

Forecasting linear dynamical systems using subspace methods

Alfredo García-Hiernaux*

October 26, 2010

Abstract

This paper presents a novel approach to predict with subspace methods. It consists in combining multiple forecasts obtained from setting a range of values for a specific parameter that is typically fixed by the user in this literature. Two procedures are proposed. The first one combines all the forecast in a particular range. The second one predicts with a restricted number of combinations previously optimized. Both methods are evaluated using Monte Carlo experiments and by forecasting the German gross domestic product.

Keywords: Forecasting, subspace methods, combining forecasts.

JEL Classification: C53,C22,E27

*Department of Quantitative Economics, Universidad Complutense de Madrid, email: agarciah@ccee.ucm.es, tel: (+34) 91 394 25 11, fax: (+34) 91 394 25 91. The author is greatly indebted to Miguel Jerez for several helpful comments and gratefully acknowledges financial support from Ministerio de Educación y Ciencia, ref. ECO2008-02588/ECON.

1 Introduction

Subspace methods are widely employed in engineering and physics and have been recently adapted to some characteristics of the economic and financial data (see, Bauer and Wagner, 2002; Bauer, 2005b; García-Hiernaux et al., 2009, 2010). In comparison with mainstream time series analysis (Box and Jenkins, 1976; Tiao and Tsay, 1989), they are flexible, as univariate and multivariate cases are treated in the same way, and fast, as iterations are not required. Consequently, they are a very interesting alternative to conventional forecasting tools such as VAR models.

Despite the extensive literature about their statistical properties (see, *e.g.*, Bauer, 2005a,b) and their increasing empirical uses (Kapetanios, 2004; Kascha and Mertens, 2009), forecasting with subspace methods still remains quite unexplored. The scarce references (Ljung, 1999; Mossberg, 2007; Schumacher, 2007) just use a state-space model estimated with these techniques to extrapolate, without exploiting the subspace properties to improve the forecasts.

In contrast, this paper explores the forecasting in- and out-of-sample properties of subspace methods and suggests two procedures based on combining multiple forecasts, obtained from setting a range of values for a specific parameter that is typically fixed by the user in the subspace literature. The first one combines a range of forecasts obtained with common subspace methods. The second one optimizes the number of forecasts to combine using the AIC (Akaike, 1976). The procedures are compared against appropriate alternatives and tested with simulated and real data, showing good results in one- and multi-step-ahead out-of-sample forecasts.

The plan of the paper is as follows. Subspace identification techniques are described in Section 2. Section 3 presents two procedures to improve the forecasts obtained through subspace methods. The usefulness of the proposals for making high quality forecasts is illustrated through Monte Carlo experiments in Section 4 and with real data in Section 5. Finally, Section 6 provides some concluding remarks.

2 Preliminaries

Consider a linear fixed-coefficients system that can be described by the following state space model,

$$\mathbf{x}_{t+1} = \mathbf{\Phi}\mathbf{x}_t + \mathbf{E}\boldsymbol{\psi}_t \quad (1a)$$

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \boldsymbol{\psi}_t \quad (1b)$$

where $\mathbf{x}_{t,t \in \mathbb{N}}$ is a state n -vector, being n the true order of the system. In addition, $\mathbf{z}_{t,t \in \mathbb{N}}$ is an observable output m -vector, $\boldsymbol{\psi}_{t,t \in \mathbb{N}}$ is a noise m -vector known as innovations, while $\mathbf{\Phi}$, \mathbf{E} and \mathbf{H} are parameter matrices. Moreover, some assumptions about the system and the noise must be established.

Assumptions:

A1. $\boldsymbol{\psi}_t$ is a sequence of independent and identically distributed random variables with $E(\boldsymbol{\psi}_t) = 0$ and $E(\boldsymbol{\psi}_t\boldsymbol{\psi}_t') = \mathbf{Q}$, being \mathbf{Q} a positive definite matrix.

