

Forecasting linear dynamical systems using subspace methods

Alfredo García-Hiernaux

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 gratefully acknowledges financial support from Ministerio de Ed-
ucación y Ciencia, ref. ECO2008-02588/ECON.

Forecasting linear dynamical systems using subspace methods

Abstract

A new procedure to predict with subspace methods is presented in this paper. It is 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 methods literature. An algorithm to compute these predictions and to obtain a suitable number of combinations is provided. The procedure is illustrated by forecasting the German gross domestic product.

Keywords: Forecasting, subspace methods, combining forecasts.

JEL Classification: C53,C22,E27

1 Introduction

Since the seminal work of Ho and Kalman (1966) system identification has concentrated in modeling a data set using a state-space representation with no a priori restrictions. From the 90s, system identification has been led by new techniques known as subspace methods. These algorithms have a wide use in fields like engineering and physics and have been recently adapted to the particular characteristics of the economic and financial data, see, for instance, Bauer and Wagner (2002); Bauer (2005); García-Hiernaux et al. (2009a,b).

In comparison with the common time series analysis, see Box and Jenkins (1976) for the univariate case or Tiao and Tsay (1989) for multiple series, the main advantages of these procedures are: a) the univariate and multivariate cases are treated in the same way, b) they allow one to specify a general linear model directly from the data, without a priori knowledge of the process structure, c)

they are based on robust and computationally efficient algebraic tools and, consequently, d) iterations are not required, avoiding convergence problems.

However, despite the extensive literature about the statistic properties of these procedures and its increasing use with different purposes (Kapetanios, 2004; Kascha and Mertens, 2009), the matter of forecasting with subspace methods still remains quite unexplored. The scarce references, as Mossberg (2007), just use the state-space model estimated with these techniques to extrapolate, but do not exploit the subspace properties in order to improve the forecasts.

This paper explores the forecasting in- and out-of-sample properties of the subspace methods and suggests a procedure 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 methods literature. The proposal is compared against alternatives and tested with real data, finding good results in one-step-ahead and mid-term out-of-sample forecasts.

The plan of the paper is as follows. Subspace identification techniques are described in Section 2. A procedure that improves the in-sample forecasts obtained through subspace methods is presented in Section 3. The usefulness of the proposal for making high quality forecasts is illustrated in Section 4 with the German Gross Domestic Product (GDP). Finally, some concluding remarks are given in Section 5.

2 Model set and subspace estimation

Consider a linear fixed-coefficients system that can be described by the State Space (SS) model,

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

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \boldsymbol{\psi}_t \tag{1b}$$

where \mathbf{x}_t is a state n -vector, being n the true order of the system. In addition, \mathbf{z}_t is an observable output m -vector, which is assumed to be zero-mean without loss of generality, $\boldsymbol{\psi}_t$ is a noise m -vector (known as innovations), while $\boldsymbol{\Phi}$, \mathbf{E} and \mathbf{H} are parametric matrices. Model (1a-1b) is called an “innovations model”, used as it is simple and general, in the sense that any fixed-coefficients SS model can be written in this specific form (see e.g., Casals et al., 1999, Theorem 1). Moreover, some assumptions about the system and the noise must be established.

Assumptions A.1. Let $\boldsymbol{\psi}_t$ be a sequence of independent and identically distributed random variable with $E(\boldsymbol{\psi}_t) = 0$ and $E(\boldsymbol{\psi}_t' \boldsymbol{\psi}_t) = \mathbf{Q}$, being \mathbf{Q} a positive definite matrix. **A.2.** Let (1a-1b) be a non-explosive system, that is all the eigenvalues of $\boldsymbol{\Phi}$ lie in or inside the unit circle, which fulfills the strictly minimum-phase condition, i.e. all the eigenvalues of $(\boldsymbol{\Phi} - \mathbf{E}\mathbf{H})$ lie inside the unit circle.

Now we will show that the subspace methods can derive from the innovations model. By substituting (1b) into (1a) in $\boldsymbol{\psi}_t$ and solving by recursion we have:

$$\mathbf{x}_t = (\boldsymbol{\Phi} - \mathbf{E}\mathbf{H})^t \mathbf{x}_0 + \sum_{j=1}^t (\boldsymbol{\Phi} - \mathbf{E}\mathbf{H})^{t-j} \mathbf{E} \mathbf{z}_{j-1} \quad (2)$$

so that the states in time t depend on the initial state and past values of the output. We will use this equation afterward.

