	9
12		1	2	3	4	5	6	7	8	9

Table 4: RMSFEs of Some Forecasting Methods With Two Lagged Dependent Variables Added.

h	Adap	UAR	Comb	CoAR	PLS : 1 to 6 components					
1	5.673	5.673	5.762	5.330	5.842	5.489	5.362	5.622	5.912	6.346
2	3.979	4.023	4.396	3.970	4.477	3.906	3.646	3.977	4.345	4.724
3	3.562	3.618	4.023	3.681	4.107	3.724	3.410	3.701	4.164	4.522
4	3.430	3.527	3.816	3.504	3.873	3.636	3.307	3.589	4.039	4.265
5	3.285	3.437	3.650	3.337	3.681	3.524	3.242	3.596	4.011	4.130
6	3.116	3.296	3.512	3.223	3.522	3.385	3.148	3.544	3.910	4.016
7	3.092	3.300	3.437	3.148	3.441	3.341	3.153	3.555	3.799	3.845
8	3.056	3.287	3.364	3.082	3.379	3.247	3.109	3.521	3.709	3.719
9	3.017	3.252	3.301	3.048	3.400	3.189	3.079	3.454	3.565	3.537
10	3.025	3.254	3.254	3.029	3.580	3.171	3.003	3.327	3.372	3.427
11	3.034	3.251	3.210	3.020	3.542	3.131	2.984	3.255	3.286	3.360
12	3.031	3.236	3.163	3.020	3.553	3.097	2.966	3.172	3.296	3.382
PCR: 1 to 10 components										
1	5.540	5.600	5.489	5.532	5.554	5.650	5.471	5.477	5.434	5.572
2	4.020	4.100	3.793	3.796	3.844	4.075	3.839	3.828	3.879	3.972
3	3.805	3.925	3.603	3.483	3.543	3.855	3.622	3.589	3.677	3.796
4	3.651	3.841	3.524	3.365	3.400	3.660	3.508	3.451	3.542	3.671
5	3.489	3.701	3.400	3.243	3.299	3.576	3.447	3.386	3.503	3.662
6	3.375	3.551	3.255	3.099	3.131	3.440	3.359	3.278	3.396	3.587
7	3.338	3.489	3.204	3.083	3.097	3.402	3.356	3.277	3.400	3.573
8	3.299	3.392	3.100	3.002	3.000	3.314	3.272	3.192	3.329	3.507
9	3.267	3.335	3.024	2.951	2.949	3.275	3.241	3.162	3.278	3.448
10	3.251	3.335	3.015	2.940	2.928	3.248	3.192	3.096	3.200	3.377
11	3.236	3.294	2.975	2.914	2.895	3.210	3.155	3.067	3.171	3.342
12	3.225	3.253	2.932	2.878	2.861	3.216	3.161	3.063	3.166	3.332
PCR: 11 to 20 components										
1	5.532	5.457	5.349	5.384	5.434	5.512	5.523	5.487	5.423	5.380
2	3.985	3.953	3.905	3.919	3.941	4.066	3.927	3.846	3.846	3.801
3	3.813	3.795	3.731	3.751	3.768	3.874	3.755	3.630	3.689	3.644
4	3.694	3.691	3.644	3.658	3.671	3.737	3.639	3.530	3.556	3.529
5	3.685	3.678	3.607	3.613	3.620	3.671	3.576	3.485	3.529	3.508
6	3.619	3.612	3.546	3.548	3.556	3.583	3.495	3.410	3.454	3.420
7	3.604	3.595	3.529	3.533	3.541	3.553	3.499	3.420	3.458	3.442
8	3.539	3.513	3.459	3.460	3.481	3.489	3.452	3.394	3.433	3.436
9	3.482	3.468	3.414	3.411	3.424	3.410	3.399	3.352	3.387	3.402
10	3.409	3.408	3.359	3.344	3.361	3.324	3.330	3.278	3.297	3.300
11	3.367	3.366	3.317	3.305	3.329	3.289	3.299	3.251	3.261	3.266
12	3.348	3.340	3.299	3.279	3.314	3.266	3.279	3.225	3.239	3.242
Ridge regression: wgt (0-90)										
1	96.940	6.300	6.190	6.105	6.035	5.976	5.926	5.881	5.842	5.808
2	76.453	4.341	4.244	4.179	4.131	4.095	4.066	4.042	4.022	4.005
3	78.892	4.292	4.161	4.072	4.007	3.957	3.916	3.883	3.855	3.830
4	76.071	3.935	3.839	3.778	3.734	3.701	3.674	3.652	3.634	3.618
5	75.558	3.754	3.677	3.626	3.589	3.561	3.538	3.520	3.505	3.492
6	69.756	3.587	3.515	3.470	3.439	3.417	3.399	3.386	3.374	3.365
7	73.540	3.325	3.281	3.258	3.245	3.237	3.231	3.227	3.224	3.222
8	74.667	3.194	3.169	3.156	3.148	3.144	3.142	3.140	3.139	3.139
9	73.505	2.990	2.985	2.989	2.994	3.000	3.006	3.012	3.017	3.021
10	72.258	2.867	2.863	2.870	2.880	2.889	2.897	2.905	2.912	2.919
11	71.253	2.801	2.807	2.820	2.832	2.844	2.854	2.864	2.872	2.879
12	68.268	2.736	2.768	2.791	2.809	2.823	2.836	2.846	2.855	2.862

