Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Econometrics 119 (2004) 291 – 321
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Forecasting with nonstationary dynamic factor models

Daniel Peña[a], Pilar Poncela[b,*]

[a]Department of Statistics and Econometrics, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe, Madrid, Spain
[b]Departamento de Análisis Económico: Economía Cuantitativa, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain

Abstract

In this paper we analyze the structure and the forecasting performance of the dynamic factor model. It is shown that the forecasts obtained by the factor model imply shrinkage pooling terms, similar to the ones obtained from hierarchical Bayesian models that have been applied successfully in the econometric literature. Thus, the results obtained in this paper provide an additional justification for these and other types of pooling procedures. The expected decrease in MSE for using a factor model versus univariate ARIMA and shrinkage models are studied for the one factor model. Monte Carlo simulations are presented to illustrate this result. A factor model is also built to forecast GNP of European countries and it is shown that the factor model can provide a substantial improvement in forecasts with respect to both univariate and shrinkage univariate forecasts.

JEL classification: C32; C53

Keywords: Cointegration; Common factors; Pooled forecasts; Prediction; Vector time series

1. Introduction

In this article, we address the important issue of forecasting a vector of time series that has been generated by a set of dynamic common factors, possibly nonstationary. Dynamic factor models have been studied by Geweke (1977), Geweke and Singleton (1981), Engle and Watson (1981), Velu et al. (1986), Peña and Box (1987), Stock and Watson (1988), Tiao and Tsay (1989), Engle and Kozicki (1993), Gonzalo and

* Corresponding author. Tel.: +34-91-3975521; fax: +34-91-3974091.
E-mail addresses: dpena@est-econ.uc3m.es (D. Peña), pilar.poncela@uam.es (P. Poncela).

0304-4076/$ - see front matter
doi:10.1016/S0304-4076(03)00198-2

Granger (1995), Vahid and Engle (1997), Forni and Reichlin (1998) and Peña and Poncela (2002), among others (see these articles for further references). When the factors are nonstationary, the problem is very related to forecasting by using cointegration relationships (see Escribano and Peña, 1994).

The question of how the presence of common factors, or equivalently cointegration relations, among a collection of variables affects forecasting is ambiguous. Engle and Yoo (1987) considered a bivariate model and found that taking into account the equilibrium relations improved the long-run predictions, but not the short-run ones. Reinsel and Ahn (1992) found that while overspecifying the number of unit roots led to worse results in the short-run forecasts, underspecifying them led to worse long-run forecasts. In the same line, Clements and Hendry (1995) found that overspecifying the number of unit roots derives in worse results in forecasting. Lin and Tsay (1996) explored simulated and real data and concluded that imposing the correct number of unit roots improves the forecasting results for simulated data, while for the real data analyzed in their article, imposing the number of unit roots suggested by the cointegration tests did not necessarily lead to better results. Christoffersen and Diebold (1998) found that the presence of cointegration relations did not outperform the long run forecasts of univariate models and similar empirical results for the UK demand for money were found by Garcìa-Ferrer and Novales (1998).

This paper has the following contributions. First, we show that the factor model forecasts incorporate a pooling term similar to the one derived from hierarchical Bayesian models. This pooling term can be, in some particular cases, identical to the shrinkage term proposed by Garcìa-Ferrer et al. (1987). Thus the factor model provides a formal justification for univariate shrinkage forecast methods and allows to derive the optimal shrinkage in each case. Second, we derive the expected gains, in the one factor case, of the factor model forecasts with respect to univariate and shrinkage models. The advantage of the factor model increases with the dimension of the time series vector and with the strength of the dynamic relationship among the components. Third, we show by Monte Carlo and a real example that in some cases we can obtain a substantial reduction in mean square forecast error from the factor model with respect to alternative forecasting approaches.

The paper is organized as follows. In Section 2 we briefly review the dynamic factor model and the generation of forecasts from it. In Section 3 we analyze the structure of the factor model forecasts and in Section 4 we study with more detail the one factor case. In Section 5 the large sample comparison of the forecast performance of the one factor model and the ARIMA and shrinkage pooled univariate models is presented. In Section 6 these results are illustrated in a Monte Carlo study and in Section 7 by forecasting the Gross National Product (GNP) of European countries. Finally, Section 8 presents the limitation of our work and some concluding remarks. The proofs of the lemmas and theorems in the text are given in the appendix.

2. The factor model

Let $\mathbf{y}_t$ be an m-dimensional vector of observed time series, generated by a set of r non observed common factors. We assume that each component of the vector of observed series, $\mathbf{y}_t$, can be written as a linear combination of common factors plus noise,

$$\underset{m\times 1}{\mathbf{y}_t} = \underset{m\times r}{\mathbf{P}}\ \underset{r\times 1}{\mathbf{f}_t} + \underset{m\times 1,}{\varepsilon_t,} \tag{1}$$

where $\mathbf{f}_t$ is the r-dimensional vector of *common factors*, $\mathbf{P}$ is the factor loading matrix, and $\varepsilon_t \sim \mathrm{N}_m(\mathbf{0}, \mathbf{\Sigma}_\varepsilon)$, with $\mathbf{\Sigma}_\varepsilon = diag(\sigma_1^2, \ldots, \sigma_m^2)$ and $\sigma_j^2 < \infty\ \forall j$. We suppose that the vector of common factors follows a VARIMA(p, d, q) model:

$$\underset{r\times r}{\mathbf{\Phi}(B)}\ \underset{r\times 1}{\mathbf{f}_t} = \underset{r\times r}{\mathbf{\Theta}(B)}\ \underset{r\times 1,}{\mathbf{a}_t,} \tag{2}$$

where $\mathbf{\Phi}(B) = \mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p$ and $\mathbf{\Theta}(B) = \mathbf{I} - \mathbf{\Theta}_1 B - \cdots - \mathbf{\Theta}_q B^q$ are polynomial matrices $r \times r$, B is the backshift operator, such that $B\mathbf{y}_t = \mathbf{y}_{t-1}$, the roots of $|\mathbf{\Phi}(B)| = 0$ are on or outside the unit circle, the roots of $|\mathbf{\Theta}(B)| = 0$ are outside the unit circle and $\mathbf{a}_t \sim \mathrm{N}_r(\mathbf{0}, \mathbf{\Sigma}_a)$ is serially uncorrelated, $\mathrm{E}(\mathbf{a}_t \mathbf{a}'_{t-h}) = \mathbf{0}$, $h \neq 0$. We assume that the noises from the common factors and the observed series are also uncorrelated for all lags, $\mathrm{E}(\mathbf{a}_t \varepsilon'_{t-h}) = \mathbf{0}\ \forall h$.

The model as stated is not identified, because for any $r \times r$ nonsingular matrix $\mathbf{H}$ the observed series $\mathbf{y}_t$ can be expressed in terms of a new set of factors and system matrices. To solve this identification problem, we can always choose either $\mathbf{\Sigma}_\mathbf{a} = \mathbf{I}$ or $\mathbf{P}'\mathbf{P} = \mathbf{I}$, but it is easy to see that the model is not yet identified under rotations. Harvey (1989) imposes the additional condition that $p_{ij} = 0$, for $j > i$, where $\mathbf{P} = [p_{ij}]$. This condition is not restrictive, since the factor model can be rotated for a better interpretation when needed (see Harvey (1989) for a brief discussion about it). In this paper we will impose that $\mathbf{\Sigma}_\mathbf{a} = \mathbf{I}$; this restriction is enough to achieve identification for the one factor model and excludes the case where a common factor is just a constant, which is not analyzed in this paper. We will add, when needed, the standard restriction in static factor analysis, that is $\mathbf{P}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{P}$ diagonal.

The model can be generalized to the case where the components in ε_t have dynamic univariate stationary structure, see Peña and Poncela (2002), but this does not affect the conclusions derived in forecasting and complicates the algebra involved. Also, it can be seen that the model is fairly general and includes the case where lagged factors are present in Eq. (1).

Estimation and forecasting can be carried out by writting the model in state space form as follows: the vector of observable time series $\mathbf{y}_t$, is given by the *measurement equation*:

$$\underset{m\times 1}{\mathbf{y}_t} = \underset{m\times s}{\tilde{\mathbf{P}}}\ \underset{s\times 1}{\mathbf{z}_t} + \underset{m\times 1}{\varepsilon_t} \tag{3}$$

and the state vector $\mathbf{z}_t$ containing the factors, forecasted factors, lagged factors or error terms (depending on the state space representation that is chosen) is driven by the *transition equation*:

$$\underset{s\times 1}{\mathbf{z}_t} = \underset{s\times s}{\mathbf{G}}\ \underset{s\times 1}{\mathbf{z}_{t-1}} + \underset{s\times 1}{\mathbf{u}_t,} \tag{4}$$

with $\mathrm{E}(\mathbf{u}_t) = \mathbf{0}$, $\mathrm{E}(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma}_u$ and $\mathrm{E}(\mathbf{u}_t\mathbf{u}_\tau') = \mathbf{0}$ if $t \neq \tau$. Both noises, ε_t and $\mathbf{u}_t$, are also uncorrelated for all lags, $\mathrm{E}(\varepsilon_t\mathbf{u}_\tau') = \mathbf{0}$ for all t and τ. Given information until time $t-1$, the well-known Kalman filter equations give the forecast of the state vector, $\mathbf{z}_{t|t-1} = \mathbf{G}\mathbf{z}_{t-1|t-1}$, with associated covariance matrix,

$$\mathbf{V}_{t|t-1} = \mathbf{G}\mathbf{V}_{t-1|t-1}\mathbf{G}' + \boldsymbol{\Sigma}_u \tag{5}$$

and the forecast for the vector of time series is computed by $\hat{\mathbf{y}}_{t|t-1} = \tilde{\mathbf{P}}\mathbf{z}_{t|t-1}$ with covariance matrix

$$\boldsymbol{\Sigma}_{t|t-1} = \tilde{\mathbf{P}}\mathbf{V}_{t|t-1}\tilde{\mathbf{P}}' + \boldsymbol{\Sigma}_\varepsilon. \tag{6}$$

When a new observation arrives, the state vector is updated by

$$\mathbf{z}_{t|t} = \mathbf{z}_{t|t-1} + \mathbf{V}_{t|t-1}\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_{t|t-1}^{-1}(\mathbf{y}_t - \tilde{\mathbf{P}}\mathbf{z}_{t|t-1}) \tag{7}$$

and its variance–covariance matrix by

$$\mathbf{V}_{t|t} = \mathbf{V}_{t|t-1} - \mathbf{V}_{t|t-1}\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_{t|t-1}^{-1}\tilde{\mathbf{P}}\mathbf{V}_{t|t-1}. \tag{8}$$

3. The structure of factor model forecasts

In this section we derive a structural form for the predictions of the factor model that shows the effect of a pooling or shrinkage term in the predictions for each component of the vector time series. Consider the factor model in state space form given by (3) and (4). The h-steps ahead forecast of the state vector with observations up to time t is obtained by $\mathbf{z}_{t+h|t} = \mathbf{G}^h\mathbf{z}_{t|t}$, with mean square error (MSE) matrix

$$\begin{aligned}\mathbf{V}_{t+h|t} &= \mathrm{E}(\mathbf{z}_{t+h} - \mathbf{z}_{t+h|t})(\mathbf{z}_{t+h} - \mathbf{z}_{t+h|t})' \\ &= \mathbf{G}^h\mathbf{V}_{t|t}(\mathbf{G}')^h + \mathbf{G}^{h-1}\boldsymbol{\Sigma}_u(\mathbf{G}')^{h-1} + \cdots + \mathbf{G}\boldsymbol{\Sigma}_u\mathbf{G}' + \boldsymbol{\Sigma}_u.\end{aligned} \tag{9}$$

The h steps ahead forecast for the observed series with origin in t is

$$\hat{\mathbf{y}}_{t+h|t} = \tilde{\mathbf{P}}\mathbf{G}^h\mathbf{z}_{t|t} \tag{10}$$

with MSE matrix

$$\boldsymbol{\Sigma}_{t+h|t} = \mathrm{E}(\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t})(\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t})' = \tilde{\mathbf{P}}\mathbf{V}_{t+h|t}\tilde{\mathbf{P}}' + \boldsymbol{\Sigma}_\varepsilon. \tag{11}$$

Using (7) in (10), we can write the forecast of the observed series as

$$\hat{\mathbf{y}}_{t+h|t} = \mathbf{A}_1\mathbf{z}_{t|t-1} + \mathbf{A}_2\mathbf{y}_t, \tag{12}$$

where

$$\mathbf{A}_1 = \tilde{\mathbf{P}}\mathbf{G}^h(\mathbf{I} - \mathbf{V}_{t|t-1}\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_{t|t-1}^{-1}\tilde{\mathbf{P}}) \tag{13}$$

and

$$\mathbf{A}_2 = \tilde{\mathbf{P}}\mathbf{G}^h\mathbf{V}_{t|t-1}\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_{t|t-1}^{-1}. \tag{14}$$

An equivalent expression of (12) can be obtained by using the well-known expression for the inverse of the sum of two matrices (see, Rao, 1973, p. 33) in (6) for the inverse of $\boldsymbol{\Sigma}_{t|t-1}$ and plugging it into (14). This leads to

$$\hat{\mathbf{y}}_{t+h|t} = \mathbf{A}_1 \mathbf{z}_{t|t-1} + \mathbf{W}_t \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{y}_t, \tag{15}$$

where $\mathbf{W}_t = \tilde{\mathbf{P}} \mathbf{N}_t \tilde{\mathbf{P}}'$ is an $m \times m$ matrix and (see the appendix)

$$\mathbf{N}_t = \mathbf{G}^h \mathbf{V}_{t|t-1} (\mathbf{I}_s - \tilde{\mathbf{P}}' \boldsymbol{\Sigma}_\varepsilon^{-1} \tilde{\mathbf{P}} \mathbf{V}_{t|t}), \tag{16}$$

an $s \times s$ matrix. Note that although if $\tilde{\mathbf{P}}' \boldsymbol{\Sigma}_\varepsilon^{-1} \tilde{\mathbf{P}} = \mathbf{P}' \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{P}$ is a diagonal matrix, by the identification restrictions, even though $\boldsymbol{\Sigma}_u$ and $\mathbf{G}$ were diagonal $\mathbf{V}_{t|t-1}$ and $\mathbf{V}_{t|t}$ do not need to be diagonal. From (15) and since $\boldsymbol{\Sigma}_\varepsilon$ is a diagonal matrix, the jth component of the forecast vector can be written as

$$\hat{y}_{j,t+h|t} = (\mathbf{A}_1 \mathbf{z}_{t|t-1})_j + \sum_{i=1}^{m} \frac{w_{ji,t}}{\sigma_i^2} y_{i,t},$$

where $(\mathbf{x})_j$ is the jth component of vector $\mathbf{x}$ and $w_{ji,t}$ is the (j,i) element of $\mathbf{W}_t$. This equation shows that the forecast of each component of the series incorporates a pooling term given by a weighted sum of all the observed series at time t with weights proportional to the elements of the $\mathbf{W}_t$ matrix and inversely proportional to their noise variances.

