FORECASTING WITH NONSTATIONARY DYNAMIC FACTOR MODELS

Daniel Peña and Pilar Poncela

00-49

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-49
Statistics and Econometrics Series 21
July 2000

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

FORECASTING WITH NONSTATIONARY DYNAMIC FACTOR MODELS

Daniel Peña and Pilar Poncela*

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 f or using a factor model versus univariate ARIMA models, shrinkage univariate models or vector ARMA models are studied f or the one factor model. It is proved that some substantial gains can be obtained in some cases with respect to the univariate forecasting. Monte Carlo simulations are presented to illustrate this result. A factor model is built to forecast GNP of European countries and it is shown that the factor model provides better forecasts than both univariate and shrinkage univariate forecasts.

Keywords: cointegration; common factors; pooled forecasts; prediction vector time series.

*Peña, Departamento de Estadistica y Econometría, Universidad Carlos III de Madrid, Spain, E-mail: dpena@est-econ.uc3m.es; Poncela, Departamento de Análisis Económico: Economía Cuantitativa. Universidad Autónoma de Madrid, Spain.

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), Gourieroux et al (1991), Engle and Kozicki (1993), Gonzalo and Granger (1995), Vahid and Engle (1997), and Peña and Poncela (1999), among others. When the factors are nonstationary, the problem is very related to forecasting by using cointegration relationships (see Escribano and Peña ,1994). Some recent applications of the factor model in the Economics field include Stock and Watson (1999) and Forni and Reichlin (1998). (See these articles for further applications).

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 units roots led to worse results in the short-run forecasts, underspecifying the number of unit roots 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 led 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 García-Ferrer and Novales (1998).

This paper has the following contributions. First we show that the forecasts from a factor model incorporate a pooling term similar to the one derived from hierarchical Bayesian models. This shrinkage terms can be, in some particular cases, identical to the pooling term proposed by García-Ferrer et al (1987) that has been shown to work well in practice. Thus the factor model provides a formal justification for shrinkage methods in univariate forecasts and allows a way to derive the optimal shrinkage in each case. Second, we derive the expected gains in the one factor case of the forecasts from the factor model with respect to univariate, shrinkage univariate and multivariate ARIMA models. The advantage of the factor model increases with the dimension of the time series vector. Third, we show by Monte Carlo and a real example the reduction in mean square forecast error obtained 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. The large sample comparison of the forecast performance of the one factor model and the ARIMA univariate models is made in section 5, and in sections 6 and 7 the forecast performance of the one factor model is compared to the shrinkage pooled univariate forecasts and to the VARMA forecasts. Section 8 presents some simulation results. In section 9 a factor model is built for forecasting the Gross National Product (GNP) of European countries and its forecasting performance is compared to the ones obtained by univariate and shrinkage univariate forecasts. Finally, section 10 presents 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,

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

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

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

where $\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ and $\boldsymbol{\Theta}(B) = \mathbf{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ are polynomial matrices $r \times r$, B is the backshift operator, such that $B\mathbf{y}_t = \mathbf{y}_{t-1}$, the roots of $|\boldsymbol{\Phi}(B)| = 0$ are on or outside the unit circle, the roots of $|\boldsymbol{\Theta}(B)| = 0$ are outside the unit circle and $\mathbf{a}_t \sim N_r(\mathbf{0}, \boldsymbol{\Sigma}_a)$, 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, $\mathrm{E}(\mathbf{a}_t \boldsymbol{\epsilon}'_{t-h}) = \mathbf{0}, \forall h$.

The model as stated is not identified, because for any $r \times r$ non singular 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 $\boldsymbol{\Sigma}_{\mathbf{a}} = \mathbf{I}$ 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 $\boldsymbol{\Sigma}_{\mathbf{a}} = \mathbf{I}$; this restriction excludes the case where a common factor is just a constant, which is not analyze in this paper. We will add the standard restriction in static factor analysis, that is $\mathbf{P}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{P}$ diagonal.

The model can be generalized to the case where the components in $\boldsymbol{\epsilon}_t$ have dynamic univariate stationary structure, see Peña and Poncela (1999), 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 equation (1).

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

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

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 $\mathrm{E}(\mathbf{u}_t) = \mathbf{0}$, $\mathrm{E}(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}_u$ 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, $\mathrm{E}(\epsilon_t\mathbf{u}_\tau') = \mathbf{0}$ for all t and τ. Given information until time $t-1$, the well-known Kalman filter equations give the forecast of the state vector

$$\mathbf{z}_{t|t-1} = \mathbf{G}\mathbf{z}_{t-1|t-1} \tag{5}$$

with associated covariance matrix,

$$\mathbf{V}_{t|t-1} = \mathbf{G}\mathbf{V}_{t-1|t-1}\mathbf{G}' + \mathbf{\Sigma}_u \tag{6}$$

and the forecast for the vector of time series is computed by

$$\hat{\mathbf{y}}_{t|t-1} = \tilde{\mathbf{P}}\mathbf{z}_{t|t-1} \tag{7}$$

with covariance matrix

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

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

$$\mathbf{z}_{t|t} = \mathbf{z}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \tilde{\mathbf{P}}\mathbf{z}_{t|t-1}) \tag{9}$$

and its variance-covariance matrix by

$$\mathbf{V}_{t|t} = \mathbf{V}_{t|t-1} - \mathbf{V}_{t|t-1}\tilde{\mathbf{P}}'\mathbf{\Sigma}_{t|t-1}^{-1}\tilde{\mathbf{P}}\mathbf{V}_{t|t-1} \tag{10}$$

where $\mathbf{K}_t$ is the filter gain, given by

$$\mathbf{K}_t = \mathbf{V}_{t|t-1}\tilde{\mathbf{P}}'\mathbf{\Sigma}_{t|t-1}^{-1}. \tag{11}$$