A2. System (1a-1b) is non-explosive, i.e., $|\lambda_i(\mathbf{\Phi})| \leq 1, \forall i = 1, \dots, n$, where $\lambda_i(\mathbf{\Phi})$ denotes the i^{th} eigenvalue of $\mathbf{\Phi}$, and fulfills the strictly minimum-phase condition, i.e., $|\lambda_i(\mathbf{\Phi} - \mathbf{E}\mathbf{H})| < 1, \forall i = 1, \dots, n$.

A3. System (1a-1b) is minimal, i.e., it uses the smallest possible state dimension, n , to represent \mathbf{z}_t .

System (1a-1b) can be expressed in a single matrix equation (see, García-Hiernaux et al., 2010, Section 2) as:

$$\mathbf{Z}_f = \mathbf{O}\mathbf{X}_f + \mathbf{V}\boldsymbol{\Psi}_f, \quad (2)$$

where $\mathbf{Z}_f := [\mathbf{z}'_t, \dots, \mathbf{z}'_{t+f-1}]$ with $t = p+1, \dots, T-f+1$; p and f are two integers chosen by the user with $p > n$. For simplicity, we will assume $p = f$, throughout the paper, denoting this value by i . \mathbf{X}_f and $\boldsymbol{\Psi}_f$ are as \mathbf{Z}_f but with \mathbf{x}_t or $\boldsymbol{\psi}_t$, respectively, instead of \mathbf{z}_t . On the other hand, the extended observability matrix, \mathbf{O} , and the lower block triangular Toeplitz matrix \mathbf{V} are known nonlinear func-

tions of the original parameter matrices Φ , E and H (see García-Hiernaux et al., 2010, Section 2, for further details).

Given A2 and for large values of i and T , \mathbf{X}_f is to a close approximation representable as a linear combination of the past of the output, $\mathbf{M}\mathbf{Z}_p$, where $\mathbf{Z}_p := [\mathbf{z}'_{t-p}, \dots, \mathbf{z}'_{t-1}]$ with $t = p + 1, \dots, T - f + 1$. Then, the relationship between the past and the future of the output can be written as:

$$\mathbf{Z}_f \simeq \mathbf{O}\mathbf{M}\mathbf{Z}_p + \mathbf{V}\Psi_f \quad (3)$$

For a given n , subspace methods estimate \mathbf{O} , \mathbf{M} and \mathbf{V} in (3) by solving a reduced-rank weighted least square problem, as the product $\mathbf{O}\mathbf{M}$, which is an im square matrix, is of rank $n < im$. There are different approaches to do this, but equation (3) is the common starting point to all of them. Here we use the Canonical Correlation Analysis (CCA), which is briefly described in the following steps:

1. Choose the integer i (or p and f).
2. Solve the reduced-rank weighted least square problem:

$$\min_{\{\hat{\mathbf{O}}, \hat{\mathbf{M}}\}} \left\| \mathbf{W}(\mathbf{Z}_f - \hat{\mathbf{O}}\hat{\mathbf{M}}\mathbf{Z}_p) \right\|_F^2 \quad (4)$$

where $\|\cdot\|_F$ denotes de Frobenius norm. Note that the weighting matrix, \mathbf{W} , and the system order, n , have to be specified. See Katayama (2005) for different \mathbf{W} and García-Hiernaux et al. (2007) to estimate n . Compute the states as $\hat{\mathbf{X}}_f = \hat{\mathbf{M}}\mathbf{Z}_p$.

3. Regress \mathbf{z}_t onto $\hat{\mathbf{x}}_t$, $t = i, \dots, T - i$, obtaining $\hat{\mathbf{H}}$ and the residuals, $\hat{\psi}_t$, from the equation (1b).
4. Regress $\hat{\mathbf{x}}_{t+1}$ onto $\hat{\mathbf{x}}_t$ and $\hat{\psi}_t$, $t = i, \dots, T - i - 1$, obtaining $\hat{\Phi}$ and \hat{E} from the equation (1a).