4. The prediction structure of the one factor model

The one factor model has special interest because many economic time series are characterized by a common trend. For example, it can be considered that the GNP of some countries of a certain area of influence are driven by the same common trend. It has also been widely used in the business cycle analysis (see, for instance, Stock and Watson (1991) and Diebold and Rudebusch (1996) among others). In this section we analyze the structure of forecasts for the one factor model, first when the factor is a common trend or a stationary AR(1) process, and second in the general case in which the factor follows an ARIMA model. We will also see that for the case of the common trend, this pooling term is permanent, while for a stationary ARMA(p,q) factor, the pooling term is transitory.

Assume first the simplest one factor model

$$\mathbf{y}_t = \mathbf{P} f_t + \varepsilon_t \tag{17}$$

with ε_t multivariate white noise with $\boldsymbol{\Sigma}_\varepsilon = diag(\sigma_1^2, \ldots, \sigma_m^2)$ and factor loading matrix $\mathbf{P} = (p_1, p_2, \ldots, p_m)'$. The equation for the factor is

$$f_t = \phi f_{t-1} + a_t \tag{18}$$

with $var(a_t) = \sigma_a^2 = 1$ (by the identification restriction), $cov(\varepsilon_{i,t} a_\tau) = 0$ for all i, t and τ and $|\phi| \leqslant 1$. This specification includes AR(1) stationary processes when $|\phi| < 1$, as well as nonstationary ones when $\phi = 1$. The model is in state space form with

$\tilde{\mathbf{P}} = \mathbf{P}$, $z_t = f_t$, $r = s = 1$ and $u_t = a_t$. Let us study in this case the h steps ahead forecast of the observed series $\mathbf{y}_{t+h}$ with information up to time t, given by (15). By (13) the matrix $\mathbf{A}_1$ is given by $\mathbf{A}_1 = \mathbf{P}\phi^h(\mathbf{1} - V_{t|t-1}\mathbf{P}'\mathbf{\Sigma}_{t|t-1}^{-1}\mathbf{P})$, where $V_{t|t-1}$ is the variance (scalar, in this case) of the factor at time t with information up to time $t-1$. Applying again the inverse lemma for the sum of two matrices (Rao, 1973) to $\mathbf{\Sigma}_{t|t-1}^{-1} = (V_{t|t-1}\mathbf{P}\mathbf{P}' + \mathbf{\Sigma}_\varepsilon)^{-1}$, $\mathbf{A}_1$ is given by

$$\mathbf{A}_1 = \mathbf{P}\frac{\phi^h}{c_t}\frac{1}{V_{t|t-1}}$$

with $c_t = 1/V_{t|t-1} + \sum_{i=1}^m p_i^2/\sigma_i^2$. The second term in Eq. (15) is $\mathbf{W}_t\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{y}_t = N_t\mathbf{P}\mathbf{P}'\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{y}_t$ where N_t is now an scalar given by $N_t = \phi^h/c_t$. Substituting in (15) $\mathbf{A}_1$ and $\mathbf{W}_t\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{y}_t$ by their expressions, the h steps ahead forecast with information up to time t is

$$\hat{\mathbf{y}}_{t+h|t} = \mathbf{P}\,\frac{\phi^h}{c_t}\left(\frac{1}{V_{t|t-1}}f_{t|t-1} + \sum_{i=1}^m \frac{p_i^2}{\sigma_i^2}\left(\frac{y_{i,t}}{p_i}\right)\right) \tag{19}$$

and the jth component of the forecasted series is $\hat{y}_{j,t+h|t} = p_j(\phi^h/c_t)\,((1/V_{t|t-1})f_{t|t-1} + \sum_{i=1}^m(p_i^2/\sigma_i^2)(y_{i,t}/p_i))$. To understand the meaning of the previous equations, first notice that $1/c_t((1/V_{t|t-1})f_{t|t-1} + \sum_{i=1}^m(p_i^2/\sigma_i^2)(y_{i,t}/p_i))$ is the estimation of the factor with information up to time t as a weighted mean. The first term is the estimation of the factor at time t with the information up to time $t-1$, and it has a weight proportional to the precision of this estimation. The second term is the estimation of the factor provided by the information contained in $\mathbf{y}_t$. To see this, note that by (17) at each time t we have m new possible independent estimates of f_t given by $\mathrm{E}(f_t|y_{j,t}) = y_{j,t}/p_j$ with variances equal to σ_j^2/p_j^2 for each $j = 1,\dots,m$. The second term is a combination of these estimates weighted by their precision.

Therefore we can conclude that the forecast of $\mathbf{y}_{t+h}$ incorporates a pooling term which is the weighted sum of all the series at time t standardized by their factor loadings, with weights inversely proportional to the noise variances in the measurement equation and directly proportional to the square of the factor loadings. The ratios p_i^2/σ_i^2 will appear throughout the article because, as we will see, they are of key importance in comparing forecasts. Let us denote by

$$\mu_i = \frac{p_i^2}{\sigma_i^2}, \quad i = 1,\dots,m, \tag{20}$$

the precision of the estimation of the factor from series y_i. These ratios determine how new information is incorporated into the forecasts. The greater μ_i is, the stronger is the signal of the common information on the series and vice versa, the smaller it is, the weaker is the common information signal in relation to the specific noise of each series.

If the common factor is stationary $|\phi| < 1$, so that $\phi^h \to 0$ as h increases, the forecast of the observed series and, in particular, the pooling term, decreases with exponential decay, until its effect disappears. This means that it is of a transitory nature. In the nonstationary case, $\phi^h = 1$, and as it was expected, the term has a permanent effect. Moreover, if the common factor affects identically to all the series, the factor loading matrix $\mathbf{P}$ is the $m \times 1$ vector $\mathbf{P} = \mathbf{1}p = p(1,1,\dots,1)'$. The forecast of the observed series $\hat{\mathbf{y}}_{t+h|t}$ is $\hat{\mathbf{y}}_{t+h|t} = \mathbf{1}(p\phi^h/c_t)((1/V_{t|t-1})f_{t|t-1} + p\sum_{i=1}^m (1/\sigma_i^2)y_{i,t})$ with $c_t = 1/V_{t|t-1} + p^2\sum_{i=1}^m 1/\sigma_i^2$. The forecast of each component of the time series vector is the same for all of them. Of course, in the case of common trends $\phi^h = 1$. Note that in the particular case in which all the series have the same noise variance, (thus $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2 = \sigma^2$) then $c_t = 1/V_{t|t-1} + p^2 m/\sigma^2$ and the optimal forecast

$$\hat{y}_{j,t+h|t} = \phi^h((1-w_t)pf_{t|t-1} + w_t\bar{y}_t), \tag{21}$$

where $w_t = p^2 m/\sigma^2 c_t$ and the forecast is obtained by shrinking the series towards the common mean, where $\bar{y}_t = 1/m\sum_{i=1}^m y_{i,t}$ is the mean at time t of the m observed variables. Note that we obtain a weighted average of the common mean and the forecast generated from the estimation of the factors with information until $t-1$. If the series are random walks, we obtain a predictor with similar structure to the one used by Garcìa-Ferrer et al. (1987):

$$\tilde{y}_{j,t+h} = (1-\eta)\hat{y}_{j,t+h|t} + \eta\hat{y}_t(h), \quad 0<\eta<1, \tag{22}$$

where $\tilde{y}_{j,t+h}$, is the h steps ahead forecast of the jth component of $\mathbf{y}_t$ at time t and $\hat{y}_t(h) = 1/m\sum_{i=1}^m \tilde{y}_{i,t+h|t} = \bar{y}_t$.

Let us consider the case in which the factor follows an ARIMA(p,d,q) process. The state space form we adopt here is the one originally proposed by Akaike (1974). Assume that the factor follows the ARMA process given by (2) for r=1. The dimension of the state vector is $s = \max\{p+d, q+1\}$, the measurement equation (3) is now

$$\mathbf{y}_t = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ p_m & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} f_t \\ \vdots \\ f_{t+s-1|t} \end{bmatrix} + \varepsilon_t$$

and the transition equation can be written as (4) with $\mathbf{z}_t' = (f_t, f_{t+1|t}, \ldots, f_{t+s-1|t})$, $\mathbf{u}_t' = a_t(1, \psi_1, \psi_2, \ldots, \psi_s)$ where the ψ_i are the coefficients obtained from $\varphi(B)\psi(B) = \varphi(B)\sum_{i=0}^{\infty}\psi_i B^i = \theta(B)$ where $\varphi(B) = \nabla^d\phi(B)$ and

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\varphi_r & -\varphi_{r-1} & -\varphi_{r-2} & \cdots & -\varphi_1 \end{bmatrix}$$

with $\varphi_i = 0$ if $i > p+d$. It is straightforward to show that the h steps ahead forecast of the jth series in $\mathbf{y}_t$, $h<q$ is again

$$y_{j,t+h|t} = (\mathbf{A}_1\mathbf{z}_{t|t-1})_j + n_{11}p_j\sum_{i=1}^m \frac{p_i}{\sigma_i^2}y_{i,t},$$

where n_{11} is the element $(1,1)$ of $\mathbf{N}_t$. As in the previous case, it is easily seen that the pooling term reduces to the common mean of the observed series at time t when the common factor affects all series in the same way (then $p_i = p\ \forall i = 1,\dots,m$) and the noises of all the series have the same variance (then $\sigma_i^2 = \sigma^2\ \forall i = 1,\dots,m$). Again, if the noise variances of the observed series are different, the pooling term is proportional to the weighted mean $\sum_{i=1}^{m} y_{i,t}/\sigma_i^2$. In the MA($q$) case, $\mathbf{G}^h = 0$ for $h > q$, for any q positive integer, so the pooling term is zero for $h > q$. For the AR(p), in the stationary case, the pooling term vanishes when $h \to \infty$. If the process is nonstationary, the pooling term has a permanent effect.

5. Comparison of univariate, pooled and factor model forecasts in the one factor model

In this section we will use the following well-known result from prediction theory. Let

$$\hat{y}^{(i)}_{t+h|t} = \mathrm{E}(y_{t+h}|I_{it})$$

be the optimal forecast with the MSE criterion of a random variable y_{t+h} given the set of information I_{it}, and let $MSE(\hat{y}^{(i)}_{t+h|t})$ the mean square error of this forecast. It is well known that if we consider two sets of information, I_{1t}, I_{2t} such that $I_{1t} \subseteq I_{2t}$ then

$$MSE(\hat{y}^{(1)}_{t+h|t}) \geqslant MSE(\hat{y}^{(2)}_{t+h|t}).$$

Thus, the forecast of any component of the vector time series from the factor model, that incorporates the past information of the other components, should be at least as precise as the univariate forecast, that uses only past information of this component. Also, when the pooled forecasts can be interpreted as the conditional expectation including some joint information from the rest of the series they should be at least as precise as the univariate forecasts. Of course this result is true if we know the parameters of the model and compute the exact conditional expectations. In practice, when the parameters are estimated from the data and the sample size is not large this increase in precision may not be observed.

In this section we quantify the increase in forecast precision of the factor model versus forecasting methods based on a smaller information set. First we obtain the univariate ARIMA models implied by the factor model and compute their MSE of prediction. Second, we compare the trace of the MSE of prediction of the factor model with the sum of MSE of prediction of the univariate ARIMA models. Third we obtain the expected gain of the pooled forecasts with respect to the univariate forecasts. Finally we compute the expected gain of the factor model forecasts with respect to the pooled forecasts.