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 applying repeatedly (5)

$$\mathbf{z}_{t+h|t} = \mathbf{G}^h \mathbf{z}_{t|t} \tag{12}$$

with mean square error (MSE) matrix $\mathbf{V}_{t+h|t} = E(\mathbf{z}_{t+h} - \mathbf{z}_{t+h|t})(\mathbf{z}_{t+h} - \mathbf{z}_{t+h|t})'$ given by

$$\mathbf{V}_{t+h|t} = \mathbf{G}^h \mathbf{V}_{t|t} (\mathbf{G}')^h + \mathbf{G}^{h-1} \boldsymbol{\Sigma}_u (\mathbf{G}')^{h-1} + ... + \mathbf{G}\boldsymbol{\Sigma}_u \mathbf{G}' + \boldsymbol{\Sigma}_u. \tag{13}$$

>From (12) and (7) in $t+h|t$, we obtain that the h steps ahead forecast for the observed series with origin in t is

$$\widehat{\mathbf{y}}_{t+h|t} = \tilde{\mathbf{P}}\mathbf{G}^{\mathbf{h}} \mathbf{z}_{t|t}, \tag{14}$$

with MSE matrix

$$\boldsymbol{\Sigma}_{t+h|t} = E(\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t})(\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t})' = \tilde{\mathbf{P}}\mathbf{V}_{t+h|t}\tilde{\mathbf{P}}' + \boldsymbol{\Sigma}_\epsilon. \tag{15}$$

Using (9) in (14), we can write the forecast of the observed series as

$$\widehat{\mathbf{y}}_{t+h|t} = \mathbf{A}_1 \mathbf{z}_{t|t-1} + \mathbf{A}_2 \mathbf{y}_t, \tag{16}$$

where

$$\mathbf{A}_1 = \tilde{\mathbf{P}}\mathbf{G}^h (\mathbf{I} - \mathbf{K}_t \tilde{\mathbf{P}}) \tag{17}$$

and

$$\mathbf{A}_2 = \tilde{\mathbf{P}}\mathbf{G}^h \mathbf{K}_t. \tag{18}$$

An equivalent expression of (16) can be obtained by using the well-known expression for the inverse of the sum of two matrices (see, Rao, 1973, p. 33) in (8) for the inverse of $\boldsymbol{\Sigma}_{t|t-1}$and plugging it into (18) through $\mathbf{K}_t$ in (11). This leads to

$$\widehat{\mathbf{y}}_{t+h|t} = \mathbf{A}_1 \mathbf{z}_{t|t-1} + \mathbf{W}_t \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{y}_t \tag{19}$$

where

$$\mathbf{W}_t = \tilde{\mathbf{P}}\mathbf{N}_t \tilde{\mathbf{P}}', \tag{20}$$

is a $m \times m$ matrix and (see the appendix)

$$\mathbf{N}_t = \mathbf{G}^h \mathbf{V}_{t|t-1}(\mathbf{I}_s - \tilde{\mathbf{P}}' \boldsymbol{\Sigma}_\epsilon^{-1} \tilde{\mathbf{P}} \mathbf{V}_{t|t}) \tag{21}$$

a $s \times s$ matrix. Note that although if $\tilde{\mathbf{P}}' \boldsymbol{\Sigma}_\epsilon^{-1} \tilde{\mathbf{P}} = \mathbf{P}' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{P}$ is a diagonal matrix, by the identification restrictions, nevertheless, $\mathbf{V}_{t|t-1}$ and $\mathbf{V}_{t|t}$ do not need to be diagonal by (6), (8) and (10), even though $\boldsymbol{\Sigma}_u$ and $\mathbf{G}$ were diagonal.

For the general case, and since $\boldsymbol{\Sigma}_\epsilon$ is a diagonal matrix, the j-th component of the forecast vector can be written as

$$\widehat{y}_{j,t+h|t} = (\mathbf{A}_1 \mathbf{z}_{t|t-1})_j + \sum_{i=1}^{m} \frac{w_{ji,t}}{\sigma_i^2} y_{i,t}, \tag{22}$$

where $(\mathbf{X})_j$ is the j-th 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 series observed at time t with weights proportional to the elements of the $\mathbf{W}_t$ matrix and inversely proportional to their noise variance.

4 The prediction structure of the one factor model

We analyze now how the observations at time t are incorporated in the forecasts with 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. 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 and 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

$$\underset{m \times 1}{\mathbf{y}_t} = \underset{m \times 1}{\mathbf{P}} \; \underset{1 \times 1}{f_t} + \underset{m \times 1}{\epsilon_t}, \tag{23}$$

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

$$\underset{1 \times 1}{f_t} = \underset{1 \times 1}{\phi} \; \underset{1 \times 1}{f_{t-1}} + \underset{1 \times 1}{a_t}, \tag{24}$$

with a_t white noise with $\text{var}(a_t) = \sigma_a^2 = 1$, (by the identification restriction) $\text{cov}(\epsilon_{i,t} a_\tau) = 0$ for all i, t and τ and $|\phi| \leq 1$. 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 the structure of (19) in this case. Now $\mathbf{A}_1$ by (17) and (11) is given by

$$\mathbf{A}_1 = \mathbf{P}\phi^h(\mathbf{I} - V_{t|t-1}\mathbf{P}'\mathbf{\Sigma}_{t|t-1}^{-1}\mathbf{P}),$$