5. Check the minimum-phase condition (A2). If A2 does not hold, a refactorization is needed to ensure it (see, Hannan and Deistler, 1988, Theorem 1.3.3).

3 Forecasting by exploiting different values of i

I will start by denoting the parameters estimated using the CCA algorithm and a particular i as, $\hat{\Xi}_i = \{\hat{\Phi}, \hat{E}, \hat{H}\}_i$. In this situation, it is known that for $i \geq i_0$ the estimates $\hat{\Xi}_i$ are consistent, where $i_0 = \text{int}(d\hat{\rho}_{bic})$ which is the integer closer to the product of d and the optimal lag length for an autoregressive approximation of \mathbf{z}_t , chosen by using the Schwarz (1978) criterion over $0 \leq \rho \leq (\log T)^a$ for some constant $0 < a < \infty$. Specifically, $d > 1$ is a sufficient condition in the stationary case (see Deistler et al., 1995), whereas $d > 2$ is required in the integrated case (see Bauer, 2005b). However, in finite samples the estimates $\hat{\Xi}_i$ differ for different i , resulting in distinct forecasts.

Ljung (1999) proposes choosing the value for i that optimizes the AIC. The procedure consists of: a) choosing a range of possible values for i , b) estimating the corresponding state space models, c) calculating, for each estimated model, the AIC, and d) choosing the model which minimizes the criterion.

The procedures proposed here also emphasize in forecasting, but adopt a combination strategy motivated by the empirical success of combination forecasts. This approach seems more promising, if only because the combination strategy allows one to diversify the risk of a potentially erroneous decision about i . Therefore, I expect the procedures to be more robust than the ones relying on a single choice for i .

Whichever procedure you choose to forecast, it should be noted that the results about consistency restrict the lower bound of the range of possible values for i to

i_0 . Then, consider the $I - i_0 + 1$ estimated models:

$$\hat{\mathbf{x}}_t^i = \hat{\mathbf{\Phi}}_i \hat{\mathbf{x}}_{t-1}^i + \hat{\mathbf{E}}_i \hat{\boldsymbol{\psi}}_{t-1} \quad (5a)$$

$$\hat{\mathbf{z}}_t^i = \hat{\mathbf{H}}_i \hat{\mathbf{x}}_t^i \quad (5b)$$

where $i = i_0, \dots, I$, being I deterministically chosen by the user and $\hat{\boldsymbol{\psi}}_t = \mathbf{z}_t^i - \hat{\mathbf{z}}_t^i$. Clearly, $\hat{\mathbf{z}}_t^i$ are highly correlated for different i and, as a consequence of consistency, the correlations will tend to the unity as the sample size grows. In very large samples, $\hat{\mathbf{z}}_t^i$ will be virtually identical for different i . Accordingly, the benefits for combining are expected to be more important in short samples.

Consider now a vector, \mathbf{z}_t^s , that contains the fitted values $\hat{\mathbf{z}}_t^i$, $i = i_0, \dots, I$, but sorted in a particular way that will be explained later, and a matrix $\mathbf{\Pi} = [\boldsymbol{\pi}_0 \ \boldsymbol{\pi}_1 \ \dots \ \boldsymbol{\pi}_{I-i_0+1}]'$ where each $\boldsymbol{\pi}_j$ is a m -vector of weights. From all of this, one can solve the least squares problem:

$$\min_{\{\hat{\mathbf{\Pi}}\}} \left\| \mathbf{z}_t - [\mathbf{1} \ \mathbf{z}_t^s] \cdot \hat{\mathbf{\Pi}} \right\|_F^2 \quad (6)$$

where $\mathbf{1}$ is a ones m -vector. As a result, $\hat{\mathbf{z}}_t^* = [\mathbf{1} \ \mathbf{z}_t^s] \cdot \hat{\mathbf{\Pi}}$ is the optimal linear prediction of \mathbf{z}_t (see Granger and Ramanathan, 1984) given the range of i . This procedure, hereafter PROC A, presents lower in-sample mean squared error than any subspace forecast obtained with a fixed value of i in the range $\{i_0, I\}$. Note that this does not guarantee more accurate out-of-sample predictions, although it could be expected in practice.