On the other hand, by recursive substitution in (1a) and replacing the result into the observation equation (1b), we get:

$$\mathbf{z}_t = \mathbf{H}\boldsymbol{\Phi}^t \mathbf{x}_0 + \mathbf{H} \sum_{j=0}^{t-1} \boldsymbol{\Phi}^j \mathbf{E} \boldsymbol{\psi}_{t-j-1} + \boldsymbol{\psi}_t \quad (3)$$

which means that the endogenous variable, \mathbf{z}_t , depends on the initial state vector, \mathbf{x}_0 , and past and present innovations, $\boldsymbol{\psi}_t$. Equation (3) can be written in matrix form as,

$$\mathbf{Z}_p = \mathbf{O}\mathbf{X}_0 + \mathbf{V}\boldsymbol{\Psi}_p \quad (4)$$

where the subscript p is an integer that denotes the dimension of the row space of \mathbf{Z}_p , see Bauer (2005) for a complete discussion about p . In the following, we will define the matrices in equation (4):

1) Block-Hankel Matrices (BHM), which dimensions are determined by the integers p and f , such that:

$$\mathbf{Z}_p = \begin{pmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_{T-p-f+1} \\ \mathbf{z}_2 & \mathbf{z}_3 & \cdots & \mathbf{z}_{T-p-f+2} \\ \vdots & \vdots & & \vdots \\ \mathbf{z}_p & \mathbf{z}_{p+1} & \cdots & \mathbf{z}_{T-f} \end{pmatrix}; \quad \mathbf{Z}_f = \begin{pmatrix} \mathbf{z}_{p+1} & \mathbf{z}_{p+2} & \cdots & \mathbf{z}_{T-f+1} \\ \mathbf{z}_{p+2} & \mathbf{z}_{p+3} & \cdots & \mathbf{z}_{T-f+2} \\ \vdots & \vdots & & \vdots \\ \mathbf{z}_{p+f} & \mathbf{z}_{p+f+1} & \cdots & \mathbf{z}_T \end{pmatrix} \quad (5)$$

In (4), Ψ_p is as \mathbf{Z}_p but with ψ_t instead of \mathbf{z}_t . For simplicity, in the following we will assume that the dimension of the past and future information sets is the same, *i.e.*, $p = f = i$.

2) The state sequence which is defined as $\mathbf{X}_t = (\mathbf{x}_t \quad \mathbf{x}_{t+1} \quad \mathbf{x}_{t+2} \quad \cdots \quad \mathbf{x}_{t+T-2i})$. Specially, we will use the past and future state sequences, denoted, respectively, by $\mathbf{X}_p = \mathbf{X}_0$ and $\mathbf{X}_f = \mathbf{X}_i$.

3) The Extended Observability matrix, which is:

$$\mathbf{O} = \left(\mathbf{H}' \quad (\mathbf{H}\Phi)'\quad (\mathbf{H}\Phi^2)'\quad \cdots \quad (\mathbf{H}\Phi^{i-1})' \right)'_{im \times n} \quad (6)$$

4) The lower block triangular Toeplitz matrix, defined as:

$$\mathbf{V} = \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{H}\mathbf{E} & \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{H}\Phi\mathbf{E} & \mathbf{H}\mathbf{E} & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{i-2}\mathbf{E} & \mathbf{H}\Phi^{i-3}\mathbf{E} & \mathbf{H}\Phi^{i-4}\mathbf{E} & \cdots & \mathbf{I}_m \end{pmatrix}_{im} \quad (7)$$

Given assumption A.2. and for large values of t , the first addend in equation

(2) is negligible and \mathbf{X}_f is to a close approximation representable as a linear combination of the past of the output, \mathbf{MZ}_p . Shifting time subscripts in (4) and substituting \mathbf{X}_f by \mathbf{MZ}_p lead to,

$$\mathbf{Z}_f = \mathbf{OMZ}_p + \mathbf{V}\Psi_f \quad (8)$$

where \mathbf{Z}_f , \mathbf{Z}_p and Ψ_f are as in (5), and \mathbf{O} and \mathbf{V} , respectively, as in (6) and (7). There are different algorithms within the subspace methods but equality (8) is the common starting point to all of them. Here we use the Canonical Correlation Analysis (CCA) algorithm, 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:

$$\min_{\{\hat{\mathbf{O}}, \hat{\mathbf{M}}\}} \left\| \mathbf{W}_1 (\mathbf{Z}_f - \hat{\mathbf{O}} \hat{\mathbf{M}} \mathbf{Z}_p) \mathbf{W}_2 \right\|_F^2 \quad (9)$$

where $\|\cdot\|_F$ denotes de Frobenius norm. Note that the order, n , and the weightings matrices, \mathbf{W}_1 and \mathbf{W}_2 , have to be specified (see, Katayama, 2005, for different weightings). Compute the states as $\hat{\mathbf{X}}_f = \hat{\mathbf{M}} \mathbf{Z}_p \mathbf{W}_2$.

3. Regress \mathbf{z}_t onto $\hat{\mathbf{x}}_t$, $t = i, \dots, T - i$, obtaining $\hat{\mathbf{H}}$ and the residuals, $\hat{\boldsymbol{\psi}}_t$, as in equation (1b).
4. Regress $\hat{\mathbf{x}}_{t+1}$ onto $\hat{\mathbf{x}}_t$ and $\hat{\boldsymbol{\psi}}_t$, $t = i, \dots, T - i - 1$, obtaining $\hat{\boldsymbol{\Phi}}$ and $\hat{\mathbf{E}}$ as in equation (1a).
5. Check the minimum-phase condition (A.2). If A.2 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

It has been proved that for $i \geq i_0$, the estimates $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\Phi}}, \hat{\mathbf{E}}, \hat{\mathbf{H}}\}$ obtained by the CCA algorithm 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 $0 < a < \infty$. In the stationary case, $d > 1$ is a sufficient condition (Deistler et al., 1995) whereas $d > 2$ is required in the integrated case (Bauer, 2005). However, $\hat{\Theta}_i$, which is the set of matrices estimated with a specific integer i , differ one from another in finite samples. This fact presents two sensible choices in order to improve the forecasts: a) choose the value of i in accordance with an in-sample forecasting criterion, or b) combine several predictions generated from different i .

The first idea is included in the MatLab *System Identification Toolbox* (Ljung, 1999) and consists of: a) selecting a range of possible values, b) estimate the corresponding state space model, c) calculate for each model an information criterion, and d) select the model which minimizes that criterion. However, the alternative seems to be more sensible and promising. This is the main idea of the paper that proposes to combine the predictions obtained from a range of possible values for i .

Firstly, it should be noted that whichever procedure you choose to predict, the results about consistency restrict the lower bound of the range of possible values for i to i_0 .

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

$$\hat{\mathbf{x}}_{t+1}^i = \hat{\Phi}_i \hat{\mathbf{x}}_t^i + \hat{\mathbf{E}}_i \psi_t \quad (10a)$$

$$\hat{\mathbf{z}}_t^i = \hat{\mathbf{H}}_i \hat{\mathbf{x}}_t^i + \psi_t \quad (10b)$$

where $i = i_0, \dots, I$, being I deterministically chosen by the user. Clearly, $\hat{\mathbf{z}}_t^i$ are highly correlated and, as a consequence of consistency, the correlations will increase as the sample size grows. This suggests that the improvement of combining will be more considerable in small than in large samples. Further, let \mathbf{z}_t^s be a vector containing the in-sample predictions $\hat{\mathbf{z}}_t^i$, $i = i_0, \dots, I$, but sorted in a particular

way that will be explained later. Finally, consider a vector of weights such that $\mathbf{\Pi} = [\pi_0 \ \pi_1 \ \dots \ \pi_{I-i_0+1}]'$. From all of this, one can solve the ordinary least squares problem:

$$\min_{\{\hat{\mathbf{\Pi}}\}} \left\| \mathbf{z}_t - [1 \ \mathbf{z}_t^s] \cdot \hat{\mathbf{\Pi}} \right\|_F^2 \quad (11)$$

getting, as a result, $\hat{\mathbf{z}}_t^* = [1 \ \mathbf{z}_t^s] \cdot \hat{\mathbf{\Pi}}$, which is the optimal linear prediction of \mathbf{z}_t given the range of i (see, *i.e.*, Granger and Ramanathan, 1984).