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 to matrices (Rao, 1973) to $\mathbf{\Sigma}_{t|t-1}^{-1} = (V_{t|t-1}\mathbf{P}\mathbf{P}' + \mathbf{\Sigma}_\epsilon)^{-1}$, $\mathbf{A}_1$ is given by

$$\mathbf{A}_1 = \mathbf{P}\frac{\phi^h}{c_t}\frac{1}{V_{t|t-1}} \tag{25}$$

with $c_t = \frac{1}{V_{t|t-1}} + \sum_{i=1}^m \frac{p_i^2}{\sigma_i^2}$. The second term in equation (19) is $\mathbf{W}_t\mathbf{\Sigma}_\epsilon^{-1}\mathbf{y}_t = N_t\mathbf{P}\mathbf{P}'\mathbf{\Sigma}_\epsilon^{-1}\mathbf{y}_t$ where N_t is now an scalar given by $N_t = \frac{\phi^h}{c_t}$. The h steps ahead forecast of the observed series $\mathbf{y}_{t+h}$ with information up to time t is, substituting in (19) $\mathbf{A}_1$ and $\mathbf{W}_t\mathbf{\Sigma}_\epsilon^{-1}\mathbf{y}_t$ by their expressions is

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

and the j-th component of the forecasted series is

$$\widehat{y}_{j,t+h|t} = p_j\frac{\phi^h}{c_t}\left(\frac{1}{V_{t|t-1}}f_{t|t-1} + \sum_{i=1}^m \frac{p_i^2}{\sigma_i^2}(\frac{y_{i,t}}{p_i})\right).$$

To understand the meaning of the previous equations, first notice that $\frac{1}{c_t}\left(\frac{1}{V_{t|t-1}}f_{t|t-1} + \sum_{i=1}^m \frac{p_i^2}{\sigma_i^2}(\frac{y_{i,t}}{p_i})\right)$ 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 on $\mathbf{y}_t$. To see this note that by (1) at each time t we have m new possible independent estimates of f_t given by $E(f_t|y_{j,t}) = \frac{y_{j,t}}{p_j}$ with variances equal to $\frac{\sigma_j^2}{p_j^2}$ for each $j = 1, ..., m$. 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 quantities $\frac{p_i^2}{\sigma_i^2}$ will appear throughout the article because as we will see they are of key importance in comparing forecasts. Let us denote by

$$\mu_i = \frac{p_i^2}{\sigma_i^2}, \quad i = 1, ..., m, \tag{26}$$

to the precision of the estimation of the factor from series y_i. These measures determine how new information is incorporated into the forecasts.

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, decrease 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 $\mathbf{P} = \mathbf{1}p = p(1, 1, \cdots, 1)'$. The forecast of the observed series $\widehat{\mathbf{y}}_{t+h|t}$ is

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

with $c_t = \frac{1}{V_{t|t-1}} + p^2\sum_{i=1}^{m}\frac{1}{\sigma_i^2}$. 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 $\phi^h = 1$.

If all the series exhibit the same noise variance, the optimal forecast is obtained shrinking the series towards the common mean. That is, if $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2 = \sigma^2$, then $c_t = \frac{1}{V_{t|t-1}} + \frac{p^2 m}{\sigma^2}$ and

$$\widehat{y}_{j,t+h|t} = \frac{p\phi^h}{c_t}\left(\frac{1}{V_{t|t-1}}f_{t|t-1} + \frac{pm}{\sigma^2}\overline{y}_t\right),$$

where $\overline{y}_t = 1/m\sum_{i=1}^{m}y_{i,t}$ is the mean at time t of the m observed variables.

In particular, suppose that a set of time series are driven by a common factor that follows a random walk process, and that the variances of all the series are approximately the same. Then the univariate forecasts derived from this model incorporates a shrinking factor towards the mean of the individual series. Therefore, we expect that in these situations incorporating this term to the standard univariate forecasts will reduce the MSE of prediction. This property was found by García-Ferrer et al (1987). One of the predictors used by these authors and that we shall denote $\widetilde{y}_{j,t+h}$, (being the forecast of the j-th component of $\mathbf{y}_t$ h steps ahead) is

$$\widetilde{y}_{j,t+h} = (1-\eta)\widehat{y}_{j,t+h|t} + \eta\widehat{y}_t(h), \qquad 0 < \eta < 1, \tag{27}$$

where $\widehat{y}_t(h) = 1/m\sum_{i=1}^{m}\widehat{y}_{i,t+h|t}$ is the mean at time t of the univariate forecasts of the components of the series collected in $\mathbf{y}_t$. If the series are random walks, both shrinkage are identical. Otherwise this last predictor has as pooling term the mean of the univariate forecasts with origin in t, instead of the sample mean at time t of the observed series. A conclusion we draw from this analysis is that if the one step ahead prediction errors are very different, when building a univariate shrinkage prediction instead of a simple mean of the forecasts it may be better to consider a weighted mean with weights inversely proportional to these variances.

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 than 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

$$\mathbf{y}_t = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ p_2 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ p_m & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} f_t \\ f_{t+1|t} \\ \vdots \\ f_{t+s-1|t} \end{bmatrix} + \epsilon_t \tag{28}$$

and the transition equation can be written as (4) with $\mathbf{z}_t' = \left(f_t, f_{t+1|t}, \cdots f_{t+s-1|t}\right)$, $\mathbf{u}_t' = a_t\left(1, \psi_1, \psi_2, ..., \psi_s\right)$ 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

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\varphi_r & -\varphi_{r-1} & -\varphi_{r-2} & \cdots & -\varphi_1 \end{bmatrix} \tag{29}$$

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

$$y_{j,t+1|t} = \left(\mathbf{A}_1 \mathbf{z}_{t|t-1}\right)_j + n_{11} p_j \sum_{i=1}^{m} \frac{p_i}{\sigma_i^2} y_{i,t}, \tag{30}$$

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 in time t when the common factor affects to all the series in the same way (then $p_i = p, \forall i = 1, ..., m$) and the noise of all the series have the same variance (then $\sigma_i^2 = \sigma^2, \forall i = 1, ..., m$). Again, if the noise variances of the observed series are different, the pooling term is proportional to the weighted mean $\sum_{i=1}^{m} \frac{y_{i,t}}{\sigma_i^2}$. 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 forecasts and forecasts from the one factor model

In this section we will show that for the one factor model, the MSE of prediction is smaller than the MSE of prediction of the univariate ARIMA models. We are interested in quantifying the efficiency improvement achieved with the factor model and finding the causes that determine it.