On the other hand, choosing a large I makes the information given by the set of explanatory variables extremely redundant, due to the high correlations among $\hat{\mathbf{z}}_t^i$, $i = i_0, \dots, I$. In order to reduce the number of inputs in regression (6), we suggest a second procedure, hereafter PROC B, which consists in sequentially increasing the dimension of \mathbf{z}_t^s and using the AIC to optimize the number of inputs to combine.

I will now motivate why \mathbf{z}_t^s has a specific structure. Vector \mathbf{z}_t^s is organized so that the first component is the \hat{z}_t^i which presents a lower correlation with the others, the second element is the second less correlated and so on. As an example, for $i_0 = 5$, $I = 7$ and the following correlations $\text{corr}(\hat{z}_t^5, \hat{z}_t^6) = .8$, $\text{corr}(\hat{z}_t^5, \hat{z}_t^7) = .6$, and $\text{corr}(\hat{z}_t^6, \hat{z}_t^7) = .7$, $\hat{\mathbf{z}}_t^s$ would be the vector $[\hat{z}_t^7 \ \hat{z}_t^5 \ \hat{z}_t^6]$. Hence, the reduction of the sum-squared-error of regression (6) will be, in principle, higher when adding the first \mathbf{z}_t^s components than when adding the last ones, as, by construction, most of the information brought by the last variables will already be in the model. In the following, I describe the proposals that compute the final out-of-sample forecasts:

1. Find i_0 as the integer closer to $d\hat{\rho}_{bic}$ and choose I .
2. Estimate $\hat{\Xi}_i$ for $i = i_0, \dots, I$ and compute the corresponding in- and out-of-sample forecasts.
3. Create \mathbf{z}_t^s with the in-sample forecasts obtained in step 2, sorted from the least correlated to the most correlated.
4. Regress \mathbf{z}_t onto $[\mathbf{1} \ \mathbf{z}_t^s]$, either: (i) once, this is PROC A, or (ii) $I - i_0 + 1$ times, increasing \mathbf{z}_t^s by one component each time and calculating the AIC in each regression, this is PROC B. Whatever method is used, keep the weights $\hat{\Pi}$.
5. Compute the combined out-of-sample forecasts as $\hat{\mathbf{z}}_{t+f}^* = [\mathbf{1} \ \mathbf{z}_{t+f}^s] \cdot \hat{\Pi}$, where f is the prediction horizon. In PROC B, the number of columns of \mathbf{z}_{t+f}^s will be determined by minimizing AIC in the previous step.

4 An empirical application

This section illustrates the application of this methodology to real data by forecasting the German GDP growth rate in a one-step and multi-step ahead framework. The forecasts are evaluated in terms of RMSFE and predictive accuracy is tested with the Diebold and Mariano (1995) test, hereafter DM.

The data employed corresponds to the quarterly German GDP in constant prices of year 2000. The sample period goes from 1991:01 until 2008:03. The exercise is divided in two parts. First, a one-step-ahead forecast evaluation is made over the period 2006:02 to 2008:03, updating the models each time with the new data. Second, a multi-step prediction analysis is presented by fitting the models for the period of 1991:01 to 2006:01 and forecasting ten periods, from 2006:02 to 2008:03.

As a result of the autoregressive approximation of German GDP $\hat{\rho}_{bic} = 8$, so i_0 is fixed to 11, assuring the consistency of the estimates. As the sample size is not large, I is fixed to $I = 20$. Consequently, $I - i_0 + 1 = 10$ models are estimated and used in the prediction exercise. Estimating the system order with the *MbC* criterion (García-Hiernaux et al., 2007) returns $\hat{n} = 7$. Finally, an AR(8) is specified and estimated to provide benchmark forecasts for comparison purposes.