5.1. Univariate ARIMA forecasts from the factor model

The univariate series generated by the one factor model with AR(1) dynamics verify

$$y_{j,t} - \phi y_{j,t-1} = p_j a_t + \varepsilon_{j,t} - \phi \varepsilon_{j,t-1} \tag{23}$$

and they will follow the ARMA(1,1) model

$$y_{j,t} - \phi y_{j,t-1} = v_{j,t} - \theta_j v_{j,t-1}. \tag{24}$$

In order to obtain the MA parameter implied by this representation, let $var(v_{j,t}) = \tilde{\sigma}_j^2$, then equating moments in the MA part in both sides and using definition (20) θ_j must satisfy (see proof of the auxiliary lemma, Part 1, in the appendix)

$$\theta_j^2\phi - \theta_j(\mu_j + 1 + \phi^2) + \phi = 0. \tag{25}$$

The invertible solution of (25) is given by

$$\theta_j = \frac{\mu_j + 1 + \phi^2 - \sqrt{(\mu_j + 1 + \phi^2)^2 - 4\phi^2}}{2\phi}. \tag{26}$$

Note first that $sign(\theta_j) = sign(\phi)$. Also, if $\mu_j = 0$, then $\theta_j = \phi$. This will happen if $p_j^2 = 0$. In this case the series is not affected by the factor and it will follow a white noise process. Since the generation of the series does not participate to the common information, the series should be dropped from the joint analysis. In what follows, and to simplify the exposition, we will eliminate this case by assuming, without loss of generality, that $\mu_j \neq 0$. Then we have that

$$|\theta_j| \leqslant \frac{\mu_j + 1 + \phi^2 - (\mu_j + 1 - \phi^2)}{2|\phi|} = |\phi| \tag{27}$$

and the absolute value of the MA parameter is always smaller than this of the AR parameter. On the other hand, if $\mu_j \to \infty$, (the signal from the common information is really strong compared to the noise), $\theta_j \to 0$ and the univariate series will be AR(1) as it is the common factor. Finally, note that if $\phi > 0$, $\mathrm{d}\theta_j/\mathrm{d}\mu_j < 0$ and if $\phi < 0$, $\mathrm{d}\theta_j/\mathrm{d}\mu_j > 0$, so θ_j always decreases in magnitude with μ_j, the intensity of the common information. From (25), it is also straightforward to show that the difference between the AR and MA parameters is given by

$$\phi - \theta_j = \frac{\mu_j\theta_j}{1 - \theta_j\phi} = \frac{p_j^2\theta_j}{\sigma_j^2(1 - \theta_j\phi)}. \tag{28}$$

Lemma 1. *Let U be the sum of MSE of the h steps ahead prediction of the univariate ARIMA models obtained from the one factor model in* (17) *and* (18). *Then*

$$U = \sum_{j=1}^{m}\left(\sigma_j^2 + \frac{p_j^2\theta_j}{\phi(1 - \phi\theta_j)} + \frac{p_j^2(\phi - \theta_j)}{(1 - \theta_j\phi)\phi}\sum_{i=0}^{h-1}\phi^{2i}\right). \tag{29}$$

Proof. The proof is given in the appendix. □

The sum of prediction MSE depends on the number of series m, the factor loadings p_j, which measure the effect of the factor on the series, the AR parameter ϕ, which gives the dynamics of the common factor and, of course, the forecast horizon h. If the factor is nonstationary, $\phi = 1$, the univariate models are IMA(1, 1), and using (28) we have

$$U = \sum_{j=1}^{m}\sigma_j^2(2 - \theta_j + \mu_j h)$$

and when the horizon of prediction goes to ∞, the sum as well as the average of prediction MSE also goes to ∞. On the other hand if the factor is stationary, $|\phi| < 1$, the univariate models are ARMA$(1,1)$ and the sum of prediction MSE can be written as

$$U = \sum_{j=1}^{m} \sigma_j^2 \left(1 + \frac{\mu_j}{\phi(1-\phi\theta_j)}\left(\theta_j + \frac{1-\phi^{2h}}{1-\phi^2}(\phi-\theta_j)\right)\right)$$

and if the prediction horizon goes to ∞, the average prediction MSE converges to

$$\lim_{h\to\infty} \frac{1}{m} U = \sigma_M^2 + \frac{1}{1-\phi^2} p_M^2,$$

where $p_M^2 = \sum_{j=1}^{m} p_j^2/m$ is the average square factor loading, and $\sigma_M^2 = \sum_{j=1}^{m} \sigma_j^2/m$. This equation shows that this limit is the sum of the average univariate measurement error plus the average induced effect by the factor model.

5.2. Comparison between univariate and factor model forecasts

Let $\Delta_{u-f} = U - F$, where $F = tr(\mathbf{\Sigma}_{t+h|t})$. The increase in precision of the factor model with respect to the univariate ARIMA forecasts is provided by the following theorem.

Theorem 2. *For the one factor model given by (17) and (18), with $\mu_j \neq 0$, let $\Delta_{u-f} = U - F$ be the increase in precision of the factor model with respect to the univariate forecasts. For the nonstationary case*

$$\Delta_{u-f} = \sum_{j=1}^{m} p_j^2 \left(\frac{\sqrt{\mu_j^2 + 4\mu_j} - \mu_j}{2\mu_j} - \frac{2}{m\bar{\mu} + \sqrt{m^2\bar{\mu}^2 + 4m\bar{\mu}}}\right) \geqslant 0, \tag{30}$$

where $\bar{\mu} = 1/m \sum_{j=1}^{m} \mu_j$. Moreover, Δ_{u-f} is strictly positive if $m > 1$. For the stationary case, Δ_{u-f} becomes

$$\begin{aligned}\Delta_{u-f} = {} & \phi^{2(h-1)} \sum_{j=1}^{m} p_j^2 \left(\frac{\sqrt{(\mu_j + 1 - \phi^2)^2 + 4\mu_j\phi^2} - (\mu_j + 1 - \phi^2)}{2\mu_j}\right.\\ & \left. + \frac{1-\phi^2}{m\bar{\mu}} - \frac{2}{m\bar{\mu} + \phi^2 - 1 + \sqrt{(m\bar{\mu} + \phi^2 - 1)^2 + 4m\bar{\mu}}}\right)\\ & \geqslant 0. \end{aligned} \tag{31}$$

Note that $\Delta_{u-f} \to 0$ if $h \to \infty$.

Proof. The proof is given in the appendix. □

Some comments on this theorem are in order. First note that in both cases the advantage of the factor model increases with the common information: the larger the sum of the factor loadings and the larger the number of series with nonzero factor loadings, the larger the reduction in MSE. Second, the difference decreases with the ratios μ_j which give us the precision at which new information enters into the factor model. It is easy to see that in both cases if $m=1$ $\Delta_{u-f}=0$, that is, in the trivial case of a simple time series the difference is zero. Third in the stationary case the difference increases with ϕ, as expected and when $h\to\infty$, $\Delta_{u-f}\to 0$ and there is no difference between both models. Finally the difference will be in general greater in the nonstationary case.

Note that Christoffersen and Diebold (1998) found that there is no precision gain in the medium and long run forecasts in considering the cointegration relations with respect to the ARIMA univariate models. A reason for their result could be that they did not take into account the number of series that satisfy the long run or cointegration relation. In this analysis, we have found that the number of series is a key factor for the forecasting improvement when explicitly modelling the common trends of the series.

The theorem provides an estimate of the expected decrease in MSE provided for the factor model when we have a large sample and, therefore, using consistent estimates for the parameters, we can assume that the parameter values are approximately known. For instance, consider a large sample generated by the simplest common random walk factor, and assume that $p_j^2=1, \sigma_a^2=\sigma_j^2=1\ \forall j=1,\dots,m$. Then $\theta_j=0.38\ \forall j=1,\dots,m$, and the relative decrease in MSE of the factor model with respect to the univariate models for the one step ahead prediction error is

$$\frac{\Delta_{u-f}}{U}=\frac{0.5\sqrt{3}-0.5-2(m+\sqrt{m^2+4m})^{-1}}{2+0.38/(1-0.38)},$$

which is equal to 0.06 for $m=4$ and goes to 0.14 when $m\to\infty$. These numbers provide some indication of the advantages that we can obtain from the factor model with respect to the univariate forecasts.

5.3. Comparison of pooled and univariate forecasts

Empirically, Garcìa-Ferrer et al. (1987) showed that the univariate forecast of a collection of variables can be improved by introducing a pooling term. In this section, we analyze when we can guarantee that there will be an increase in precision of the pooled forecasts with respect to the univariate ARIMA forecasts when the series are driven by a common factor. Then, we compute the increase in precision in this case.

Suppose that we forecast a vector of time series by using the pooling model (22) and extend this model for the extreme cases $\eta=0$ and 1. Note that for $\eta=0$, this forecast collapses to the univariate ARIMA forecasts and for $\eta=1$, we are using the mean of the univariate forecasts for all the series. The forecast error of the pooling predictions is $y_{j,t+h}-\tilde{y}_{j,t+h}=y_{j,t+h}-\hat{y}_{j,t+h|t}+\eta(\hat{y}_{j,t+h|t}-\hat{y}_t(h))$ and its prediction MSE

$$\begin{aligned}MSE(y_{j,t+h}-\tilde{y}_{j,t+h})&=MSE(y_{j,t+h}-\hat{y}_{j,t+h|t})+\eta^2\mathrm{E}(\hat{y}_{j,t+h|t}-\hat{y}_t(h))^2\\&\quad+2\eta\mathrm{E}[(y_{j,t+h}-\hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t}-\hat{y}_t(h))].\end{aligned}\qquad(32)$$

The first term of the right-hand side in (32) is the prediction MSE of the univariate ARIMA forecasts and it is given by (A.5) in the appendix. The second term measures how far is an univariate forecast from the mean of the univariate forecasts. From (A.3) in the appendix, and substituting $y_{j,t}$ by its expression as a function of the common factor

$$\hat{y}_{j,t+h|t} - \hat{y}_t(h) = \phi^h y_{j,t} - \phi^{h-1}\theta_j v_{j,t} - (\phi^h \bar{y}_t - \phi^{h-1}\theta\bar{v}_t)$$
$$= \phi^h((p_j - \bar{p})f_t + (\varepsilon_{j,t} - \bar{\varepsilon}_t)) - \phi^{h-1}(\theta_j v_{j,t} - \theta\bar{v}_t). \tag{33}$$

If, for instance, the common factor is a random walk, its variance $\mathrm{E}(f_t^2)$ grows linearly with t. It can be easily checked that the term $\mathrm{E}[(y_{j,t+h} - \hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t} - \hat{y}_t(h))]$ in (32) is bounded. Therefore, to guarantee that the pooled forecasts can provide a smaller MSE than the univariate forecasts, the factor loadings must be equal, so $(p_j - \bar{p}) = 0$. Note that this is in agreement with the results of Section 4 in which we showed that when $p_j = p\ \forall j = 1,\ldots,m$ the conditional expectation of the factor model incorporates in all the components the same pooling term with the same weight. For that reason, from now on we will assume this condition. Then, we have the following:

Theorem 3. *For the one factor model, given by (17) and (18) with $p_j = p\ \forall j = 1,\ldots,m$, let us call $P = \sum_{j=1}^m MSE(y_{j,t+h} - \tilde{y}_{j,t+h})$ to the sum of MSE predictions of the pooled forecasts, and U to the sum of MSE of the ARIMA univariate models. The change in precision of the pooled forecasts with respect to the univariate forecasts, $\Delta_{u-p} = U - P$, is given by*

$$\Delta_{u-p} = U - 2\eta\frac{\phi^{2(h-1)}}{m}\sum_{j=1}^{m}(\phi - \theta_j)\sigma_j^2 \sum_{i=1, i\neq j}^{m}\frac{\phi - \theta_i}{1 - \theta_i\theta_j}$$
$$+ \frac{\eta^2\phi^{2(h-1)}}{m^2}\sum_{j=1}^{m}\sum_{i\neq j}^{m}\left([(m-1)\sigma_j^2 + \sigma_i^2]\frac{(\phi - \theta_i)(\phi - \theta_j)}{1 - \theta_i\theta_j}\right.$$
$$\left. + p^2\sum_{k\neq i,j}^{m}\left(\frac{1}{1 - \theta_i\theta_k} - \frac{1}{1 - \theta_i\theta_j}\right)\right).$$

If additionally $\sigma_j^2 = \sigma^2,\ \forall j = 1,\ldots,m$ the increase in precision of the pooled forecasts with respect to the univariate forecasts is given by

$$\Delta_{u-p} = \eta(2 - \eta)(m - 1)\phi^{2(h-1)}\sigma^2\frac{(\phi - \theta)^2}{1 - \theta^2} \geqslant 0. \tag{34}$$

Proof. The proof is given in the appendix. □

The theorem shows that we can only guarantee that we can obtain reductions in the prediction MSE of the pooled forecasts with respect to the ARIMA univariate forecasts if all the series behave equally. If $p_j = p$ and $\sigma_j^2 = \sigma^2\ \forall_j = 1,\ldots,m$, the MA parameters θ_j given by (26) are equal for all the univariate ARMA(1,1) processes and the last term in Δ_{u-p} disappears since the differences $1/(1 - \theta_i\theta_k) - 1/(1 - \theta_i\theta_j) = 0$ for all i, j, k. Then, the difference of prediction MSE between the ARIMA univariate model and the pooled forecasts is always positive (as it is shown by 34) and the prediction MSE using pooled forecasts reaches its minimum for $\eta = 1$, giving the highest possible weight to the pooling term. In this case, the best predictor is the mean of the univariate ARIMA forecasts $\tilde{y}_{j,t+h} = 1/m \sum_{i=1}^{m} \hat{y}_{i,t+h|t}$. From (34) we can also conclude that for this restrictive case, the difference of prediction MSE between the ARIMA univariate model and the pooled forecasts increases with the number of series, with the value of the AR parameter and with the weight given to the pooling term. The difference between both models gets smaller as the precision μ decreases, then $\theta \to \phi$, the second term in (34) vanishes and the univariate series tend to behave as white noise and nothing is gain, in terms of precision, from using a linear combination of them.

So, in order to obtain an increase in precision with the pooled forecasts with respect to the univariate ARIMA forecasts, the factor loadings and the errors variances of each individual series should be very close. Nevertheless, our limited experience with Monte Carlo simulations reveals that the condition of equal noise variances for the observed series is not so important, while the condition of equal factor loadings is crucial for obtaining these results in practice. This will be confirmed through some simulations in the next section.