First, we will obtain the univariate ARIMA models implied by the factor model. Second we will compute the MSE of prediction of the factor model and then, we will compare the trace of the MSE of prediction of the factor model with the sum of MSE of prediction of the univariate ARIMA models.

5.1 Univariate ARIMA forecasts

>From (23) and (24), the univariate series generated by the one factor model verify

$$y_{j,t} = \phi y_{j,t-1} + p_j a_t + \epsilon_{j,t} - \phi \epsilon_{j,t-1}, \tag{31}$$

and they will follow an ARMA(1,1) model with the same AR parameter, as

$$y_{j,t} = \phi y_{j,t-1} + v_{j,t} - \theta_j v_{j,t-1}. \tag{32}$$

In order to obtain the MA parameter implied by this representation, let var($v_{j,t}$)=$\tilde{\sigma}_j^2$, then from (31) and (32)

$$p_j a_t + \epsilon_{j,t} - \phi \epsilon_{j,t-1} = v_{j,t} - \theta_j v_{j,t-1} \tag{33}$$

and equating moments in both sides and using definition (26)(see proof of Lemma 1 in the appendix), θ_j must satisfy

$$\theta_j^2 \phi - \theta_j(\mu_j + 1 + \phi^2) + \phi = 0, \tag{34}$$

The invertible solution to the previous equation is given by

$$\theta_j = \frac{\mu_j + 1 + \phi^2 - \sqrt{(\mu_j + 1 + \phi^2)^2 - 4\phi^2}}{2\phi}. \tag{35}$$

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. In what follows, and to simplify the exposition, we will eliminate this case by assuming, without loss of generality, that $\mu_j \neq 0$. On the other hand, if $\mu_j \to \infty, \theta_j \to 0$ and the univariate series will be AR(1). From (34), it is also straightforward to show that the difference between the AR and MA parameters is given by

$$\phi - \theta_j = \frac{\theta_j \mu_j}{1 - \theta_j \phi} = \frac{\theta_j p_j^2}{\sigma_j^2 (1 - \theta_j \phi)} \tag{36}$$

and as $(\mu_j + 1 + \phi^2)^2 - 4\phi^2 = (\mu_j + 1 - \phi^2)^2 + 4\mu_j\phi^2$, from (35) we have that

$$|\theta_j| \,\Box\, \left| \frac{\mu_j + 1 + \phi^2 - (\mu_j + 1 - \phi^2)}{2\phi} \right| = |\phi|$$

Next, we will proof some auxiliary results to characterize the new error term of the univariate ARIMA models, $v_{j,t}$, and to establish the correlation between this new set of errors and the ones from the factor model

Lemma 1 *For the one factor model given by (23) and (24), if (i)* $E(a_t v_{j,0}) = 0$ *and (ii)* $E(\epsilon_{j,t} v_{i,0}) = 0$, $\forall j, i = 1, ..., m,$ *then*

1. $Var(v_{j,t}) = \tilde{\sigma}_j^2 = \sigma_j^2 + \frac{p_j^2}{1-\phi\theta_j}$, $\forall j = 1, ..., m,$

2. $E(a_t v_{j,\tau}) = 0$, $\forall j = 1, ..., m, \forall \tau < t,$

3. $E(a_t v_{j,t}) = p_j$, $\forall j, i = 1, ..., m;\ j \neq i,$

4. $E(\epsilon_{j,t} v_{j,t}) = \sigma_j^2$, $\forall j = 1, ..., m,$

5. $E(\epsilon_{j,t} v_{j,t+h}) = -\theta_j^{h-1}(\phi - \theta_j)\sigma_j^2$, $\forall j, i = 1, ..., m,\ j \neq i, \forall j = 1, ..., m, \forall h > 0,$

6. $E(\epsilon_{j,t} v_{i,\tau}) = 0$, $\forall j, i = 1, ..., m;\ j \neq i, \forall \tau$ *integer*,

7. $Cov(v_{j,t}, v_{i,t}) = \tilde{\sigma}_{ji} = \frac{p_i p_j}{1-\theta_j\theta_i}$, $\forall j, i = 1, ..., m,$ *and*

8. $Cov(v_{j,t+h}, v_{i,t}) = \frac{\theta_j^h p_i p_j}{1-\theta_j\theta_i}$, $\forall j, i = 1, ..., m, \forall h > 0.$

Proof: The proof is given in the appendix.

Once the new innovations have been characterized, we can compute the MSE of prediction of the ARMA univariate model as a function of the parameters of the factor model.

Lemma 2 *Let U be the sum of MSE of the h step ahead prediction of the ARIMA univariate models obtained from the one factor model in (23) and (24). Then*

$$U = \sum_{j=1}^{m} \left(\sigma_j^2 + \frac{p_j^2 \theta_j}{\phi(1 - \phi\theta_j)} + \frac{p_j^2(\phi - \theta_j)}{(1 - \theta_j\phi)\phi} \sum_{i=0}^{h-1} \phi^{2i} \right). \tag{37}$$

Proof: The proof is given in the appendix.