Table 1 presents the RMSFEs, forecast accuracy ranking and results of the DM for the one-step-ahead predictions obtained from: a) PROC A, the combination of the forecasts of the whole vector \mathbf{z}_t^s , b) PROC B, the combination of the forecasts using the AIC to decrease the dimension of \mathbf{z}_t^s , c) the alternative AR(8) model, and d) the subspace single forecasts obtained with $i = 11, \dots, 20$.

[TABLE 1 AND FIGURE 1 SHOULD BE AROUND HERE]

The results show that PROC A clearly outperforms the rest of the methods. The gain in RMSFE with respect to PROC B is 9% and the DM suggests that PROC A forecasts are statistically more accurate at 12% level. This result coincides with those obtained in the simulation experiments. On the other hand, the improvement of the RMSFE with respect the rest of the (non-combined) subspace models ranges from 6-22%, depending on the choice of i . The DM considers all these predictions significantly less precise than those resulting from PROC A at about 10% level, except for the non-combined subspace forecast with $i = 11$.

PROC A particularly outperforms the AR(8), which is 26% worse in terms of RMSFE. This improvement is significant at 2% level in terms of the predictive accuracy. The performances can be observed in Figure 1, which depicts the prediction errors for the combined procedures, the best subspace single forecast and the AR(8).

[TABLE 2 AND FIGURE 2 SHOULD BE AROUND HERE]

In a second exercise, ten out-of-sample forecasts are computed from 2006:02 to 2008:03. The models are the same used before, although this time data and estimates are not updated. Table 2 reports the results. In this case, PROC B clearly beats the other alternatives. The gain in RMSFE with respect to PROC A is 27.7% and the predictions are statistically more accurate at 1% level. PROC A worse behavior is not completely unexpected in multi-step forecasting, as it has been devised to minimize one-step-ahead errors. More surprising is the positive performance of PROC B. The precision improvement with respect to the other (non-combined) subspace models is quite remarkable, ranging from 16.5-58.2% in terms of RMSFE. Further, DM considers all the predictions significantly less precise at 1% level than those obtained with PROC B, except for $i = 13, 14$, which can be considered less accurate at 10% level. The alternative AR(8), whose RMSFE is 2.1 and 1.6 times those of PROC B and PROC A, respectively, is widely defeated by both proposals. Finally, despite the positive results of PROC B in this exercise, further research is needed in the multi-step forecast to draw more general conclusions.

5 Concluding remarks

I propose two procedures to forecast linear dynamical systems using subspace methods. They are based on combining multiple predictions obtained from setting a range of values for a parameter that is commonly fixed by the user in the subspace methods literature. The experiments with Monte Carlo and real data show

that they generally outperform the (non-combined) subspace methods and (vector) autoregressive models in one-step and multi-step ahead predictions, although further research is open for future in the second case.

These algorithms are implemented in a MATLAB toolbox for time series modeling called E^4 . The source code is freely provided under the terms of the GNU General Public License and can be downloaded at www.ucm.es/info/icae/e4. This site also includes a complete user manual and other reference materials.

References

- Akaike, H. (1976). *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*. Academic Press.
- Bauer, D. (2005a). Comparing the CCA subspace method to pseudo maximum likelihood methods in the case of no exogenous inputs. *Journal of Time Series Analysis*, 26(5):631–668.
- Bauer, D. (2005b). Estimating linear dynamical systems using subspace methods. *Econometric Theory*, 21:181–211.
- Bauer, D. and Wagner, M. (2002). Estimating cointegrated systems using subspace algorithms. *Journal of Econometrics*, 111:47–84.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 2nd ed. edition.
- Deistler, M., Peterzell, K., and Scherrer, W. (1995). Consistency and relative efficiency of subspace methods. *Automatica*, 31(12):1865–1875.
- Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.
- García-Hiernaux, A., Jerez, M., and Casals, J. (2007). Estimating the system order by subspace methods. *Statistics and Econometrics Working Papers*, 070301.