5.4. Pooled versus one factor model forecasts

The increase in precision of the pooled forecasts is only guaranteed if the series come from similar processes. In this later case (the most favorable for the pooling technique), the following theorem provides the increase in precision of the factor model with respect to the pooled forecasts.

Theorem 4. *For the one factor model given by (17) and (18), with $\mu_j = \mu \ \forall_j = 1,\dots,m, \mu_j \neq 0$, let $\Delta_{p-f} = P - F = \sum_{j=1}^{m}(MSE(y_{j,t+h} - \tilde{y}_{j,t+h})) - tr(\mathbf{\Sigma}_{t+h|t})$ be the increase in precision of the factor model with respect to the univariate pooled forecasts. For the nonstationary case*

$$\Delta_{p-f} \geqslant \Delta_{u-f} - (m-1)\sigma^2 \frac{1-\theta}{1+\theta} \tag{35}$$

$$= p^2 \left(\frac{1}{2} \left(\sqrt{1 + \frac{4}{\mu}} - 1 \right) \frac{m\theta + 1}{1+\theta} - \frac{2}{\mu + \sqrt{\mu^2 + 4\frac{\mu}{m}}} \right) \geqslant 0. \tag{36}$$

Δ_{p-f} is strictly positive if $m > 1$. For the stationary case, Δ_{p-f} becomes

$$\Delta_{p-f} \geqslant \Delta_{u-f} - (m-1)\phi^{2(h-1)} \frac{p^2(\phi-\theta)\theta}{(1-\theta^2)(1-\phi\theta)} \tag{37}$$

$$= p^2 m \left(\frac{\theta}{\phi(1-\phi\theta)}\beta - \frac{1}{Vm\mu\phi^2} + \frac{1-\theta^2}{\phi^2 m\mu} \right) \phi^{2h}, \tag{38}$$

where $\beta = 1 - ((m-1)/m)((\phi-\theta)/\phi(1-\theta^2))$. Note that $\Delta_{p-f} \to 0$ if $h \to \infty$.

Proof. The proof is given in the appendix. □

Two equivalent expressions ((35) and (36) for the nonstationary case and (37) and (38) for the stationary one) have been given in the theorem for the lower bound of the increase in forecasting precision from using the factor model rather than the univariate pooled forecasts. Eqs. (35) and (37) show that for the trivial case of one series there is no difference between the univariate and pooled forecasts. In this case $\Delta_{p-f} = \Delta_{u-f} = 0$. Expressions (36) and (38) are useful to study the dependence of the results on the parameters. The $\geqslant$ signs comes from the fact that under the hypothesis of the theorem, the best pooled forecasts, in terms of prediction MSE, are given for $\eta = 1$, using as pooled forecasts the means of the univariate forecasts. For any other value of η, $0 \leqslant \eta \leqslant 1$, the increase in forecast precision when using the factor model will be greater if $m > 1$. In both cases, the stationary and nonstationary ones, the equality holds for $\eta = 1$. A few other remarks are in order.

First note that in both cases the advantage of the factor model increases with the number of series m and the common information through the square of the factor loading p. Second, in the nonstationary case the advantages of the pooling forecasts decrease if μ decrease and in the limit case when $\mu \to 0$, $\theta \to \phi = 1$, $\Delta_{p-f} \to \Delta_{u-f}$ and the pooled forecasts do not show any advantages over the univariate ARIMA forecasts. Notice that in this case the univariate series will behave as white noise and there is no improvement in terms of the prediction MSE with the pooled forecasts with respect to the ARIMA univariate forecasts. Third, in the stationary case the difference increases with ϕ, as expected and when $h \to \infty$, $\Delta_{p-f} \to 0$ and there is no difference between both models. Fourth, notice also that in this case, the first term is always positive since θ and ϕ are of the same sign and β is always positive (since $|\theta| \leqslant |\phi|$ and $(\phi-\theta)/\phi(1-\theta^2) = ((1-\theta)/\phi)/(1-\theta^2) = ((1-\theta)/\phi)/(1-|\theta|)(1+|\theta|) < 1/(1+|\theta|) < 1$).

As it was previously pointed out, this is the best result we can obtain for pooled forecasts, when series behave in a similar way giving the highest weight to the pooled term. In any other case, the advantages of the factor model over the pooled forecasts should be greater. This will be confirmed in the next section through some simulation results.

6. Some simulation results

The previous analysis has been obtained by assuming that the parameters are known. To check these results in the usual case in which the parameters are estimated from the sample, we have carried out three Monte Carlo experiments. In the three experiments we compare the forecasts generated by five different models. First, we fitted an ARI(3,1) for each of the series as an approximation to the true univariate model, ARIMA(0,1,1), and computed the univariate forecasts. Then we computed pooled forecasts using (22) and the three values of η, ($\eta = 0.25, 0.50$ and 0.75), which were used by Garcìa-Ferrer et al. (1987). Finally, we have estimated the factor model through the EM algorithm via the Kalman filter. For all these models we made forecasts $h = 1, \ldots, 20$ steps ahead and computed the prediction MSE for horizons 1, 5, 10 and 20. We made 1000 replications.

Table 1
Comparison of RMSE of prediction for the univariate ARIMA, pooled and factor models

Horiz.	μ	UNIV	$\eta = 0.25$	$\eta = 0.50$	$\eta = 0.75$	FACT
$h = 1$	0.1	1.1877	1.1466	1.1165	1.0984	1.1097
$h = 5$	0.1	1.3515	1.3114	1.2817	1.2633	1.2638
$h = 10$	0.1	1.5089	1.4745	1.4499	1.4355	1.4380
$h = 20$	0.1	1.8066	1.7768	1.7552	1.7421	1.7406
$h = 1$	1	1.6567	1.6066	1.5701	1.5481	1.5337
$h = 5$	1	2.5751	2.5420	2.5181	2.5034	2.4977
$h = 10$	1	3.3195	3.2942	3.2760	3.2650	3.2668
$h = 20$	1	4.5538	4.5362	4.5239	4.5168	4.5164
$h = 1$	10	3.5331	3.4892	3.4581	3.4403	3.4566
$h = 5$	10	7.3283	7.3056	7.2892	7.2793	7.2650
$h = 10$	10	9.8499	9.8328	9.8206	9.8133	9.8297
$h = 20$	10	14.2844	14.2722	14.2633	14.2578	14.2339

Data consist of 4 time series generated by a common random walk for three different values of μ=0.1, 1, 10.

The sample size for each of the simulated series was 124 observations. The first 104 observations were used to estimate the model and the last 20 were reserved to compute the forecasts.

In the first experiment we assume the most favorable situation for shrinking forecasts: a common nonstationary factor (common trend) for $m = 4$ observed series, $\mathbf{P} = p(1,1,1,1)'$, $\mathrm{E}(\varepsilon_t) = \mathbf{0}$, $var(\varepsilon_t) = \mathbf{I}_4$, $\mathrm{E}(a_t) = 0$ and $var(a_t) = \sigma_a^2 = 1$. We have used three different values of $p\,(1, \sqrt{10}$ and $\sqrt{0.1})$. In this case $\mu_j = \mu\ \forall j = 1,\dots,m$ and $\mu = p^2$. Notice that this is the same as keeping $p = 1$ in all cases and letting the noise variance of the factor to vary according to $\sigma_a^2 = 1, 10, 0.1$.

Table 1 shows the root mean square error (RMSE) of prediction for $\mu = 0.1, 1$ and 10. Each model is characterized by a single value of μ, which indicates that all the series behave in a similar way within a model. The first column shows the horizon of prediction; the second one shows the precision of generation of each series, μ; and columns third to seventh show the RMSE obtained for the univariate ARI(3,1), for the pooling technique with the three values of η previously indicated, and for the factor model.

Note that for $\mu = 1$ the factor model improves the univariate forecasts by $(1.6567 - 1.5337)/1.6567 = 0.0742$ for $h = 1$, by $(3.3195 - 3.2668)/3.3195 = 0.0159$ for $h = 10$ and by $(4.5538 - 4.5164)/4.5538 = 0.0082$ for $h = 20$. The best forecasts from the pooling method correspond to $\eta = 0.75$ and the gains with respect to the univariate forecasts is, for one step ahead, $(1.6567 - 1.5481)/1.6567 = 0.0656$ similar to the one from the factor model. We see that the pooling forecast with $\eta = 0.75$ performs very similar to the factor model for all horizons.

In the second experiment we allow different precisions μ_i. First we take $var(\varepsilon_t) = diag(0.1^2, 0.5^2, 1^2, 3^2)$ and set the other parameters as in the first experiment. Second, we additionally change $\mathbf{P} = (1, 0.5, 0.2, 0.05)'$. The results of this second experiment are

Table 2
Comparison of RMSE of prediction for the univariate ARIMA, pooled and factor models

Horiz.	p	UNIV	$\eta = 0.25$	$\eta = 0.50$	$\eta = 0.75$	FACT
$h = 1$	E	2.1451	2.0675	2.0103	1.9752	1.9502
$h = 5$	E	2.9308	2.8749	2.8344	2.8100	2.7799
$h = 10$	E	3.5882	3.5410	3.5066	3.4855	3.4585
$h = 20$	E	4.7594	4.7251	4.7007	4.6862	4.6653
$h = 1$	D	1.8987	1.9819	2.3761	2.9595	1.7206
$h = 5$	D	2.2157	2.2681	2.6030	3.1310	2.0569
$h = 10$	D	2.5147	2.5373	2.8198	3.2962	2.3537
$h = 20$	D	3.1215	3.1366	3.3663	3.7716	2.9916

Data consist of 4 time series generated by a common random walk factor with different values of the noise variances (rows 2–5) and with different factor loadings and noise variances of the observed series (rows 6–9).

given in Table 2. The first part of the table indicates the results for the first simulation in which the factor loadings are equal, and this is indicated by the E in the second column. The second part of the table shows the result when the loadings are different, and this is indicated by D in the second column. The rest of the table has the same structure as Table 1.

The results of Table 2 show that when the series have different factor loadings and noise variances (rows 6–9 of the table), the pooling forecast are no longer successful. The best results are obtained when the pooling is the smallest, $\eta = 0.25$, and as η increases the behavior of the pooling method worsens. However, when the factor loadings are equal, the results (rows 2–5) are qualitative similar to the ones of Table 1, and the models rank in the same way.

The third experiment was designed to check if the previous results can be generalized to the case of several factors. We assume $r = 3$ common factors and $m = 8$ observed series. Three models are considered and in all of them the factor loading matrix is $\mathbf{P} = [\mathbf{p}_1 \vdots \mathbf{p}_2 \vdots \mathbf{p}_3]$, where $\mathbf{p}_2 = (0, 1, 1, -1, -1, 0, 0, 0)'$ and $\mathbf{p}_3 = (0, 0, 0, 1, -1, 1, -1, 0)'$, $\mathrm{E}(\varepsilon_t) = \mathbf{0}$; the dynamics of the common factors are given by Eq. (2) with $\mathbf{\Phi}_1 = diag(1, 0.7, 0.3)$, $\mathbf{\Phi}_j = \mathbf{0}$, $j > 1$ and $\mathbf{\Theta}_i = \mathbf{0}$ for all i, $\mathrm{E}(\mathbf{a}_t) = \mathbf{0}$ and $var(\mathbf{a}_t) = \mathbf{I}_3$. The first model, Model I, has $\mathbf{p}_1 = (1, 1, 1, 1, 1, 1, 1, 1)'$ and $var(\varepsilon_t) = \mathbf{I}_8$. The second model, Model II, has also $\mathbf{p}_1 = (1, 1, 1, 1, 1, 1, 1, 1)'$ but now $var(\varepsilon_t) = diag(0.01^2, 0.05^2, 0.1^2, 0.5^2, 1, 1.5^2, 3^2, 5^2)$. The third model, Model III, has different factor loadings associated to the common nonstationary factor, $\mathbf{p}_1 = (1, -1, 1, -1, 1, -1, 1, -1)'$ and $var(\varepsilon_t) = diag(0.01^2, 0.05^2, 0.1^2, 0.5^2, 1, 1.5^2, 3^2, 5^2)$.

The results are presented in Table 3. Note that for Model I, in which the common trend affects in the same way to all the components the results are similar to the one factor model. The factor model improves the univariate forecasts by 0.083 for $h = 1$, by 0.0451 for $h = 10$ and by 0.0270 for $h = 20$. The best forecasts from the pooling method correspond to $\eta = 0.75$ and the gains with respect to the univariate forecasts is, for one step ahead, 0.0617 similar to the one from the factor model. The same qualitative results

Table 3
Comparison of RMSE of prediction for the univariate ARIMA, pooled and factor models

Horiz.	Model	UNIV	$\eta = 0.25$	$\eta = 0.50$	$\eta = 0.75$	FACT
$h = 1$	I	1.9783	1.8883	1.8466	1.8563	1.8141
$h = 5$	I	3.0446	2.9447	2.8745	2.8363	2.8475
$h = 10$	I	3.8830	3.7965	3.7355	3.6951	3.7080
$h = 20$	I	5.0103	4.9426	4.8933	4.8631	4.8751
$h = 1$	II	2.9434	2.8169	2.7424	2.7240	2.6720
$h = 5$	II	3.7244	3.5906	3.4963	3.4449	3.4768
$h = 10$	II	4.4711	4.3478	4.2580	4.2039	4.2355
$h = 20$	II	5.4835	5.3799	5.3040	5.2571	5.2958
$h = 1$	III	2.9424	8.7517	16.6965	24.8033	2.6330
$h = 5$	III	3.7175	9.0849	16.8978	24.9564	3.4178
$h = 10$	III	4.4546	9.3108	19.9656	24.9652	4.1606
$h = 20$	III	5.5538	9.6196	16.9859	24.8757	5.2774

Data consist of 8 time series generated by 3 common factors.

are obtained from Model II, confirming the results of the second experiment. However, for model III, the results from the pooled forecasts greatly deteriorates, as expected. Recall from the results of Section 5.3, Eq. (33) that there is an unbounded term in the expression of the MSE of the pooled forecasts, when there is a nonstationary common factor with different values of the factor loadings for each of the series associated to it.