The sum of MSE of prediction depends on the number of series m, the ratio between the noise variances , the coefficients p_j, which measure the effects on the series of the factor, the AR parameter ϕ which gives the dynamics of the common factor and, of course, the forecast horizon h.

Next, we will analyze the previous expression for the two cases of interest: the nonstationary case with a random walk as the common factor and the case of an AR(1) stationary factor.

CASE 1:Nonstationary common or $I(1)$ factor.

In this case $\phi = 1$ and the univariate models are IMA(1,1). From (36) for $\phi = 1$,

$$(1 - \theta_j)^2 = \theta_j \mu_j. \tag{38}$$

The sum of MSE of prediction for the observed series h steps ahead can be written using (38) in (37) as

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

Note that, as the series have a unit root, when the horizon of prediction goes to ∞, the sum as well as the average of MSE of prediction also goes to ∞.

CASE 2: Stationary common or $I(0)$ common factors.

If $|\phi| < 1$, the univariate models are ARMA(1,1). From (37), the sum of MSE of prediction is

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

If the horizon of prediction goes to ∞, the average prediction MSE of 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 with the one factor model

In this subsection, we will prove that the one factor model provides prediction with an average smaller MSE of prediction than the univariate ARIMA models. Let $\Delta_{u-f} = U - F$, where $F = \Sigma_{t+h|t}$. The better forecasting performance of the factor model is to be expected since we assume that it is the data generating process, however it is not clear in advance how much we can win over the ARIMA univariate models. The answer to this question is provided by the following theorem.

Theorem 3 *For the one factor model given by (23) and (24), with $\mu_j \neq 0$, let $\Delta_{u-f} = U - F$. For the nonstationary case*

$$\Delta_{u-f} = \sum_{j=1}^{m} p_j^2 \left(\frac{\sqrt{\mu_j^2 + 4\mu_j} - \mu_j}{2\mu_j} - \frac{2}{m\overline{\mu} + \sqrt{m^2\overline{\mu}^2 + 4m\overline{\mu}}} \right) \geq 0, \qquad (39)$$

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

$$\Delta_{u-f} = \sum_{j=1}^{m} p_j^2 \left(\frac{\theta_j}{\phi(1-\phi\theta_j)} - V_{t|t} \right) \phi^{2h}. \qquad (40)$$

and if $h \to \infty$ $\Delta_{u-f} \to 0$. For finite h, $\Delta_{u-f} > 0$ if

$$\frac{\phi^2 - 1 - \mu_j + \sqrt{(\mu_j + 1 + \phi^2)^2 - 4\phi^2}}{2\mu_j} + \frac{1-\phi^2}{m\overline{\mu}} > \frac{2}{m\overline{\mu} + \phi^2 - 1 + \sqrt{(m\overline{\mu} - 1 + \phi^2)^2 + 4m\overline{\mu}}}. \qquad (41)$$

Proof: The proof is given in the appendix.

Some comments on this theorem are in order. Consider the case of a common nonstationary factor. First note that Δ_{u-f} increases with m, and so we can conclude that the average gain increases with the number of series in the system. In fact, it is easy to see that if $m = 1$ $\Delta_{u-f} = 0$, that is, in the trivial case of a simple time series the difference is zero as expected. Second, the difference also increases with the square of the factor loading matrix p_j^2. Third, the difference decreases with the ratios μ_j which give us the precision at which new information enters into the factor model.

In the stationary case when $h \to \infty$, $\Delta_{u-f} \to 0$ and there is no difference between the two models. For finite h note that in (41), the first term does not depend on m whereas the other two decrease when m increases. As shown in the appendix the first term is positive, so note that this inequality will be always true for large m.

We conclude that both for the stationary case, as well as for the nonstationary case, there is an efficiency gain from using the factor model instead of ARIMA forecasts. In both cases, this difference decreases with the precision at which new information enters into the factor model, therefore increases with the square of the corresponding factor loading and decreases with the noise variance of each series. It also increases with the number of series and the value of the AR parameter. The difference will be greater in the nonstationary case and when the univariate models are less informative about their dynamics. This is a contradiction with the result in Christoffersen and Diebold (1998), who found that there is no efficiency gain in the medium and long run forecasts in considering the cointegration relations with respect to the ARIMA univariate models. A reason for that could be that they do 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 $p_j^2 = 1, \sigma_a^2 = \sigma_j^2 = 1, \forall j = 1, ..., m$. Then $\theta_j = .38$ $\forall j = 1, ..., m$, and the relative decrease in MSE of the factor model with respect to the univariate models for the one step ahead prediction error is

$$\frac{\Delta_{u-f}}{U} = \frac{.5\sqrt{3} - .5 - 2(m + \sqrt{m^2 + 4m})^{-1}}{2 + .38/(1 - .38)}$$

which is equal to .06 for $m = 4$ and goes to .14 when $m \to \infty$. These numbers provides some indication of the advantages that we can obtain from the factor model with respect to the univariate forecasts.

6 Comparison of pooled forecasts and forecasts from univariate and the one factor model

Empirically, García Ferrer et al (1987) showed that the univariate forecast of a collection of variables improves, in the sense that the MSE of prediction decreases, when a pooling term is used in forecasting. In this section, we will show that, for the series driven by a common factor, pooled forecasts always outperform univariate ARIMA forecasts, and that factor forecasts outperforms both of them. We will show that the factor model outperforms the pooling technique even when the shrinkage is performed in more favorable conditions. Recall that the equation for pooling is (27) where $\widehat{y}_t(h) = 1/m \sum_{i=1}^{m} \widehat{y}_{i,t+h|t}$ is just a simple mean of the univariate forecasts. Therefore, the shrinkage technique should work better for the case that all the series have a similar behavior. (Otherwise a weighted mean should be preferable in order to build the pooling term). For that reason, we will assume that $\mu_j = \mu, \forall j = 1, ..., m$ (and therefore $\theta_j = \theta, \forall j = 1, ..., m$) throughout this section. The previous conjecture is showed in subsections 6.1 (theorem 7) and 6.2. Lemmas 4, 5 and 6 are simply auxiliary results needed to prove the main conjecture.