- García-Hiernaux, A., Jerez, M., and Casals, J. (2009). Fast estimation methods for time series models in state-space form. *Journal of Statistical Computation and Simulation*, 79(2):121–134.
- García-Hiernaux, A., Jerez, M., and Casals, J. (2010). Unit roots and cointegration modeling through a family of flexible information criteria. *Journal of Statistical Computation and Simulation*, 80(2):173–189.
- Granger, C. W. J. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3:197–204.
- Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. John Wiley, New York.
- Kapetanios, G. (2004). A note on modelling core inflation for the uk using a new dynamic factor estimation method and a large disaggregated price index dataset. *Economic Letters*, 85:63–69.
- Kascha, C. and Mertens, K. (2009). Business cycle analysis and VARMA models. *Journal of Economics Dynamics and Control*, 33(2):267–282.
- Katayama, T. (2005). *Subspace Methods for System Identification*. Springer Verlag, London.
- Ljung, L. (1999). *System Identification, Theory for the User*. PTR Prentice Hall, New Jersey, second edition.
- Mossberg, M. (2007). Forecasting electric power consumption using subspace algorithms. *International Journal of Power and Energy Systems*, 27:369–386.
- Schneider, T. and Neumaier, A. (2001). Algorithm 808: Arfit - a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27:58–65.
- Schumacher, C. (2007). Forecasting german gdp using alternative factor models based on large datasets. *Journal of Forecasting*, 26:271–302.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society, B Series*, 51(2):157–213.

Tables and Figures

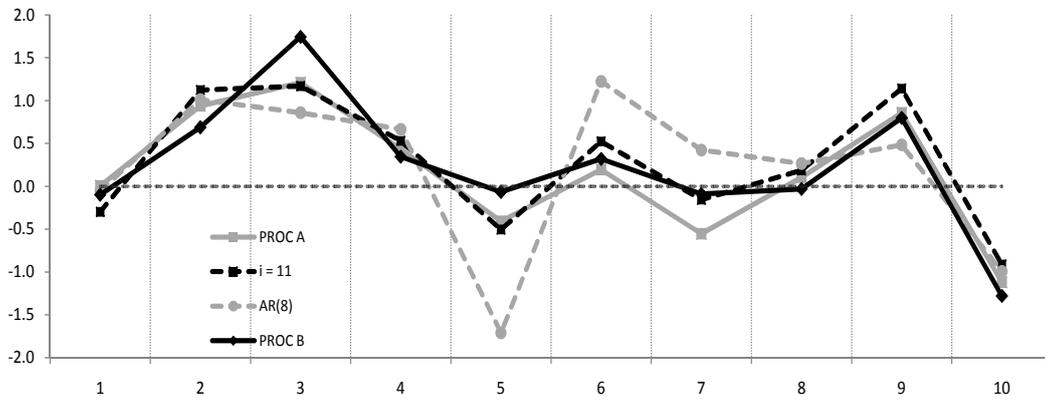


Figure 1: One-step-ahead out-of-sample forecast errors.

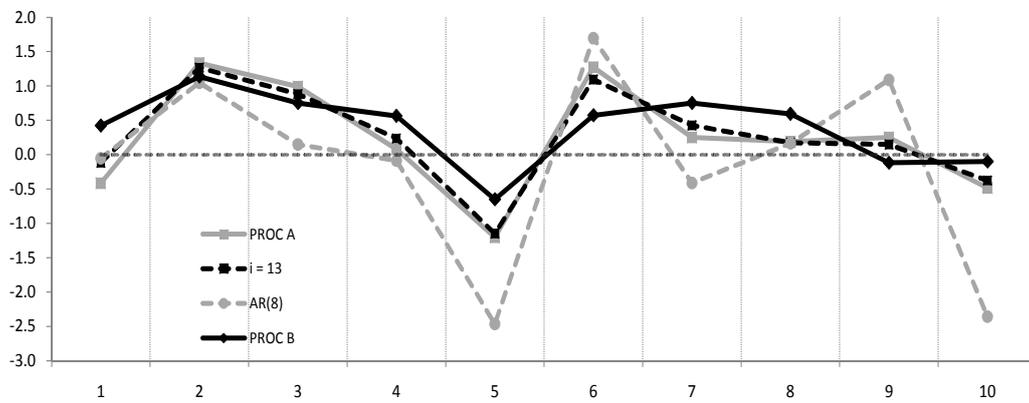


Figure 2: 1-to-10 out-of-sample forecast errors.