However, if we let the user choose I as bigger as she wants, the information given by the set of explanatory variables will be 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 (11), we suggest increasing the dimension of \mathbf{z}_t^s one by one and using the AIC (Akaike, 1976) to select the best model. I will now motivate why \mathbf{z}_t^s has a specific structure. Vector \mathbf{z}_t^s is organized such that the first component is the $\hat{\mathbf{z}}_t^i$ which presents a lower correlation with the others, the second element is the second less correlated and so on. In this way, the reduction of the sum-squared-error of regression (11) 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 short, the algorithm to compute the final out-of-sample forecasts may be described as follows:

1. Find i_0 as the integer closer to $d\hat{\rho}_{bic}$ and choose I .
2. Estimate $\hat{\Theta}_i$ 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 that which is less correlated (to the rest) to that which is more correlated.
4. Regress \mathbf{z}_t onto $[1 \ \mathbf{z}_t^s]$, $I - i_0 + 1$ times, increasing \mathbf{z}_t^s by one component each time and calculating the AIC in each regression. Keep the weights $\hat{\mathbf{\Pi}}$.
5. Compute the combined out-of-sample forecasts as $\hat{\mathbf{z}}_{t+f}^* = [1 \ \mathbf{z}_{t+f}^s] \cdot \hat{\mathbf{\Pi}}$, where f

is the prediction horizon. The number of columns of \mathbf{z}_{t+f}^s will be determined by minimizing AIC in the previous step.

It is straightforward to see that the proposal presents lower in-sample mean squared error than any common subspace forecast with a fixed value of i in the range (i_0, I) . However, although it could be expectable, this does not guarantee more accurate out-of-sample prediction.

4 An empirical application

In this section we illustrate the methodology by modeling and forecasting the growth rate of the German GDP, hereafter z_t . Specifically, data used 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. Firstly, a one-step-ahead forecast evaluation will be made over the period 2006:02 to 2008:03, updating the models each time with the new data. Secondly, a mid-term prediction analysis is presented by fitting the models for the period of 1991:01 to 2006:01 and forecasting 10 periods, from 2006:02 to 2008:03.

As a result of the autoregressive approximation of z_t , $\hat{\rho}_{bic} = 8$ and i_0 is fixed to 11, assuring the consistency of the estimates. As the sample size is not very large, we decide to fix $I = 20$. Consequently, $I - i_0 + 1 = 10$ models are estimated and used in the prediction exercise. An alternative autoregressive of order 8 (AR(8)) is chosen in order to compare its forecasts with the proposals. Vector AR models should be fair rivals as they present similar properties to subspace methods: 1) speed and stability (no iterations are required) and 2) simplicity (both can be easily automatised).

The forecasting errors are evaluated in terms of Root Mean Squared Error (RMSE) and the predictive accuracy is tested with the Diebold and Mariano (1995) test.

4.1 One-step-ahead forecast evaluation

Table 1 presents the RMSE, ranking and results of the Diebold and Mariano test of the one-step-ahead out-of-sample prediction errors obtained from: a) the combination of the forecasts of the whole vector \mathbf{z}_t^s (called *All combination*), b) the sample mean of the forecasts compute from \mathbf{z}_t^s (called *Mean of all*), c) the combination of the forecasts proposed in Section 3 using the AIC to decrease the dimension of \mathbf{z}_t^s (called *AIC combination*), d) the sample mean of the predictions compute from this reduced \mathbf{z}_t^s (called *Mean of AIC*), e) the usual subspace methods forecasts got with $i = 11, \dots, 20$, and f) the alternative AR(8) model.

[TABLE 1 AND FIGURE 1 SHOULD BE AROUND HERE]

The results show that both combination, either with the whole \mathbf{z}_t^s or with its reduced form by AIC, clearly outperform the rest of models. The differences between the combined procedures are not quite significant (7% better in terms of RMSE) in favour of the *All combination*. The Diebold and Mariano test suggests that both forecasts are not statistically different at 10% of significance. This can also be observed in Figure 1, where the combined predictions move very similarly, which points out the good behaviour of the *AIC combination*. On the other hand, the improvement with respect the rest of the (non-combined) subspace models is very remarkable, as its RMSEs range from 1.9 to 79 times those of the combined models. The Diebold and Mariano test considers all these predictions significantly less precise than those got with the proposed models at 5%. The combined models also outperform the AR(8) in terms of RMSE and the pvalue related to the predictive accuracy test also remains relatively small (.049 and .119 for the *All combination* and the *AIC combination*, respectively).