6.1 The pooled forecasts versus the univariate forecasts

Suppose that we forecast a vector of time series by using the pooling model (27) and extend this model for the extreme cases $\eta = 0, 1$. 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

$$y_{j,t+h} - \widetilde{y}_{j,t+h} = y_{j,t+h} - \hat{y}_{j,t+h|t} + \eta(\hat{y}_{j,t+h|t} - \widehat{y}_t(h)),$$

and its MSE of prediction

$$\begin{aligned} MSE(y_{j,t+h} - \widetilde{y}_{j,t+h}) &= MSE(y_{j,t+h} - \hat{y}_{j,t+h|t}) + \eta^2 E(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))^2 \\ &\quad + 2\eta E\left[(y_{j,t+h} - \hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))\right]. \end{aligned} \qquad (42)$$

The first term of the right hand side in this equality is the MSE of prediction of the ARIMA univariate forecasts and it is given by (61) in the appendix. To show that $MSE(y_{j,t+h} - \widetilde{y}_{j,t+h}) < MSE(y_{j,t+h} - \hat{y}_{j,t+h|t})$ is equivalent to show that the sum of the second and third term of the right hand side of (42) is negative. We will prove some preliminary lemmas

Lemma 4 *If $\mu_j = \mu$ and $p_j = p, \forall j = 1, ..., m$,*
(i) $\phi(\epsilon_{j,t} - \overline{\epsilon}_t) - \theta(v_{j,t} - \overline{v}_t) = (\epsilon_{j,t+1} - \overline{\epsilon}_{t+1}) - (v_{j,t+1} - \overline{v}_{t+1})$,
(ii) $E(\epsilon_{j,t+1} - \overline{\epsilon}_{t+1})^2 = \frac{m-1}{m}\sigma^2$,
(iii) $E(v_{j,t+1} - \overline{v}_{t+1})^2 = \theta^2 \frac{m-1}{m}\sigma^2 \left[1 + \frac{(\phi-\theta)^2}{1-\theta^2}\right]$,
and (iv) $E[(\epsilon_{j,t+1} - \overline{\epsilon}_{t+1})(v_{j,t+1} - \overline{v}_{t+1})] = \frac{m-1}{m}\sigma^2$,
where $\overline{\epsilon}_t = 1/m \sum_{i=1}^{m} \epsilon_{i,t}$ and $\overline{v}_t = 1/m \sum_{i=1}^{m} v_{i,t}$.

Proof: The proof is given in the appendix.

Lemma 5 *If $\mu_j = \mu$ and $p_j = p, \forall j = 1, ..., m$, $E(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))^2 = \phi^{2(h-1)} \frac{m-1}{m} \sigma_j^2 \frac{(\phi-\theta_j)^2}{1-\theta_j^2}$.*

Proof: The proof is given in the appendix.

Lemma 6 *If $\mu_j = \mu$ and $p_j = p, \forall j = 1, ..., m$, then*
(a) $E[(v_{j,t+\tau}(\epsilon_{j,t} - \overline{\epsilon}_t)] = -\frac{m-1}{m}\theta^{\tau-1}(\phi - \theta)\sigma^2, \forall h > 0;$
(b) $E[(v_{j,t+h}(v_{j,t} - \overline{v}_t)] = -\frac{m-1}{m}\theta^{\tau-1}\frac{(\phi-\theta)(1-\phi\theta)}{1-\theta^2}$
and (c) $E\left[v_{j,t+\tau}(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))\right] = -\frac{m-1}{m}\theta^{\tau-1}\phi^{h-1}\frac{(\phi-\theta)^2}{1-\theta^2}, \forall \tau > 0,$
where $\overline{e}_t = 1/m \sum_{i=1}^{m} e_{i,t}$ *and* $\overline{v}_t = 1/m \sum_{i=1}^{m} v_{i,t}$.

Proof: The proof is given in the appendix.

With these lemmas we can prove the following theorem:

Theorem 7 *For the one factor model, given by (23) and (24), the MSE of prediction obtained through pooled forecasts is smaller than the MSE of prediction obtained only through ARIMA univariate forecasts. The advantage increases with the number of series, the value of the autoregressive parameter and with the weight given to the pooling term in the pooled forecasts. Calling* $P = \sum_{j=1}^{m} MSE(y_{j,t+h} - \widetilde{y}_{j,t+h})$ *to the sum of MSE predictions of the pooled forecasts, we have*

$$P = U - \eta(2-\eta)(m-1)\phi^{2(h-1)}\sigma^2\frac{(\phi-\theta)^2}{1-\theta^2}. \tag{43}$$

Proof: The proof is given in the appendix.

The MSE of prediction 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 $\widetilde{y}_{j,t+h} = 1/m \sum_{i=1}^{m} \hat{y}_{i,t+h|t}$. When $\eta = 1$, $2\eta - \eta^2$ reaches its maximum value of 1, so we subtract the maximum possible quantity to the univariate ARIMA forecasts. Obviously, if $\eta = 0$, the predictor and the MSE of prediction are the ones corresponding to the univariate ARIMA case.