Table 5: Ratios of RMSFE w.p.t. Adaptive Forecast When Two Lagged Dependent Variables Are Added.

h	Adap	UAR	Comb	CoAR	PLS : 1 to 6 components					
1	5.673	1.000	1.016	0.940	1.030	0.967	0.945	0.991	1.042	1.119
2	3.979	1.011	1.105	0.998	1.125	0.982	0.916	1.000	1.092	1.187
3	3.562	1.016	1.130	1.034	1.153	1.046	0.957	1.039	1.169	1.270
4	3.430	1.028	1.113	1.021	1.129	1.060	0.964	1.046	1.177	1.243
5	3.285	1.046	1.111	1.016	1.120	1.073	0.987	1.094	1.221	1.257
6	3.116	1.058	1.127	1.034	1.130	1.086	1.010	1.137	1.255	1.289
7	3.092	1.067	1.111	1.018	1.113	1.080	1.020	1.150	1.229	1.243
8	3.056	1.076	1.101	1.009	1.106	1.062	1.017	1.152	1.214	1.217
9	3.017	1.078	1.094	1.011	1.127	1.057	1.021	1.145	1.182	1.172
10	3.025	1.076	1.076	1.001	1.183	1.048	0.993	1.100	1.115	1.133
11	3.034	1.072	1.058	0.996	1.168	1.032	0.983	1.073	1.083	1.107
12	3.031	1.068	1.044	0.996	1.172	1.022	0.979	1.047	1.088	1.116
PCR: 1 to 10 components										
1	0.977	0.987	0.967	0.975	0.979	0.996	0.964	0.965	0.958	0.982
2	1.010	1.031	0.953	0.954	0.966	1.024	0.965	0.962	0.975	0.998
3	1.068	1.102	1.012	0.978	0.995	1.082	1.017	1.008	1.032	1.066
4	1.064	1.120	1.027	0.981	0.991	1.067	1.023	1.006	1.033	1.070
5	1.062	1.127	1.035	0.987	1.004	1.088	1.049	1.031	1.066	1.115
6	1.083	1.140	1.045	0.995	1.005	1.104	1.078	1.052	1.090	1.151
7	1.080	1.128	1.036	0.997	1.001	1.100	1.085	1.060	1.099	1.156
8	1.080	1.110	1.014	0.982	0.982	1.085	1.071	1.045	1.090	1.148
9	1.083	1.106	1.003	0.978	0.978	1.086	1.074	1.048	1.087	1.143
10	1.074	1.102	0.997	0.972	0.968	1.074	1.055	1.023	1.058	1.116
11	1.067	1.086	0.980	0.960	0.954	1.058	1.040	1.011	1.045	1.102
12	1.064	1.073	0.967	0.950	0.944	1.061	1.043	1.011	1.045	1.099
PCR: 11 to 20 components										
1	0.975	0.962	0.943	0.949	0.958	0.972	0.973	0.967	0.956	0.948
2	1.001	0.994	0.981	0.985	0.991	1.022	0.987	0.967	0.967	0.955
3	1.071	1.065	1.048	1.053	1.058	1.088	1.054	1.019	1.036	1.023
4	1.077	1.076	1.062	1.066	1.070	1.090	1.061	1.029	1.037	1.029
5	1.122	1.120	1.098	1.100	1.102	1.117	1.088	1.061	1.074	1.068
6	1.162	1.159	1.138	1.139	1.141	1.150	1.122	1.094	1.108	1.098
7	1.165	1.163	1.141	1.143	1.145	1.149	1.132	1.106	1.118	1.113
8	1.158	1.150	1.132	1.132	1.139	1.142	1.130	1.111	1.123	1.124
9	1.154	1.150	1.132	1.131	1.135	1.131	1.127	1.111	1.123	1.128
10	1.127	1.126	1.110	1.105	1.111	1.099	1.101	1.084	1.090	1.091
11	1.110	1.110	1.093	1.089	1.097	1.084	1.087	1.072	1.075	1.076
12	1.105	1.102	1.088	1.082	1.094	1.078	1.082	1.064	1.069	1.070
Ridge regression: wgt (0-90)										
1	17.087	1.110	1.091	1.076	1.064	1.053	1.044	1.037	1.030	1.024
2	19.216	1.091	1.067	1.050	1.038	1.029	1.022	1.016	1.011	1.007
3	22.151	1.205	1.168	1.143	1.125	1.111	1.100	1.090	1.082	1.075
4	22.176	1.147	1.119	1.101	1.089	1.079	1.071	1.065	1.059	1.055
5	22.999	1.143	1.119	1.104	1.093	1.084	1.077	1.071	1.067	1.063
6	22.387	1.151	1.128	1.114	1.104	1.097	1.091	1.087	1.083	1.080
7	23.783	1.075	1.061	1.054	1.049	1.047	1.045	1.044	1.043	1.042
8	24.436	1.045	1.037	1.033	1.030	1.029	1.028	1.028	1.027	1.027
9	24.367	0.991	0.990	0.991	0.993	0.995	0.997	0.998	1.000	1.002
10	23.884	0.948	0.946	0.949	0.952	0.955	0.958	0.960	0.963	0.965
11	23.487	0.923	0.925	0.929	0.934	0.937	0.941	0.944	0.947	0.949
12	22.525	0.903	0.913	0.921	0.927	0.932	0.936	0.939	0.942	0.944