7. An example: forecasting GNP

The data we considered are annual observations of the real GNP, from 1949 to 1997, for some European OECD countries. An extended data base, from 1948 to 1986, was analyzed by Garcìa-Ferrer et al. (1987), who considered several alternatives to forecast the output growth rates defined as $g_t = \ln(O_t/O_{t-1})$, where O_t is real GNP, for several OECD countries. Forecasts were compared by the root mean-square error of prediction for one step ahead forecasts. The problem was further studied by Zellner and Hong (1989), Mittnik (1990), Zellner et al. (1991), Min and Zellner (1993), Li and Dorfman (1996), Zellner and Min (1998) and Garcìa-Ferrer and Poncela (2002).

A factor model with a common trend and a common stationary factor was built for this European group of countries that includes Belgium, France, Italy, the Netherlands and Spain. A graph of the logs of real GNP of these countries is shown in Fig. 1.

The difficulty in forecasting this data set is associated to the presence of several turning points. We will show that the prediction MSE decreases in a dynamic factor model with respect to an ARIMA univariate model and pooled forecasts, when increasing the number of countries considered. Each of the models was estimated with data from 1949 to 1980, then we generated one step ahead forecasts. We reestimated the models adding one observation at the time and made new forecasts. Finally, we compute the RMSE of prediction for each country.

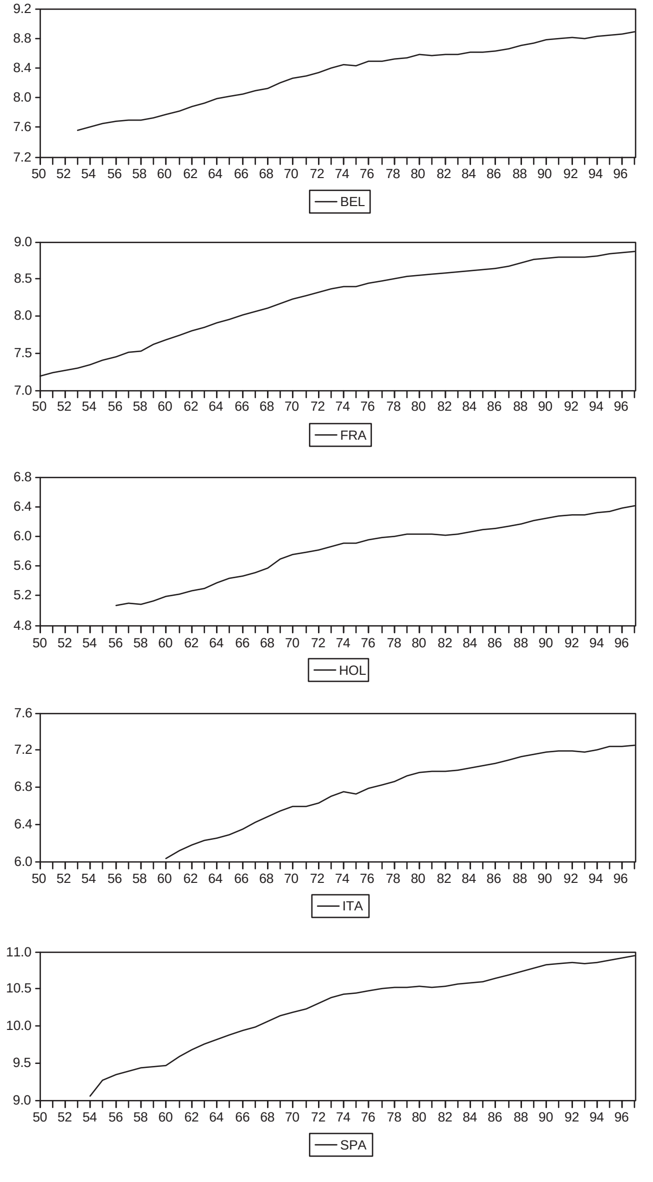

Fig. 1. Logs of real GNP of Belgium, France, Holland, Italy and Spain.

In order to achieve a systematic procedure for our comparison, (see, for instance, Garcìa-Ferrer et al., 1987, among others), we fitted and AR(3) model for each growth rate,

$$g_{it} = \beta_{0i} + \beta_{1i} g_{it-1} + \beta_{2i} g_{it-2} + \beta_{3i} g_{it-3} + \varepsilon_{it}. \tag{39}$$

Table 4
RMSE of prediction for each country for a factor model built by using 5, 4, 3 and 2 series

m	BEL	SPA	FRA	HOL	ITA	Mean	Median
5	1.27	1.52	1.20	1.96	1.48	1.49	1.48
4	1.45	1.57	1.31	1.90		1.56	1.51
3	1.75	1.51	1.40			1.56	1.51
2	1.82	2.55				2.18	2.18

Then, we built a factor model. Applying the results in Peña and Poncela (2002), we found a common trend plus a common AR(1) stationary factor. Let $\mathbf{y}_t=(y_{1,t},\ldots,y_{m,t})'$, $y_{i,t}=\log(O_{i,t})$, $i=1,\ldots,m$, $t=1,\ldots,T$, the factor model could be written as $\mathbf{y}_t=\mathbf{P}\mathbf{f}_t+\mathbf{n}_t$, where $\mathbf{f}_t$ is the r-dimensional vector of *common factors*, $\mathbf{P}$ is the factor loading matrix, and $\mathbf{n}_t$ is the vector of *specific components*. In our case, $\mathbf{f}_t' = [T_t' \vdots f_{2,t}']$, where T_t is a common trend and $f_{2,t}$, is an AR(1) stationary common factor,

$$\begin{bmatrix} T_t \\ f_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} T_{t-1} \\ f_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \tag{40}$$

where $\mathbf{a}_t=(a_{1,t}\ a_{2,t})' \sim \mathrm{N}_2(\mathbf{0},\mathbf{\Sigma}_a)$ is serially uncorrelated, $\mathrm{E}(\mathbf{a}_t\mathbf{a}_{t-h}')=\mathbf{0},\ h \neq 0$. After extracting the common dynamic structure, we fitted an univariate AR(3) for each of the specific components in order to capture the remaining dynamic structure

$$n_{it} = \alpha_{0i} + \alpha_{1i}n_{it-1} + \alpha_{2i}n_{it-2} + \alpha_{3i}n_{it-3} + e_{it}. \tag{41}$$

The sequence of vectors $\mathbf{e}_t=(e_{1,t},\ldots,e_{m,t})'$ are normally distributed, have zero mean and diagonal covariance matrix $\mathbf{\Sigma}_e$. We assume that the noises from the common factors and specific components are also uncorrelated for all lags, $\mathrm{E}(\mathbf{a}_t\mathbf{e}_{t-h}')=\mathbf{0}\ \forall h$.

We repeated the process discarding one country at the time. The series were discarded in Spanish alphabetical order (Spain starts with ESP in Spanish). The results are shown in Table 4. The first column has the number of series considered in each of the factor models. Columns 2–6 have the RMSE of prediction for each of the countries with each of the factor models and columns 6 and 7 show the mean and median of the RMSE of prediction of each model.

It is clear that the mean and median of the RMSE of prediction decreases with the number of series. By using 5 series the mean RMSE decreases 31.65% with respect to the case in which only 2 series are used. Pooled forecasts were also built for values of η equal to 0.25, 0.50 and 0.75. Table 5 shows a comparison of the factor model, the ARI(3,1) model and the pooled forecasts. It is clearly seen in Table 5 that the best forecasting results are achieved through the factor model. Also, when we increase the value of η, from 0.25 to 0.75, giving greater weight to the pooling term the mean and, therefore, the sum of RMSE of predictions slightly decreases. The ARI(3,1) model gives the overall worse results.

Finally, we want to check the influence of the number of series m in the forecasting performance of the different models and if these models rank in a similar way when we decrease the number of series. In order to do so, we will repeat Table 5, discarding one

Table 5
RMSE of prediction for each country for the factor model (FM) with five series, pooled forecasts (P) for three values of η, 0.75, 0.50 and 0.25 and univariate ARI(3,1) model

Model	BEL	SPA	FRA	HOL	ITA	Mean	Median
FM	1.27	1.52	1.20	1.96	1.48	1.49	1.48
P, $\eta = 0.75$	1.86	1.69	1.64	1.79	1.82	1.76	1.79
P, $\eta = 0.50$	1.88	1.62	1.63	1.77	1.93	1.77	1.77
P, $\eta = 0.25$	1.90	1.57	1.63	1.77	2.06	1.79	1.77
ARI(3,1)	1.92	1.54	1.64	1.77	2.22	1.82	1.77

Table 6
RMSE of prediction for each country for the factor model (FM) with four series, pooled forecasts (P) for three values of η, 0.75, 0.50 and 0.25 and univariate ARI(3,1) model

Model	BEL	SPA	FRA	HOL	Mean	Median
FM	1.45	1.57	1.31	1.90	1.56	1.51
P, $\eta = 0.75$	1.79	1.66	1.59	1.72	1.69	1.69
P, $\eta = 0.50$	1.83	1.61	1.60	1.73	1.69	1.67
P, $\eta = 0.25$	1.87	1.55	1.62	1.75	1.70	1.69
ARI(3,1)	1.92	1.54	1.64	1.77	1.72	1.71

Table 7
RMSE of prediction for each country for the factor model (FM) with three series, pooled forecasts (P) for three values of η, 0.75, 0.50 and 0.25 and univariate ARI(3,1) model

Model	BEL	SPA	FRA	Mean	Median
FM	1.75	1.51	1.40	1.56	1.51
P, $\eta = 0.75$	1.80	1.67	1.60	1.69	1.67
P, $\eta = 0.50$	1.83	1.61	1.60	1.68	1.61
P, $\eta = 0.25$	1.88	1.57	1.62	1.69	1.62
ARI(3,1)	1.92	1.54	1.64	1.70	1.64

series at the time. The series are discarded in Spanish alphabetical order. The results are shown in Tables 6–8.

For all the cases but the last one, the factor model outperforms the remaining univariate models and the different models rank in a similar way as in the case of $m = 5$ series. Only in the last case, when there are only two series, the behavior of the factor model deteriorates and all the univariate models provide very similar results and better than the factor model.

Table 8
RMSE of prediction for each country for the factor model (FM) with two series, pooled forecasts (P) for three values of η, 0.75, 0.50 and 0.25 and univariate ARI(3,1) model

Model	BEL	SPA	Mean	Median
FM	1.82	2.55	2.18	2.18
P, $\eta = 0.75$	1.82	1.65	1.74	1.74
P, $\eta = 0.50$	1.85	1.60	1.73	1.73
P, $\eta = 0.25$	1.88	1.57	1.73	1.73
ARI(3,1)	1.92	1.54	1.73	1.73

8. Conclusions

We have shown that the forecasting equations for each component of a vector of time series that follows a factor model incorporate a pooling term of a weighted sum of all the variables observed in t. In particular, for the one factor model and under some very restrictive assumptions we can obtain as the pooling term the sample mean of the observed series.

For the AR(1) one factor model and assuming that the parameters are known, we have shown that the gain in precision, in terms of the prediction MSE, of the factor model with respect to univariate ARIMA and pooled forecasts depends on the common information and increases with the number of time series and the sum of the relative sizes of the factor loadings. Thus, we have that when the parameters are known and under some restrictive assumptions (equal factor loadings and equal noise variance for the observed series)

$$MSE_{\mathrm{u}} > MSE_{\mathrm{p}} > MSE_{\mathrm{f}}, \tag{42}$$

where MSE_{f} is the trace of MSE of predictions of the factor model, and MSE_{u} and MSE_{p} are the sum of MSE of prediction of the univariate ARIMA models and pooled forecasts. If there exists a common nonstationary factor with different loadings associated to it, the expression for MSE_{p} is unbounded and the pooled forecasts can be worse than the univariate forecasts, although MSE_{u} and MSE_{p} always remain bigger than MSE_{f} (if we discard the estimation error). This was confirmed by Monte Carlo experiments. Our limited Monte Carlo experience reveals that the result $MSE_{\mathrm{u}} > MSE_{\mathrm{p}}$ depends much more on the hypothesis of equal factor loadings than on the one of equal noise variances.

A limited Monte Carlo experiment seems to reveal that we can broadly draw the same conclusions for the multifactor model. The larger the common information, the larger the advantages of the factor model with respect to the univariate forecasts. The advantages of pooling forecast with respect to univariate forecast can be important if there is a main factor with similar effects on all the series, but this advantage may disappear if the loading coefficients of the main factor are very different for the different series. This result is not surprising because then a common pooling term will be unable to approximate the conditional expectation for each component given the past information for all of them.

Acknowledgements

The authors acknowledge financial support from research projects BEC2000-0167 and BEC2002-00081. We also acknowledge Prof. A. Garcìa-Ferrer for very helpful comments.