4.2 Mid-term forecast evaluation

In this subsection, 10 out-of-sample forecasts are computed from 2006:02 to 2008:03. The models are the same used before although this time no update is carried out.

Results are depicted in Table 2.

[TABLE 2 AND FIGURE 2 SHOULD BE AROUND HERE]

As in the previous analysis, models that combine several predictions are ranked in the first positions, presenting almost identical performance. However, in this case the gain of the combined models is not so substantial with respect to, for instance, those estimated with $i = 14$ (see Figure 2) or $i = 19$. Here the enhancement in terms of RMSE is ranged from 1.05 to 1.91. This result is not unexpected as the combinations are constructed to minimize the one-step-ahead error. Even so, the proposals clearly outperform, also from the Diebold and Mariano point of view, several models such as those estimated with $i = 11, 12, 17$ and the AR(8).

5 Concluding remarks

A new procedure to forecast linear dynamical systems using subspace methods has been put forward. It is based on combining multiple predictions obtained from setting a range of values for a parameter that is commonly fixed by the user in the literature of subspace methods. An algorithm which provides a suitable number of combinations is also proposed. Finally, an empirical application using the German GDP shows that the procedures outperform the out-of-sample forecasts got with subspace methods (as they are so far computed) and with an alternative autoregressive model in one-step-ahead and mid-term.

These algorithms are implemented in a MatLab toolbox for time series modeling called (*This information has been deliberately omitted in the blind version*). The source code of this toolbox is freely provided under the terms of the GNU General Public License and can be downloaded at (*This information has been deliberately omitted in the blind version*).

References

- Akaike, H. (1976). *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*. Academic Press.
- Bauer, D. (2005). 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.
- Casals, J., Sotoca, S., and Jerez, M. (1999). A fast and stable method to compute the likelihood of time invariant state space models. *Economics Letters*, 65(3):329–337.
- Deistler, M., Peternell, 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. (2009a). 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. (2009b). Unit roots and cointegration modeling through a family of flexible information criteria. *Journal of Statistical Computation and Simulation*, page In press.
- 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.

- Ho, B. and Kalman, R. (1966). Effective construction of linear state-variable models from input-output functions. *Regelungstechnik*, 14:545–548.
- 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*, 33:267–282.
- 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.
- 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

Table 1: Evaluation of the one-step-ahead prediction errors*

Model	RMSE			Diebold and Mariano Test			
	Value	Relative	Rk	Rk(1) vs Rk(j)		Rk(2) vs Rk(j)	
				Statistic	Pvalue	Statistic	Pvalue
All combination	0.789	100	1	-	-	0.984	0.838
Mean of all	6.960	882	13	-3.333	0.000	-3.284	0.001
AIC combination	0.847	107	2	-0.984	0.162	-	-
Mean of AIC	9.855	1249	14	-3.554	0.000	-3.525	0.000
$i = 11$	62.660	7944	15	-3.499	0.000	-3.499	0.000
$i = 12$	3.393	430	12	-2.368	0.009	-2.333	0.010
$i = 13$	2.407	305	7	-1.999	0.023	-1.918	0.028
$i = 14$	1.745	221	6	-2.805	0.003	-2.664	0.004
$i = 15$	2.442	310	8	-2.860	0.002	-2.659	0.004
$i = 16$	2.666	338	9	-2.842	0.002	-2.688	0.004
$i = 17$	1.677	213	5	-5.102	0.000	-4.383	0.000
$i = 18$	1.504	191	4	-2.374	0.009	-2.098	0.018
$i = 19$	2.925	371	10	-2.094	0.018	-2.018	0.022
$i = 20$	3.245	411	11	-2.080	0.019	-2.019	0.022
AR(8)	0.945	120	3	-1.656	0.049	-1.178	0.119

* Prediction errors are multiplied by 100 in order to facilitate the comparison. Diebold and Mariano 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 .