Table 6: Using out-of-sample forecasts to determine the number of factors in two subsamples.

Subsample	Selected PLS components: 1- to 12- step forecasts											
300-430	3	3	3	3	3	3	3	3	3	3	3	3
200-300	2	3	3	3	4	4	4	4	4	4	4	4
Subsample	Selected PCR components: 1- to 12- step forecasts											
300-430	31	24	4	4	4	4	4	4	4	4	4	4
200-300	1	1	10	1	11	11	11	11	11	11	11	11

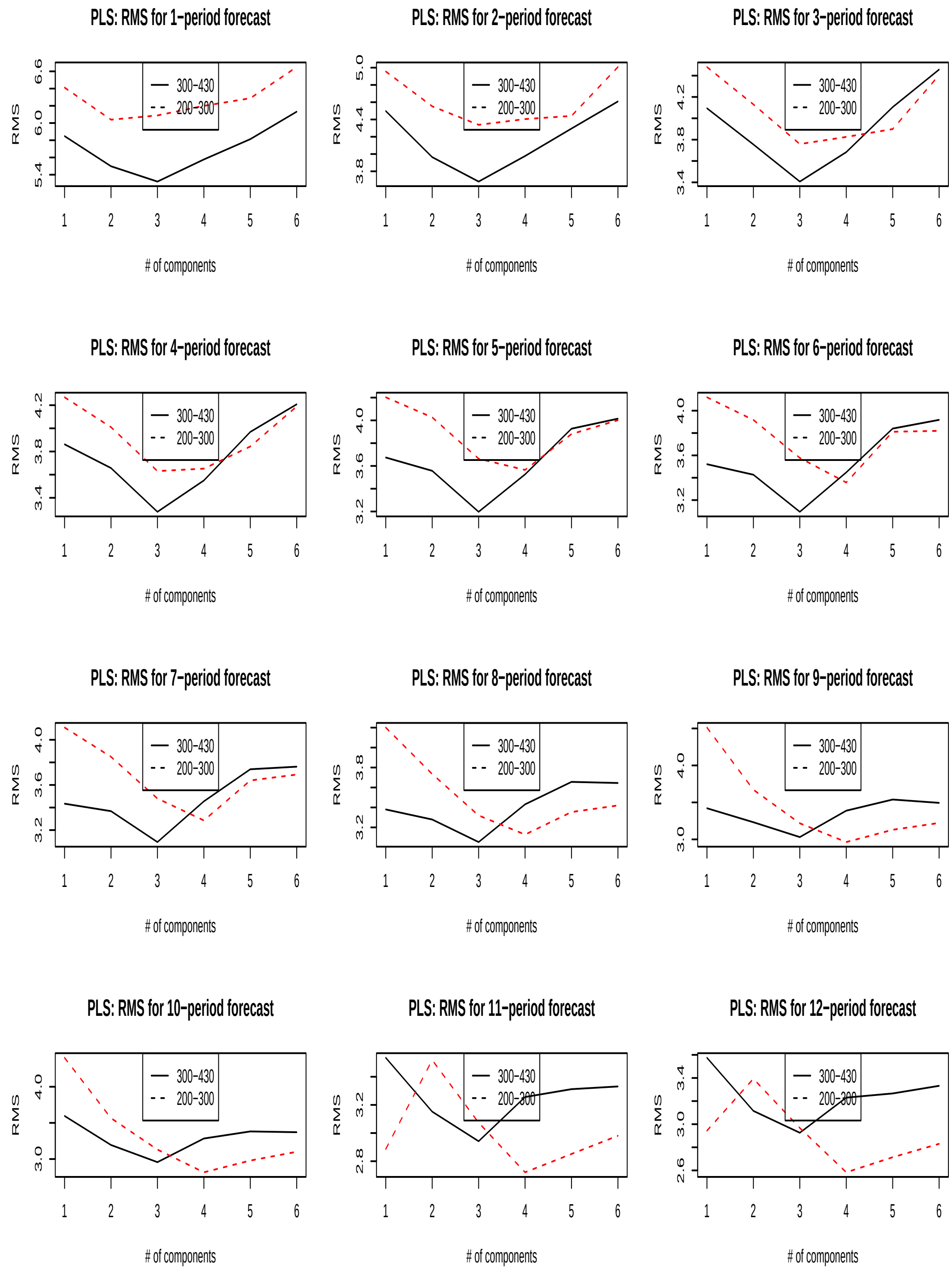

Figure 4: Number of Factors Selected for PLS

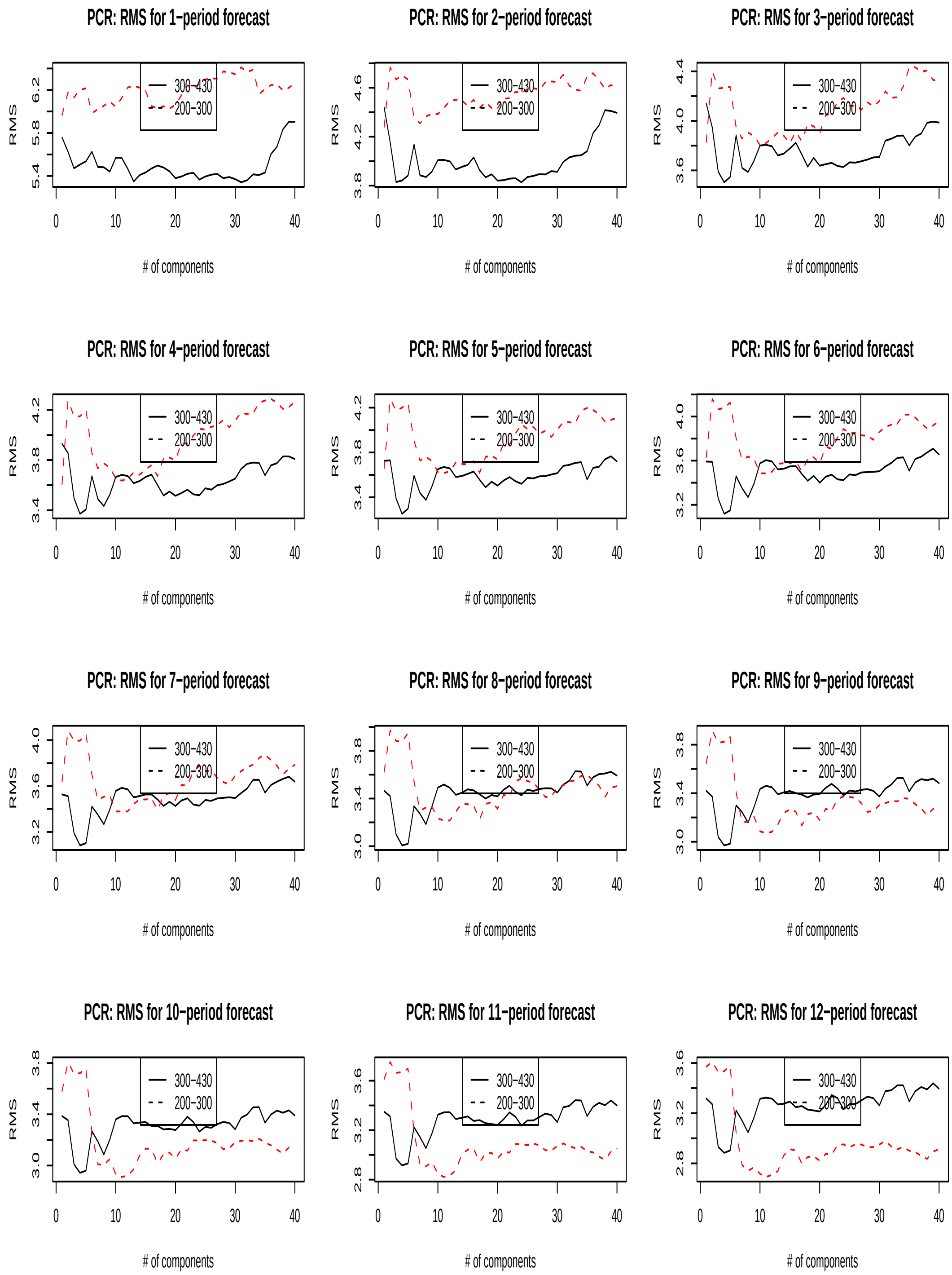

Figure 5: Number of Factors Selected for PCR

References

Almoy, T. (1996), "A simulation study on comparison of prediction methods when only a few components are relevant," *Computational Statistics and Data Analysis*, 21, 87-107.

Frank, I and J. Friedman (1993), "A statistical view of some chemometrics regression tools," *Technometrics*, 35, 109-135.

Garthwaite, P. (1994), "An interpretation of partial least squares", *Journal of the American Statistical Association*, 89, 122-127.

Helland, I. (1990), " Partial least squares regression and statistical models", *Scandinavian Journal of Statistics*, 17, 97-114.

Helland, I and T. Almoy (1994), "Comparison of prediction methods when only a few components are relevant", *Journal of the American Statistical Association*, 89, 583-591.