Appendix

Proof of (16). Starting from (14) and using (6) we have that

$$\mathbf{A}_2 = \tilde{\mathbf{P}}\mathbf{G}^h\mathbf{V}_{t|t-1}\tilde{\mathbf{P}}'(\boldsymbol{\Sigma}_\varepsilon^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\tilde{\mathbf{P}}(\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_\varepsilon^{-1}\tilde{\mathbf{P}} + \mathbf{V}_{t|t-1}^{-1})^{-1})\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_\varepsilon^{-1})$$

and calling $\mathbf{A}_2 = \tilde{\mathbf{P}}\mathbf{N}_t\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_\varepsilon^{-1}$, we can write $\mathbf{N}_t = \mathbf{G}^h\mathbf{V}_{t|t-1}(\mathbf{I}_s - \tilde{\mathbf{P}}'\boldsymbol{\Sigma}_\varepsilon^{-1}\tilde{\mathbf{P}}(\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_\varepsilon^{-1}\tilde{\mathbf{P}} + \mathbf{V}_{t|t-1}^{-1})^{-1})$ and computing again the inverse of $(\tilde{\mathbf{P}}'\boldsymbol{\Sigma}_\varepsilon^{-1}\tilde{\mathbf{P}} + \mathbf{V}_{t|t-1}^{-1})$, we obtain (16). □

Proof of Lemma 1. We first proof the following auxiliary lemma that characterizes the noise processes involved and will be used in the proof of other results of Section 5 as well.

Auxiliary lemma. *For the one factor model given by* (17) *and* (18), *and for* $v_{i,t}$ *defined in Eq.* (24), *if* (i) $\mathrm{E}(a_t v_{j,0}) = 0$ *and* (ii) $\mathrm{E}(\varepsilon_{j,t} v_{i,0}) = 0$ $\forall j, i = 1, \ldots, m$, *then*

1. $Var(v_{j,t}) = \tilde{\sigma}_j^2 = \sigma_j^2 + p_j^2/(1 - \phi\theta_j)$ $\forall j = 1, \ldots, m$,
2. $\mathrm{E}(a_t v_{j,\tau}) = 0$ $\forall j = 1, \ldots, m$ $\forall \tau < t$,
3. $\mathrm{E}(a_t v_{j,t}) = p_j$ $\forall j, i = 1, \ldots, m,$ $j \neq i$,
4. $\mathrm{E}(\varepsilon_{j,t} v_{j,t}) = \sigma_j^2$ $\forall j = 1, \ldots, m$,
5. $\mathrm{E}(\varepsilon_{j,t} v_{j,t+h}) = -\theta_j^{h-1}(\phi - \theta_j)\sigma_j^2$ $\forall j = 1, \ldots, m$ $\forall h > 0$,
6. $\mathrm{E}(\varepsilon_{j,t} v_{i,\tau}) = 0$ $\forall j, i = 1, \ldots, m,$ $j \neq i$ $\forall \tau$ *integer*,
7. $Cov(v_{j,t}, v_{i,t}) = \tilde{\sigma}_{ji} = p_i p_j/(1 - \theta_j\theta_i)$ $\forall j, i = 1, \ldots, m,$ $i \neq j$ *and*
8. $Cov(v_{j,t+h}, v_{i,t}) = \theta_j^h p_i p_j/(1 - \theta_j\theta_i)$ $\forall j, i = 1, \ldots, m,$ $i \neq j$ $\forall h > 0$.

Proof of the auxiliary lemma. 1. By (23) and (24)

$$p_j a_t + \varepsilon_{j,t} - \phi\varepsilon_{j,t-1} = v_{j,t} - \theta_j v_{j,t-1} \tag{A.1}$$

and equating variances in both sides

$$\widetilde{\sigma}_j^{\,2} + \theta_j^2\widetilde{\sigma}_j^{\,2} = p_j^2 + \sigma_j^2 + \phi^2\sigma_j^2 \tag{A.2}$$

and for the equality of the first-order autocovariances, $\theta_j\widetilde{\sigma}_j^{\,2} = \phi\sigma_j^2$. From both equations we obtain that θ_j must satisfy (26), also solving for $\widetilde{\sigma}_j^{\,2}$ then $\widetilde{\sigma}_j^{\,2} = p_j^2 + (1 + (\phi - \theta_j)\phi)\sigma_j^2$. Introducing (28) in the last equation $\widetilde{\sigma}_j^{\,2} = \sigma_j^2 + (1 + \phi\theta_j/(1 - \phi\theta_j))p_j^2 = \sigma_j^2 + p_j^2/(1 - \phi\theta_j)$.

2. To show that $\mathrm{E}(a_t v_{j,\tau}) = 0$, we solve for $v_{j,\tau}$ in (A.1) and by backward substitution we get that $\mathrm{E}(a_t v_{j,\tau}) = \theta^{\tau}\mathrm{E}(a_t v_{j,0}) = 0$, by hypothesis (i).

3. We solve for $v_{j,\tau}$ in (A.1) and introduce it in $\mathrm{E}(a_t v_{j,t}) = \mathrm{E}(a_t(p_j a_t + \varepsilon_{j,t} - \phi\varepsilon_{j,t-1} + \theta_j v_{j,t-1})) = p_j$ $\forall j = 1, \ldots, m$.

4. $\mathrm{E}(\varepsilon_{j,t}v_{j,t}) = \mathrm{E}(\varepsilon_{j,t}(p_j a_t + \varepsilon_{j,t} - \phi\varepsilon_{j,t-1} + \theta_j v_{j,t-1})) = \sigma_j^2 \;\forall j = 1,\dots,m.$

5. Applying (A.1) in time $j+h$ and by backward substitution $v_{j,t+h}$ h times, $\mathrm{E}(\varepsilon_{j,t}v_{j,t+h}) = -\theta_j^{h-1}(\phi-\theta_j)\sigma_j^2 \;\forall j,i = 1,\dots,m,\; j\neq i \;\forall j = 1,\dots,m \;\forall h>0.$

6. To show $\mathrm{E}(\varepsilon_{j,t}v_{i,\tau}) = 0 \;\forall j,i = 1,\dots,m,\; j\neq i \;\forall\tau$ integer, is immediately from backward substitution of $v_{i,\tau}$ τ times from (A.1) and applying hypothesis (ii).

7. From (A.1) $\widetilde{\sigma_{ji}} = \mathrm{E}(v_{j,t}, v_{i,t}) = \mathrm{E}[(p_j a_t + \varepsilon_{j,t} - \phi\varepsilon_{j,t-1} + \theta_j v_{j,t-1})(p_i a_t + \varepsilon_{i,t} - \phi\varepsilon_{i,t-1} + \theta_i v_{i,t-1})] = p_j p_i + \theta_j\theta_i\widetilde{\sigma_{ji}}$. Solving for $\widetilde{\sigma_{ji}}$, we get $\widetilde{\sigma_{ji}} = p_i p_j/(1-\theta_j\theta_i) \;\forall j,i = 1,\dots,m$, $j\neq i$.

8. Applying recursively (A.1) h times and from Parts 2 and 6 of this lemma, $\mathrm{E}(v_{j,t+h}, v_{i,t}) = \mathrm{E}[(p_j a_{t+h} + \varepsilon_{j,t+h} - \phi\varepsilon_{j,t+h-1} + \theta_j v_{j,t+h-1})v_{i,t}] = \theta_j \mathrm{E}(v_{j,t+h-1}v_{i,t}) = \theta_j^h \mathrm{E}(v_{j,t}v_{i,t}) = \theta_j^h(p_i p_j/(1-\theta_j\theta_i))$. □

We proceed now to prove Lemma 1. The forecast of the observed series h steps ahead is given by

$$\hat{y}_{j,t+h|t} = \phi^h y_{j,t} - \phi^{h-1}\theta_j v_{j,t} \tag{A.3}$$

and the true value in $t+h$ can be written as $y_{j,t+h} = \phi^h y_{j,t} + v_{j,t+h} + \sum_{i=1}^{h-1}\phi^{i-1}(\phi-\theta_j)v_{j,t+h-i} - \phi^{h-1}\theta_j v_{j,t}$, so the forecast error is

$$y_{j,t+h} - \hat{y}_{j,t+h|t} = v_{j,t+h} + \sum_{i=1}^{h-1}\phi^{i-1}(\phi-\theta_j)v_{j,t+h-i} \tag{A.4}$$

and the prediction MSE

$$\begin{aligned} MSE_j = MSE(y_{j,t+h} - \hat{y}_{j,t+h|t}) &= \widetilde{\sigma}_j^{\,2}\left(1 + (\phi-\theta_j)^2\sum_{i=1}^{h-1}\phi^{2(i-1)}\right) \\ &= \left(\sigma_j^2 + \frac{p_j^2}{1-\phi\theta_j}\right)\left(1 + (\phi-\theta_j)^2\sum_{i=1}^{h-1}\phi^{2(i-1)}\right), \end{aligned} \tag{A.5}$$

where the last equality is obtained replacing $\widetilde{\sigma}_j^{\,2}$ by its expression given in the auxiliary lemma, Part 1. Then, the sum of MSE for all the series, $U = \sum_{j=1}^m (MSE)_j$, will be

$$U = \sum_{j=1}^m\left(\sigma_j^2 + \frac{p_j^2}{1-\phi\theta_j}\right)\left(1 + (\phi-\theta_j)^2\sum_{i=1}^{h-1}\phi^{2(i-1)}\right)$$

and using (28) and after some straightforward algebra we finally have

$$U = \sum_{j=1}^m\left(\sigma_j^2 + \frac{p_j^2\theta_j}{\phi(1-\phi\theta_j)} + \frac{\phi-\theta_j}{(1-\theta_j\phi)\phi}p_j^2\sum_{i=0}^{h-1}\phi^{2i}\right). \qquad \square \tag{A.6}$$

Proof of Theorem 2. For the one factor model, substituting (9) in (11), with $\mathbf{\Sigma}_u = 1$ by the identification restriction and $\mathbf{G} = \phi$, we obtain that the trace of the MSE of predictions matrix is

$$tr(\mathbf{\Sigma}_{t+h|t}) = \sum_{j=1}^m p_j^2\left(V_{t|t}\phi^{2h} + \sum_{i=0}^{h-1}\phi^{2i}\right) + \sum_{j=1}^m\sigma_j^2. \tag{A.7}$$

The difference between (29) and (A.7), after some straightforward algebra, is

$$\Delta_{u-f} = \sum_{j=1}^{m} p_j^2 \left(\frac{\theta_j}{\phi(1-\phi\theta_j)} - V_{t|t}\phi^{2h} \right) - \sum_{j=1}^{m} \left(p_j^2 \frac{\theta_j(1-\phi^2)}{(1-\theta_j\phi)\phi} \sum_{i=0}^{h-1} \phi^{2i} \right), \quad \text{(A.8)}$$

that can be written, both for $|\phi| = 1$ as well as for $|\phi| < 1$ as

$$\Delta_{u-f} = \sum_{j=1}^{m} p_j^2 \left(\frac{\theta_j}{\phi(1-\phi\theta_j)} - V_{t|t} \right) \phi^{2h}. \quad \text{(A.9)}$$

To compute Δ_{u-f}, first notice that the model written in state space form is a detectable and stabilisable system. Therefore, it reaches a steady state for any initial condition $V_{1|0}$ of the Kalman filter (see, for instance, Harvey, 1989, p. 119). Pre- and post-multiplying $\boldsymbol{\Sigma}_{t|t-1}^{-1}$ by $\mathbf{P}'$ and $\mathbf{P}$ and from (6) and using again the lemma for the inverse of the sum of two matrices (Rao, 1973), $\mathbf{P}'\boldsymbol{\Sigma}_{t|t-1}^{-1}\mathbf{P} = m\bar{\mu}/(1+m\bar{\mu}V_{t|t-1})$. Now, from (8), $V_{t|t}$ can be written as $V_{t|t} = V_{t|t-1}/(1+m\bar{\mu}V_{t|t-1})$. Substituting $V_{t-1|t-1}$ as given by (8) in (5) and assuming that the filter has already reached the steady state, $V_{t|t-1} = V_{t-1|t-2} = V$, we obtain the algebraic Riccati equation

$$m\bar{\mu}V^2 - (m\bar{\mu} + \phi^2 - 1)V - 1 = 0 \quad \text{(A.10)}$$

and the steady state of $V_{t|t}$, denoted by $\tilde{V}$, is

$$\tilde{V} = \frac{V}{1+m\bar{\mu}V} = \frac{1}{Vm\bar{\mu}\phi^2} - \frac{1-\phi^2}{m\bar{\mu}\phi^2}, \quad \text{(A.11)}$$

which introduced in (A.9) gives

$$\Delta_{u-f} = \phi^{2(h-1)} \sum_{j=1}^{m} p_j^2 \left(\frac{\phi\theta_j}{(1-\phi\theta_j)} - \frac{1}{Vm\bar{\mu}} + \frac{1-\phi^2}{m\bar{\mu}} \right). \quad \text{(A.12)}$$

Also inserting the positive solution of (A.10) in (A.12) and using (28) and (26), we obtain (31). Note that in (31) the first term is positive because $(\mu_j + 1 + \phi^2)^2 - 4\phi^2 = (\mu_j + 1 - \phi^2)^2 + 4\mu_j\phi^2$. To prove that Δ_{u-f} in (31) is always positive, first notice that in the stationary case $(1-\phi^2)/m\bar{\mu} > 0$. Then it is sufficient to prove that

$$\left(\sqrt{(\mu_j + 1 - \phi^2)^2 + 4\mu_j\phi^2} - (\mu_j + 1 - \phi^2) \right) \times \left(m\bar{\mu} + \phi^2 - 1 + \sqrt{(m\bar{\mu} + \phi^2 - 1)^2 + 4m\bar{\mu}} \right) - 4\mu_j > 0.$$

Calling $A = \sqrt{(m\bar{\mu} + \phi^2 - 1)^2 + 4m\bar{\mu}}$ and $B = \sqrt{(\mu_j + 1 - \phi^2)^2 + 4\mu_j}$, and noting that $B - (\mu_j + 1 - \phi^2) > 0$, $m\bar{\mu} > \mu_j$ and $A > B$ if $m > 1$, this expression can be written as $(B - (\mu_j + 1 - \phi^2))((m\bar{\mu} + \phi^2 - 1) + A) - 4\mu_j > (B - (\mu_j + 1 - \phi^2))(B + (\mu_j + 1 - \phi^2)) - 4\mu_j = (\mu_j - (1 - \phi^2))^2 \geqslant 0$, and the result is proved.