Table 1: Forecast accuracy comparison: Evaluation of the one-step-ahead prediction errors.

Procedure	RMSFE			Diebold-Mariano	
	Value	Relative	Rk	Rk(1) vs Rk(j) Statistic	Pvalue
PROC A	<u>.746</u>	<u>100</u>	<u>1</u>	-	-
PROC B	.815	109.0	4	-1.208	.114
AR(8)	.945	126.7	13	-2.082	.019
SM (11)	.791	106.0	2	-1.221	.111
SM (12)	.855	114.6	9	-2.726	.003
SM (13)	.832	111.4	6	-3.117	.001
SM (14)	.805	107.8	3	-1.252	.105
SM (15)	.827	110.8	5	-1.577	.057
SM (16)	.858	115.0	10	-1.295	.098
SM (17)	.842	112.8	8	-1.391	.082
SM (18)	.858	115.0	11	-2.144	.016
SM (19)	.911	122.1	12	-2.775	.018
SM (20)	.837	112.1	7	-1.358	.087

Notes: The best RMSFE is underlined. SM (i) corresponds to the forecasts obtained from non-combined subspace methods estimated with i . Prediction errors are multiplied by 100 in order to facilitate the comparison. Diebold and Mariano's test computed with a squared error loss. Hypothesis defined as $H_0 : E[(\epsilon_{t+1|t}^1)^2] \geq E[(\epsilon_{t+1|t}^j)^2]$ and $H_1 : E[(\epsilon_{t+1|t}^1)^2] < E[(\epsilon_{t+1|t}^j)^2]$, where $\epsilon_{t+1|t}^j$ is the one-step-ahead forecast error obtained from the model ranked in position j .

Table 2: Forecast accuracy comparison: Evaluation of 1-to-10 prediction errors.

Procedure	RMSFE			Diebold-Mariano	
	Value	Relative	Rk	Rk(1) vs Rk(j) Statistic	Pvalue
PROC A	.840	127.7	10	-4.840	.000
PROC B	<u>.657</u>	<u>100</u>	<u>1</u>	-	-
AR(8)	1.375	209.1	13	-4.520	.000
SM (11)	.820	124.7	4	-3.596	.000
SM (12)	1.040	158.2	12	-3.094	.001
SM (13)	.766	116.5	2	-1.325	.092
SM (14)	.780	118.7	3	-1.454	.073
SM (15)	.876	133.3	11	-2.230	.003
SM (16)	.830	126.3	8	-4.023	.000
SM (17)	.835	127.0	9	-8.857	.000
SM (18)	.822	125.1	6	-4.549	.000
SM (19)	.820	124.7	5	-3.436	.009
SM (20)	.827	125.8	7	-2.376	.009

Notes: The best RMSFE is underlined. SM (i) corresponds to the forecasts obtained from non-combined subspace methods estimated with i . Prediction errors are multiplied by 100 in order to facilitate the comparison. Diebold and Mariano's test computed with a squared error loss. Hypothesis defined as $H_0 : E[(\epsilon_{t+k|t}^1)^2] \geq E[(\epsilon_{t+k|t}^j)^2]$ and $H_1 : E[(\epsilon_{t+k|t}^1)^2] < E[(\epsilon_{t+k|t}^j)^2]$, where $\epsilon_{t+k|t}^j$ is the one-step-ahead forecast error obtained from the model ranked in position j and $k = 1, 2, \dots, 10$.