Hoerl, A. E. and R. W. Kennard (1970), "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, 8, 27-51.

Hwang, J.T. and D. Nettleton (2003), " Principal components regression with data-chosen components and related methods", *Technometrics*, 45, 70-79.

Johnson, R. A. and Wichern, D. W. (2002), *Applied Multivariate Statistical Analysis*, 5th Edition, Prentice-Hall: Upper Saddle River, New Jersey.

Lin, J and C. Granger (1994), "Forecasting from non-linear models in practice," *Journal of Forecasting*, 13, 1-9.

Marcellino, M, J. Stock, and W. Watson (2004), "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," Working paper available at http://www.wws.princeton.edu/m̃watson/wp.html

Naes, T and I. Helland (1993), " Relevant components in regression", *Scandinavian Journal of Statistics* 20, 239-250.

Naik, P and C. Tsai (2000), "Partial least squares estimator for single-index models", *Journal of Royal Statistical Society, B*, 62, 763-771.

Stock, J. and M. Watson (1989), "New indexes of coincident and leading economic indicators," *NBER Macroeconomics Annual*, 351-393.

Stock, J. and M. Watson (1991), "A probability model of the coincident economic indicators," in G. Moore and K. Lahiri, eds. *The Leading Economic Indicators: New Approaches and Forecasting Methods*, Cambridge University Press, 63-90.

Stock, J. and M. Watson (1999), "Forecasting inflation," *Journal of Monetary Economics*, 44, 293-335.

Stock, J. and M. Watson (2002a), "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, 20, 147-162.

Stock, J. and M. Watson (2002b), "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167-1179.

Stock, J. and M. Watson (2004), "Forecasting with many predictors," forthcoming *Handbook of Economic Forecasting*.

Stone, M. and R. Brooks, (1990), "Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression", *Journal of Royal Statistical Society, B*, 52, 237-269.

Tiao, G. C. and Tsay, R. S. (1994) "Some advances in nonlinear and adaptive modeling in time series analysis" (with discussion), *Journal of Forecasting*, 13, 109-140.

Wold, H (1966), "Estimation of principal components and related models by iterative least squares," in P. Krishnaiah ed. *Multivariate Analysis*, New York: Academic Press. 391-420.

Wold, H. (1975), "Soft modeling by latent variables: the nonlinear iterative partial least squares approach," in *Perspectives in Probability and Statistics*, Papers in Honor of M.S. Bartlett, ed. J. Gani, London: Academic Press.

Wold, S, L. Eriksson, J. Trygg and N. Kettaneh (2004), "The PLS method-partial least squares projections to latent structures and its applications in industrial RDP," presented at the COMPSTAT 2004, 16th Symposium of IASC, PRAGUE, August 23-27, 2004.