To obtain (30), just insert $\phi = 1$ in (31), and to prove now that this difference is strictly positive for $m > 1$, we just have to prove that in this case

$$\left(\sqrt{\mu_j^2 + 4\mu_j} - \mu_j \right) \left(m\bar{\mu} + \sqrt{m^2\bar{\mu}^2 + 4m\bar{\mu}} \right) - 4\mu_j > 0.$$

Calling $A=\sqrt{m^2\bar{\mu}^2+4m\bar{\mu}}$ and $B=\sqrt{\mu_j^2+4\mu_j}$, and noting that $B-\mu_j>0$, $m\bar{\mu}>\mu_j$ and $A>B$, this expression can be written as $(B-\mu_j)(m\bar{\mu}+A)-4\mu_j>(B-\mu_j)(B+\mu_j)-4\mu_j=0$, and the result is proved. It is easy to check that if $m=1$ then $\Delta_{u-f}=0$. □

Proof of Theorem 3. We first prove three auxiliary lemmas and corollaries.

Lemma 5. *For the one factor model given by* (17) *and* (18), *with* $p_j=p\ \forall j=1,\dots,m$. *Let* $\bar{\theta v}_t=(1/m)\sum_{i=1}^m\theta_i v_{i,t}$, $\bar{v}_t=(1/m)\sum_{i=1}^m v_{i,t}$ *and* $\bar{\varepsilon}_t=(1/m)\sum_{i=1}^m\varepsilon_{i,t}$ *for all* $t=1,\dots,T$. *Then*

(i) $\phi(\varepsilon_{j,t}-\bar{\varepsilon}_t)-(\theta_j v_{j,t}-\bar{\theta v}_t)=(\varepsilon_{j,t+1}-\bar{\varepsilon}_{t+1})-(v_{j,t+1}-\bar{v}_{t+1})$,
(ii) $\mathrm{E}(\varepsilon_{j,t+1}-\bar{\varepsilon}_{t+1})^2=((m-1)/m)^2\sigma_j^2+(1/m^2)\sum_{i=1,i\neq j}^m\sigma_i^2$,
(iii) $\mathrm{E}(v_{j,t+1}-\bar{v}_{t+1})^2=(1/m^2)\sum_{i\neq j}^m[(m-1)\sigma_j^2+\sigma_i^2][1+((\phi-\theta_i)(\phi-\theta_j))/(1-\theta_i\theta_j)]+(p^2/m^2)\sum_{i\neq j}^m\sum_{k\neq i,j}^m(1/(1-\theta_i\theta_k)-1/(1-\theta_i\theta_j))$, *and*
(iv) $\mathrm{E}[(\varepsilon_{j,t+1}-\bar{\varepsilon}_{t+1})(v_{j,t+1}-\bar{v}_{t+1})]=((m-1)/m)^2\sigma_j^2+(1/m^2)\sum_{i=1,i\neq j}^m\sigma_i^2$.

Proof. (i) From (A.1),

$$\phi\varepsilon_{j,t}-\theta v_{j,t}=pa_{t+1}+\varepsilon_{j,t+1}-v_{j,t+1}, \tag{A.13}$$

applying this equation for $j=1,\dots,m$, summing up and dividing by m, we get that

$$\phi\bar{\varepsilon}_t-\bar{\theta v}_t=pa_{t+1}+\bar{\varepsilon}_{t+1}-\bar{v}_{t+1}, \tag{A.14}$$

where $\bar{\varepsilon}_t=(1/m)\sum_{i=1}^m\varepsilon_{i,t}$ and $\bar{v}_t=(1/m)\sum_{i=1}^m v_{i,t}$ and $\bar{\theta v}_t=(1/m)\sum_{i=1}^m\theta_i v_{i,t}$. Subtracting (A.14) from (A.13) we get the desired result.

(ii) Since $\mathbf{\Sigma}_\varepsilon$ is diagonal $\mathrm{E}(\varepsilon_{j,t+1}-\bar{\varepsilon}_{t+1})^2=\mathrm{E}((m-1)/m\ \varepsilon_{j,t+1}-(1/m)\sum_{i=1,i\neq j}^m\varepsilon_{i,t+1})^2=((m-1)/m)^2\sigma_j^2+(1/m^2)\sum_{i=1,i\neq j}^m\sigma_i^2$.

(iii) $\mathrm{E}(v_{j,t+1}-\bar{v}_{t+1})^2=\mathrm{E}(((m-1)/m)v_{j,t+1}-(1/m)\sum_{i=1,i\neq j}^m v_{i,t+1})^2$; expanding the square

$$\begin{aligned}\mathrm{E}(v_{j,t+1}-\bar{v}_{t+1})^2&=\left(\frac{m-1}{m}\right)^2\mathrm{E}(v_{j,t+1}^2)+\frac{1}{m^2}\mathrm{E}\left(\sum_{i=1,i\neq j}^{2}v_{i,t+1}\right)^2\\&\quad-2\frac{m-1}{m^2}\mathrm{E}\left(v_{j,t+1}\sum_{i=1,i\neq j}^{m}v_{i,t+1}\right)\\&=\left(\frac{m-1}{m}\right)^2\tilde{\sigma}_j^2+\frac{1}{m^2}\sum_{i=1,i\neq j}^{m}\tilde{\sigma}_i^2+\frac{1}{m^2}\sum_{i=1,i\neq j}^{m}\sum_{k=1,k\neq i,j}^{m}\tilde{\sigma}_{ik}\\&\quad-2\frac{m-1}{m^2}\sum_{i=1,i\neq j}^{m}\tilde{\sigma}_{ij}.\end{aligned}$$

By the auxiliary lemma, Parts 1 and 7, and rearranging terms

$$\begin{aligned}
\mathrm{E}(v_{j,t+1}-\bar{v}_{t+1})^2 &= \left(\frac{m-1}{m}\right)^2\left(\sigma_j^2+\frac{p^2}{1-\phi\theta_j}\right)+\frac{1}{m^2}\sum_{i=1,i\neq j}^{m}\left(\sigma_i^2+\frac{p^2}{1-\phi\theta_i}\right)\\
&\quad+\frac{p^2}{m^2}\sum_{i=1,i\neq j}^{m}\sum_{k=1,k\neq i,j}^{m}\frac{1}{1-\theta_i\theta_k}-2\frac{m-1}{m^2}\sum_{i=1,i\neq j}^{m}\frac{1}{1-\theta_i\theta_j}\\
&=\frac{1}{m^2}\sum_{i=1,i\neq j}^{m}[(m-1)\sigma_j^2+\sigma_i^2]\\
&\quad+\frac{p^2}{m^2}\sum_{i=1,i\neq j}^{m}\sum_{k=1,k\neq i,j}^{m}\left(\frac{1}{1-\theta_i\theta_k}-\frac{1}{1-\theta_i\theta_j}\right)\\
&\quad+\frac{p^2}{m^2}\sum_{i=1,i\neq j}^{m}\left[(m-1)\left(\frac{1}{1-\phi\theta_j}-\frac{1}{1-\theta_i\theta_j}\right)\right.\\
&\quad\left.+\left(\frac{1}{1-\phi\theta_i}-\frac{1}{1-\theta_i\theta_j}\right)\right].
\end{aligned}$$

Finally, after some straight forward algebra and taking into account (28)

$$\begin{aligned}
\mathrm{E}(v_{j,t+1}-\bar{v}_{t+1})^2 &= \frac{1}{m^2}\sum_{i=1,i\neq j}^{m}[(m-1)\sigma_j^2+\sigma_i^2]\left[1+\frac{(\phi-\theta_i)(\phi-\theta_j)}{1-\theta_i\theta_j}\right]\\
&\quad+\frac{p^2}{m^2}\sum_{i=1,i\neq j}^{m}\sum_{k=1,k\neq i,j}^{m}\left(\frac{1}{1-\theta_i\theta_k}-\frac{1}{1-\theta_i\theta_j}\right).
\end{aligned}$$

(iv) Taking into account, the auxiliary lemma, Part 6, $\mathrm{E}[(\varepsilon_{j,t+1}-\bar{\varepsilon}_{t+1})(v_{j,t+1}-\bar{v}_{t+1})]=\mathrm{E}(\varepsilon_{j,t+1}v_{j,t+1}-\bar{\varepsilon}_{t+1}v_{j,t+1}-\varepsilon_{j,t+1}\bar{v}_{t+1}+\bar{\varepsilon}_{t+1}\bar{v}_{t+1})=\sigma_j^2-\sigma_j^2/m-\sigma_j^2/m+(1/m^2)\sum_{i=1}^{m}\sigma_i^2=((m-1)/m)^2\sigma_j^2+(1/m^2)\sum_{i=1,i\neq j}^{m}\sigma_i^2$. □

Lemma 6. *For the one factor model given by* (17) *and* (18), *with* $p_j=p\ \forall j=1,\dots,m$, *then* $\mathrm{E}(\hat{y}_{j,t+h|t}-\hat{y}_t(h))^2=\phi^{2(h-1)}/m^2(\sum_{i\neq j}^{m}[(m-1)\sigma_j^2+\sigma_i^2](\phi-\theta_i)(\phi-\theta_j)/(1-\theta_i\theta_j)+p^2\sum_{i\neq j}^{m}\sum_{k\neq i,j}^{m}(1/(1-\theta_i\theta_k)-1/(1-\theta_i\theta_j)))$. *If additionally* $\sigma_j^2=\sigma^2\ \forall j=1,\dots,m$, *the previous expression reduces to* $\mathrm{E}(\hat{y}_{j,t+h|t}-\hat{y}_t(h))^2=\phi^{2(h-1)}((m-1)/m)\sigma_j^2((\phi-\theta)^2/(1-\theta^2))$.

Proof. First, we will calculate the difference between the univariate forecast and the mean of the univariate forecasts. From (33), and since $p_j=p\ \forall j=1,\dots,m$

$$\begin{aligned}
\hat{y}_{j,t+h|t}-\hat{y}_t(h) &= \phi^h(\varepsilon_{j,t}-\bar{\varepsilon}_t)-\phi^{h-1}(\theta_j v_{j,t}-\overline{\theta v}_t)\\
&=\phi^{h-1}(\varepsilon_{j,t+1}-\bar{\varepsilon}_{t+1}-(v_{j,t+1}-\bar{v}_{t+1})),
\end{aligned} \tag{A.15}$$

where the last equality comes from Part (i) of Lemma 5. Take the expectation of the square of this expression

$$E(\hat{y}_{j,t+h|t} - \hat{y}_t(h))^2 = \phi^{2(h-1)}[E(\varepsilon_{j,t+1} - \bar{\varepsilon}_{t+1})^2 + E(v_{j,t+1} - \bar{v}_{t+1})^2 - 2E(\varepsilon_{j,t+1} - \bar{\varepsilon}_{t+1})(v_{j,t+1} - \bar{v}_{t+1})].$$

Parts (ii)–(iv) of Lemma 5 allows to write

$$E(\hat{y}_{j,t+h|t} - \hat{y}_t(h))^2 = \frac{\phi^{2(h-1)}}{m^2}\left(\sum_{i\neq j}^{m}[(m-1)\sigma_j^2 + \sigma_i^2]\frac{(\phi-\theta_i)(\phi-\theta_j)}{1-\theta_i\theta_j} + p^2\sum_{i\neq j}^{m}\sum_{k\neq i,j}^{m}\left(\frac{1}{1-\theta_i\theta_k} - \frac{1}{1-\theta_i\theta_j}\right)\right). \tag{A.16}$$

If additionally $\sigma_j^2 = \sigma^2\ \forall j = 1,\ldots,m$, then $\theta_j = \theta\ \forall j = 1,\ldots,m$, and (A.16) reduces to $E(\hat{y}_{j,t+h|t} - \hat{y}_t(h))^2 = \phi^{2(h-1)}((m-1)/m)\sigma_j^2((\phi-\theta)^2/1-\theta^2)$. □

Lemma 7. *For the one factor model given by* (17) *and* (18), *with* $p_j = p\ \forall j = 1,\ldots,m$, *then*

(a) $E[(v_{j,t+\tau}(\varepsilon_{j,t} - \bar{\varepsilon}_t)] = -((m-1)/m)\theta_j^{\tau-1}(\phi-\theta_j)\sigma_j^2,\ \forall\tau > 0$,
(b) $E[(v_{j,t+\tau}(v_{j,t} - \bar{v}_t)] = -(1/m)\theta_j^{\tau-1}(\phi-\theta_j)(1-\phi\theta_j)\sigma_j^2\sum_{i=1,i\neq j}^{m}(1/(1-\theta_i\theta_j))$ *and*
(c) $E[v_{j,t+\tau}(\hat{y}_{j,t+h|t} - \hat{y}_t(h))] = -\phi^{h-1}(1/m)\theta_j^{\tau-1}(\phi-\theta_j)\sigma_j^2\sum_{i=1,i\neq j}^{m}((\phi-\theta_i)/(1-\theta_i\theta_j))\ \forall\tau > 0$, *where* $\bar{e}_t = 1/m\sum_{i=1}^{m}\bar{e}_{i,t}$ *and* $\bar{v}_t = (1/m)\sum_{i=1}^{m}v_{i,t}$.

Proof. To prove (a), applying Parts 5 and 6 of the auxiliary lemma

$$E[(v_{j,t+\tau}(\varepsilon_{j,t} - \bar{\varepsilon}_t)] = E\left(\frac{m-1}{m}v_{j,t+\tau}\varepsilon_{j,t} - \frac{1}{m}v_{j,t+\tau}\sum_{i=1,i\neq j}^{m}\varepsilon_{i,t}\right) = E\left(\frac{m-1}{m}v_{j,t+\tau}\varepsilon_{j,t}\right) = -\frac{m-1}{m}\theta_j^{\tau-1}(\phi-\theta_j)\sigma_j^2,$$

we get the desired result.