The theorem shows that the difference of MSE of prediction 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, since in this case $\theta \to \phi$ and the second term in (43) vanishes. In this case, the univariate series tend to behave as white noise and nothing is gain, in terms of efficiency, from using a linear combination of them.

6.2 Pooled versus one factor model forecasts

In this subsection, we will compare the forecast obtained from the factor model to the pooled forecasts. Let Δ_{p-f} be the difference between the sum of MSE of prediction of the pooled forecasts and the trace of the MSE of prediction matrix of the factor model, that is, $\Delta_{p-f} = P - F = \sum_{j=1}^{m}(MSE(y_{j,t+h} - \widetilde{y}_{j,t+h})) - tr(\Sigma_{t+h|t})$.

We want to show that $\Delta_{p-f} > 0$, for any η, 0 □ η □ 1. The case $\eta = 0$, collapses into the ARIMA univariate forecasts. The worst case to show the previous claim is $\eta = 1$, since it gives the minimum MSE of prediction of the pooled forecasts. We already know the improvement of the pooled forecasts with respect to the univariate ARIMA forecasts.

First, consider the nonstationary case, we can write for $\phi = 1$,

$$\Delta_{p-f} = \Delta_{u-f} - (m-1)\sigma^2 \frac{1-\theta}{1+\theta}. \tag{44}$$

>From (35) and since $\phi^2 = 1$

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

Substituting Δ_{u-f} in (44), and after some straightforward algebra, we have

$$\Delta_{p-f} = p^2 \left(\frac{1}{2} \left(\sqrt{1 + \frac{4}{\mu}} - 1 \right) \frac{m\theta + 1}{1+\theta} - \frac{2}{\mu + \sqrt{\mu^2 + 4\frac{\mu}{m}}} \right). \tag{45}$$

For $m = 1 \Rightarrow \Delta_{p-f} = 0$. In (45) 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 $1/\mu$. Therefore as the number of series considered increases the efficiency lost in prediction for not using the factor model increases. The pooled forecasts lose their advantages over the ARIMA univariate forecasts when μ decreases, since in this case $\theta \to \phi = 1$ and the second term in (44) will vanish, so $\Delta_{p-f} \to \Delta_{u-f}$. Notice that in this case the univariate series will behave as white noise and there is no improvement in terms of the MSE of prediction with the pooled forecasts with respect to the ARIMA univariate forecasts.

In the stationary case, $|\phi| < 1$, and from (43) and (36), 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 (69) from the appendix and since $\mu_j = \mu$ and $p_j = p, \forall j = 1, ..., m$

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

where $\beta = 1 - \frac{m-1}{m} \frac{\phi-\theta}{1-\theta^2}$. Notice that the first term is always positive since θ and ϕ are of the same sign and β is always positive because:(1) if $\phi = 0$, then $\theta = 0$ and $\beta = 1$; (2) if $\phi > 0, \frac{\phi-\theta}{1-\theta^2} \square \frac{1}{1+\theta} < 1$ and $\beta > 0$; (3) finally, if $\phi < 0, \frac{\phi-\theta}{1-\theta^2} < 0$ and $\beta > 1$.

To show that $\Delta_{p-f} > 0$, we have to show that

$$\frac{\phi^2 - 1 - \mu + \sqrt{(\mu + 1 + \phi^2)^2 - 4\phi^2}}{2\mu} \beta + \frac{1-\phi^2}{m\mu}$$
$$> \frac{2}{m\mu + \phi^2 - 1 + \sqrt{(m\mu - 1 + \phi^2)^2 + 4m\mu}}. \tag{46}$$

When m increases, β converges to a constant, so the first term of the left hand side in (46) also converges to a constant. The second term decreases $m\mu$, while the right hand side (46) decreases with $2m\mu$.

7 Comparison of the VARIMA forecast and the forecast of the one factor model

In this section we analyze the VARMA models derived from the factor model and it is shown that, when the parameters are known, both models have the same forecast performance. This result is to be expected, as in both cases we are taking into account the common structure. From a practical point of view, the parameters of the model will be estimated and due to the estimation problem with VARIMA models in the presence of common structures (some matrices may not be identified, see Peña and Box, 1987), we expect better results in practice with the factor model. In this section we derive the VARIMA model form of the factor model and compute their MSE of prediction.

The VARMA model that is derived from the factor model is

$$\begin{aligned} \mathbf{y}_t &= \widetilde{\mathbf{P}}\mathbf{z}_t + \boldsymbol{\epsilon}_t = \widetilde{\mathbf{P}}(\mathbf{G}\mathbf{z}_{t-1} + \mathbf{a}_t) + \boldsymbol{\epsilon}_t \\ &= \widetilde{\mathbf{P}}\mathbf{G}\widetilde{\mathbf{P}}^{-}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t + \widetilde{\mathbf{P}}\mathbf{a}_t - \widetilde{\mathbf{P}}\mathbf{G}\widetilde{\mathbf{P}}^{-}\boldsymbol{\epsilon}_{t-1} \end{aligned} \tag{47}$$

since $\mathbf{z}_{t-1} = \widetilde{\mathbf{P}}^{-}(\mathbf{y}_{t-1} - \boldsymbol{\epsilon}_{t-1})$, where $\widetilde{\mathbf{P}}^{-}$ is the generalized Moore-Penrose inverse of $\widetilde{\mathbf{P}}$, such that $\widetilde{\mathbf{P}}\widetilde{\mathbf{P}}^{-}\widetilde{\mathbf{P}} = \widetilde{\mathbf{P}}$. So, we will observe

$$\mathbf{y}_t = \boldsymbol{\Phi} y_{t-1} + \mathbf{v}_t - \boldsymbol{\theta}\mathbf{v}_{t-1} \tag{48}$$