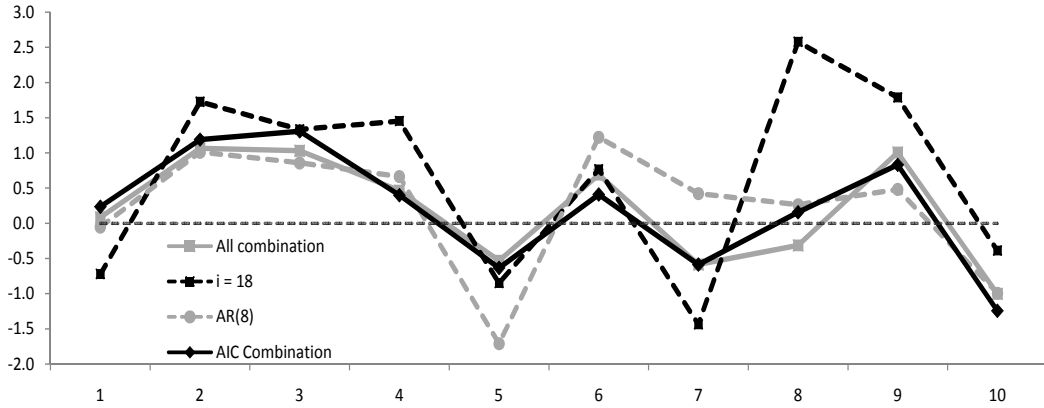


Figure 1: Best one-step-ahead prediction errors. It includes those obtained from the combined models (*All* and *AIC*), $i = 18$ and AR(8) model.

Table 2: Evaluation of the horizon 1-to-10 prediction errors*

Model	RMSE			Diebold and Mariano Test			
	Value	Relative	Rk	Rk(1) vs Rk(j)		Rk(2) vs Rk(j)	
				Statistic	Pvalue	Statistic	Pvalue
All combination	0.819	100	1	-	-	0.088	0.535
Mean of all	0.876	107	4	-1.155	0.124	-1.083	0.139
AIC combination	0.829	101	2	-0.088	0.465	-	-
Mean of AIC	0.914	112	6	-0.814	0.208	-3.343	0.000
$i = 11$	1.568	191	15	-2.691	0.004	-2.791	0.003
$i = 12$	1.081	132	13	-2.269	0.012	-1.408	0.080
$i = 13$	1.053	128	12	-0.960	0.169	-1.040	0.149
$i = 14$	0.863	105	3	-0.187	0.426	-0.252	0.400
$i = 15$	1.046	128	10	-0.954	0.170	-0.975	0.165
$i = 16$	0.948	116	7	-0.915	0.180	-0.621	0.267
$i = 17$	1.047	128	11	-1.475	0.070	-1.377	0.084
$i = 18$	0.990	121	9	-1.260	0.104	-0.840	0.200
$i = 19$	0.888	108	5	-0.395	0.346	-0.647	0.259
$i = 20$	0.964	118	8	-1.709	0.044	-0.932	0.176
AR(8)	1.375	168	14	-1.559	0.059	-1.553	0.060

* Prediction errors are multiplied by 100 in order to facilitate the comparison. Diebold and Mariano 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 $k = 1, 2, \dots, 10$.

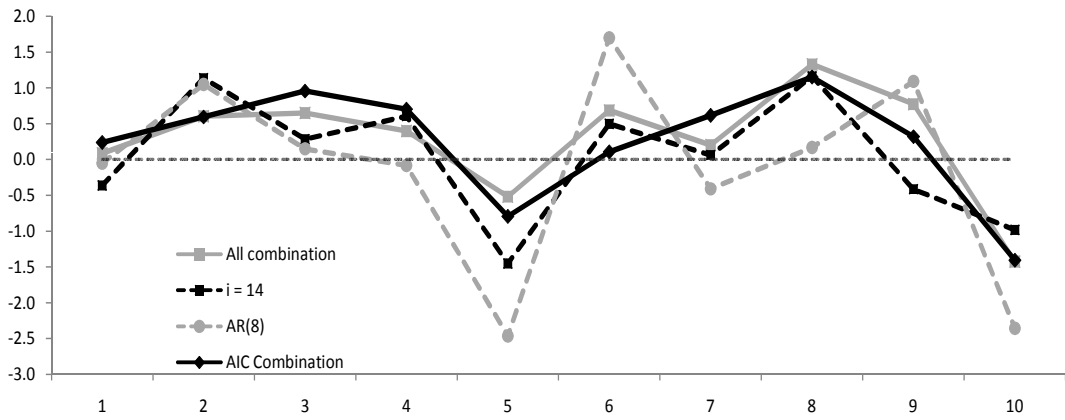


Figure 2: Best horizon 1-to-10 prediction errors (combined models and $i = 14$) compared with those obtained from the AR(8) model.