To proof (b), after some straightforward algebra and applying Part 8 of the auxiliary lemma

$$E[(v_{j,t+\tau}(v_{j,t} - \bar{v}_t)] = E\left(\frac{m-1}{m}v_{j,t+\tau}v_{j,t} - \frac{1}{m}v_{j,t+\tau}\sum_{i=1,i\neq j}^{m}v_{i,t}\right) = -\frac{1}{m}\sum_{i=1,i\neq j}^{m}E(v_{j,t+\tau}v_{i,t}) = -\frac{1}{m}\theta_j^{\tau}\sum_{i=1,i\neq j}^{m}\frac{p^2}{1-\theta_i\theta_j} = -\frac{1}{m}\theta_j^{\tau-1}(\phi-\theta_j)(1-\phi\theta_j)\sigma_j^2\sum_{i=1,i\neq j}^{m}\frac{1}{1-\theta_i\theta_j},$$

where the last equality is obtained using (28).

From (A.15), taking into account (a) and (b), and after some straightforward algebra

$$
\begin{aligned}
\mathrm{E}[v_{j,t+\tau}(\hat{y}_{j,t+h|t}-\hat{y}_t(h))] &= \phi^{h-1}(\mathrm{E}(v_{j,t+\tau}(\varepsilon_{j,t+1}-\bar{\varepsilon}_{t+1}))-\mathrm{E}(v_{j,t+\tau}(v_{j,t+1}-\bar{v}_{t+1}))) \\
&= -\phi^{h-1}\frac{1}{m}\theta_j^{\tau-2}(\phi-\theta_j)\sigma_j^2 \\
&\quad\times\left(m-1-(1-\phi\theta_j)\sum_{i=1,i\neq j}^{m}\frac{1}{1-\theta_i\theta_j}\right) \\
&= -\phi^{h-1}\frac{1}{m}\theta_j^{\tau-1}(\phi-\theta_j)\sigma_j^2\sum_{i=1,i\neq j}^{m}\frac{\phi-\theta_i}{1-\theta_i\theta_j}. \qquad \square
\end{aligned}
$$

Now the proof of the theorem is structured in two parts. First, from Lemma 7 we will show that $\mathrm{E}[(y_{j,t+h}-\hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t}-\hat{y}_t(h))]<0,\ \forall h>0$ and will give its expression. And second, we will show that the pooling forecasts methods produces a prediction MSE smaller than the one obtained with the ARIMA univariate forecasts.

(i) Proof of $\mathrm{E}[(y_{j,t+h}-\hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t}-\hat{y}_t(h))]<0,\ \forall h>0$. From (A.4) and Part (c) of Lemma 7 and after some straightforward algebra,

$$
\begin{aligned}
&\mathrm{E}[(y_{j,t+h}-\hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t}-\hat{y}_t(h))] \\
&=\mathrm{E}\left[\left(v_{j,t+h}+\sum_{k=1}^{h-1}\phi^{k-1}(\phi-\theta_j)v_{j,t+h-k}\right)(\hat{y}_{j,t+h|t}-\hat{y}_t(h))\right] \\
&=-\,\phi^{h-1}\frac{1}{m}(\phi-\theta_j)\sigma_j^2\sum_{i=1,i\neq j}^{m}\frac{\phi-\theta_i}{1-\theta_i\theta_j}\left(\theta_j^{h-1}+\sum_{k=1}^{h-1}\theta_j^{h-k-1}\phi^{k-1}(\phi-\theta_j)\right) \\
&=-\,\phi^{2(h-1)}\frac{1}{m}(\phi-\theta_j)\sigma_j^2\sum_{i=1,i\neq j}^{m}\frac{\phi-\theta_i}{1-\theta_i\theta_j}, \qquad \text{(A.17)}
\end{aligned}
$$

which is negative as we have expected, since by (27) $|\theta_j|\leqslant|\phi|\ \forall j=1,\dots,m$.

Finally the expression of the MSE of prediction when we use pooled forecasts is obtained substituting (A.16) and (A.17) in (32),

$$
\begin{aligned}
MSE(y_{j,t+h}-\tilde{y}_{j,t+h}) &= MSE(y_{j,t+h}-\hat{y}_{j,t+h|t})+\eta^2E(\hat{y}_{j,t+h|t}-\hat{y}_t(h))^2 \\
&\quad+2\eta E((y_{j,t+h}-\hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t}-\hat{y}_t)) \\
&= MSE(y_{j,t+h}-\hat{y}_{j,t+h|t}) \\
&\quad-2\eta\,\frac{\phi^{2(h-1)}}{m}(\phi-\theta_j)\sigma_j^2\sum_{i=1,i\neq j}^{m}\frac{\phi-\theta_i}{1-\theta_i\theta_j}
\end{aligned}
$$

$$+\frac{\eta^2\phi^{2(h-1)}}{m^2}\sum_{i\neq j}^{m}\left([(m-1)\sigma_j^2+\sigma_i^2]\,\frac{(\phi-\theta_i)(\phi-\theta_j)}{1-\theta_i\theta_j}\right.$$

$$\left.+\,p^2\sum_{k\neq i,j}^{m}\left(\frac{1}{1-\theta_i\theta_k}-\frac{1}{1-\theta_i\theta_j}\right)\right).$$

The previous expression can be bigger than the MSE for the univariate forecasts depending on the values of the $\sigma_j^2,\ j=1,\ldots,m$. If we additionally assume that $\sigma_j^2=\sigma^2,\ j=1,\ldots,m$ the previous expression reduces to

$$MSE(y_{j,t+h}-\tilde{y}_{j,t+h})=MSE(y_{j,t+h}-\hat{y}_{j,t+h|t})$$

$$-\eta(2-\eta)\frac{m-1}{m}\,\phi^{2(h-1)}\sigma^2\frac{(\phi-\theta)^2}{1-\theta^2}.$$

Since η is between 0 and 1 and the last term of the previous equality is always subtracting, the MSE of prediction obtained with pooling forecasts is always smaller than the one obtained only from the ARIMA univariate forecasts, since we always add a negative amount to this last one. Summing up for all $j=1,\ldots,m$ we finally have (34). □

Proof of Theorem 4. First, notice that $\Delta_{p-f}=P-F=P-U+U-F=\Delta_{u-f}-\Delta_{u-p}$, where Δ_{u-f} and Δ_{u-p} have been given in Theorems 2 and 3, respectively. Second, notice from Theorem 3 that the expression given for Δ_{u-p} reaches its maximum for $\eta=1$. Therefore, it suffices to prove that $\Delta_{p-f}\geqslant 0$, for $\eta=1$, since for any other case the difference will be greater. Consider now the nonstationary case, we can write for $\phi=1$,

$$\Delta_{p-f}=\Delta_{u-f}-(m-1)\sigma^2\frac{1-\theta}{1+\theta} \tag{A.18}$$

and from (26) and since $\phi^2=1$

$$1-\theta=\frac{\sqrt{\mu^2+4\mu}-\mu}{2}.$$

Substituting Δ_{u-f} in (A.18) and after some straightforward algebra, we get (35). It is easy to check that for $m=1\Rightarrow\Delta_{p-f}=0$. In (35) and for a given μ, the first term is always positive (recall that θ is positive since it has the same sign as ϕ) and increases with m, whereas the second term decreases with m. In fact as $m\to\infty$ the first term diverges to ∞ while the second term converges to $1/\mu$.

In the stationary case, $|\phi|<1$, and from (34) and (28), for $\eta=1$, Δ_{p-f} can be written as

$$\Delta_{p-f}=\Delta_{u-f}-(m-1)\phi^{2^{(h-1)}}\frac{p^2(\phi-\theta)\theta}{(1-\theta^2)(1-\phi\theta)}$$

and taking into account (A.12) from the appendix and since $\mu_j=\mu$ and $p_j=p\ \forall j=1,\ldots,m$

$$\Delta_{p-f}=p^2m\left(\frac{\theta}{\phi(1-\phi\theta)}\,\beta-\frac{1}{Vm\mu\phi^2}+\frac{1-\phi^2}{\phi^2m\mu}\right)\phi^{2h},$$

where $\beta = 1 - ((m-1)/m)((\phi-\theta)/\phi(1-\theta^2))$. When $h \to \infty$, $\phi^{2h} \to 0$ and $\Delta_{p-f} \to 0$. To see the behavior of Δ_{p-f} with m, notice that the inside parenthesis, which is multiplied by m, of the previous equation is given by

$$\frac{\phi^2 - 1 - \mu + \sqrt{(\mu+1+\phi^2)^2 - 4\phi^2}}{2\mu}\beta + \frac{1-\phi^2}{m\mu} - \frac{2}{m\mu + \phi^2 - 1 + \sqrt{(m\mu - 1 + \phi^2)^2 + 4m\mu}}. \tag{A.19}$$

When m increases, β converges to a constant, so the first term in (A.19) also converges to a constant. The second term decreases when $m\mu$ increases, while the third one decreases with $2m\mu$. □

References

Akaike, H., 1974. Markovian representations of stochastic processes and its application to the analysis of autoregressive moving average processes. Annals of the Institute of Statistical Mathematics 26, 363–387.

Christoffersen, P.F., Diebold, F.X., 1998. Cointegration and long-horizon forecasting. Journal of Business and Economic Statistics 16, 450–458.

Clements, M.P., Hendry, D.F., 1995. Forecasting in cointegrated systems. Journal of Applied Econometrics 10, 127–146.

Diebold, F.X., Rudebusch, G.D., 1996. Measuring business cycles: a modern perspective. The Review of Economics and Statistics 78, 67–77.

Engle, R.F., Kozicki, S., 1993. Testing for common features. Journal of Business and Economic Statistics 11, 369–380.

Engle, R.F., Watson, M.W., 1981. A one-factor multivariate time series model of metropolitan wage rates. Journal of the American Statistical Association 76, 774–781.

Engle, R.F., Yoo, S., 1987. Forecasting and testing in cointegrated systems. Journal of Econometrics 35, 143–159.

Escribano, A., Peña, D., 1994. Cointegration and common factors. Journal of Time Series Analysis 15, 577–786.

Forni, M., Reichlin, L., 1998. Let's get real: a dynamic factor analytical approach to disaggregated business cycles. Review of Economic Studies 65, 453–474.

Garcìa-Ferrer, A., Novales, A., 1998. Forecasting with money demand functions: the U.K. case. Journal of Forecasting 17, 125–145.

Garcìa-Ferrer, A., Poncela, P., 2002. Forecasting European GNP data through common factor models and other procedures. Journal of Forecasting 21, 225–244.

Garcìa-Ferrer, A., Highfield, R.A., Palm, F., Zellner, A., 1987. Macroeconomic forecasting using pooled international data. Journal of Business and Economic Statistics 5, 53–67.

Geweke, J., 1977. The dynamic factor analysis of economic time series models. In: Aigner, D.I., Goldberger, A.S. (Eds.), Latent Variables In: Socio-economic Models. North Holland, New York.

Geweke, J.F., Singleton, K.J., 1981. Maximum likelihood "confirmatory" factor analysis of economic time series. International Economic Review 22, 37–54.

Gonzalo, J., Granger, C.W.J., 1995. Estimation of common long-memory components in cointegrated systems. Journal of Business and Economic Statistics 13, 27–36.

Harvey, A.C., 1989. Forecasting Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Li, D., Dorfman, J., 1996. Predicting turning points through the integration of multiple models. Journal of Business and Economic Statistics 14, 421–429.

Lin, J.L., Tsay, R., 1996. Co-integration constraint and forecasting. An empirical examination. Journal of Applied Econometrics 114, 519–538.

Min, C., Zellner, A., 1993. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Journal of Econometrics 56, 89–118.

Mittnik, S., 1990. Macroeconomic forecasting using pooled international data. Journal of Business and Economic Statistics 8, 205–208.

Peña, D., Box, G., 1987. Identifying a simplifying structure in time series. Journal of the American Statistical Association 82, 836–843.

Peña, D., Poncela, P., 2002. Nonstationary dynamic factor models. Mimeo, Universidad Carlos III.

Rao, C.R., 1973. Linear Statistical Inference and its Applications. Wiley, New York.

Reinsel, G.C., Ahn, S.K., 1992. Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting. Journal of Time Series Analysis 13, 353–375.

Stock, J.H., Watson, M.W., 1988. Testing for common trends. Journal of the American Statistical Association 83, 1097–1107.

Stock, J.H., Watson, M.W., 1991. A probability model of the coincident economic indicators. In: Lahiri, K., Moore, G.H. (Eds.), Leading Economic Indicators: New Approaches and Forecasting Records. Cambridge University Press, Cambridge, pp. 63–89.

Tiao, G.C., Tsay, R.S., 1989. Model specification in multivariate time series. Journal of the Royal Statistical Society B 51, 157–213.

Vahid, F., Engle, R., 1997. Codependent cycles. Journal of Econometrics 80, 199–221.

Velu, R.P., Reinsel, G.C., Wichern, D.W., 1986. Reduced rank models for multiple time series. Biometrika 73, 105–118.

Zellner, A., Hong, C., 1989. Forecasting international growth rates using Bayesian shrinkage and other procedures. Journal of Econometrics 40, 183–200.

Zellner, A., Min, C., 1998. Forecasting turning points in countries' output growth rates: a response to Milton Friedman. Journal of Econometrics 88, 203–206.

Zellner, A., Hong, C., Min, C., 1991. Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregressions, time-varying techniques and other procedures. Journal of Econometrics 49, 275–304.