where $\text{var}(\mathbf{v}_t) = \boldsymbol{\Sigma}_v$. From (47) and (48), we obtain the following relations between second moments

$$\boldsymbol{\Sigma}_v = \widetilde{\mathbf{P}}\boldsymbol{\Sigma}_a\widetilde{\mathbf{P}}' + \boldsymbol{\Sigma}_\epsilon + \boldsymbol{\Phi}\boldsymbol{\Sigma}_v\boldsymbol{\theta}' - \boldsymbol{\theta}\boldsymbol{\Sigma}_v\boldsymbol{\theta}',$$

$$\boldsymbol{\Sigma}_\epsilon\boldsymbol{\Phi}' = \boldsymbol{\Sigma}_v\boldsymbol{\theta}',$$

and for the AR parameters

$$\boldsymbol{\Phi}\widetilde{\mathbf{P}} = \widetilde{\mathbf{P}}\mathbf{G}.$$

For the one factor model, $\mathbf{G} = \phi$, $\mathbf{z}_t = f_t$, $\widetilde{\mathbf{P}} = \mathbf{P}$, the general solution of this system of equations can be found in Peña and Box (1987)

$$\boldsymbol{\Phi} = \phi\mathbf{P}\mathbf{P}^{-} + \mathbf{C}(\mathbf{I} - \mathbf{P}\mathbf{P}^{-}), \tag{49}$$

where $\mathbf{C}$ is an arbitrary matrix, with the only restriction that the eigenvalues of $\boldsymbol{\Phi}$ must be outside or over the unit circle. Equation (49) has not an unique solution, but once a certain solution is selected, $\boldsymbol{\Phi}$ is considered fixed and we can solve for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}_v$. It can be verified that the solution of these equations is

$$\boldsymbol{\theta} = \boldsymbol{\Phi} - \phi\frac{\lambda}{1 + \lambda m\overline{\mu}}\mathbf{P}\mathbf{P}'\boldsymbol{\Sigma}_\epsilon^{-1} \tag{50}$$

and

$$\mathbf{\Sigma}_v = \lambda \mathbf{P}\mathbf{P}' + \mathbf{\Sigma}_\epsilon, \tag{51}$$

where $m\overline{\mu} = \mathbf{P}'\mathbf{\Sigma}_\epsilon^{-1}\mathbf{P}$ and λ verifies

$$m\overline{\mu}\lambda^2 - \lambda(m\overline{\mu} + \phi^2 - 1) - 1 = 0. \tag{52}$$

>From equation (50), the difference between both parameter matrices $\mathbf{\Phi} - \boldsymbol{\theta}$ is given by

$$\mathbf{\Phi} - \boldsymbol{\theta} = \phi \frac{\lambda}{1 + \lambda m\overline{\mu}} \mathbf{P}\mathbf{P}'\mathbf{\Sigma}_\epsilon^{-1}. \tag{53}$$

Next, let $\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t}$ the forecast error with the VARMA model and $\mathbf{D}$ the MSE matrix of prediction associated to the previous error and $\Delta_{v-f} = tr(\mathbf{D} - \mathbf{\Sigma}_{t+h|t})$ the trace of the difference between the MSE matrices of the VARMA and factor models. In the appendix it is shown that when the Kalman filter reaches the steady state $\Delta_{v-f} = 0$.

8 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 some Monte Carlo studies. In a first set of simulations, we assume the most favorable situation for shrinking forecasts: a common non stationary factor (common trend) for $m = 4$ observed series, $\mathbf{P} = \mathbf{1} = p(1,1,1,1)'$, $E(\epsilon_t) = \mathbf{0}$, $\text{var}(\epsilon_t) = \mathbf{I}_4$, $E(a_t) = 0$ and $\text{var}(a_t) = \sigma_a^2 = 1$ for different values of p (1, 10 and 0.1). In this case $\mu_j = \mu, \forall j = 1, ..., m$ 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 $\sigma_a^2 = 1, 10, 0.1$. We made 1000 replications. The sample size for each of the simulated series was 124 observations. The first 104 observations were used to estimate the models and the last 20 were reserved to compute the forecasts.

We analyzed the behavior of 7 different models. First, we fitted an ARI(3) for each of the series as an approximation to the true univariate model, ARIMA(0,1,1), and compute the univariate forecasts. Then we computed pooled forecasts using (27) and the three values of η, (η = 0.25,0.50 and 0.75), which were used by García Ferrer et al (1987). We also fitted two VARMA models: a VAR(3) in levels and also a VAR(3) in first differences. Finally, we have estimated the factor model through the EM algorithm via the Kalman filter. For all these models we made forecasts $h = 1, ..., 20$ steps ahead and computed the MSE of predictions for horizons 1, 5, 10 and 20. 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 the 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 of the series μ; the third column shows the results for the univariate ARI(3) model and columns 4 to 6 show the RMSE obtained with the pooling technique for the three values of η previously indicated; columns 7 and 8 show the RMSE of the VAR(3) for the series in first differences and in levels, respectively, and finally, the last column shows the results for the factor model.

horiz.	μ	UNIV	$\eta = .25$	$\eta = .50$	$\eta = .75$	VAR diff.	VAR lev.	FACT
h=1	.1	1.1877	1.1466	1.1165	1.0984	1.2393	1.1686	1.1097
h=5	.1	1.3515	1.3114	1.2817	1.2633	1.3649	1.3011	1.2638
h=10	.1	1.5089	1.4745	1.4499	1.4355	1.5116	1.5183	1.4380
h=20	.1	1.8066	1.7768	1.7552	1.7421	1.8130	1.9063	1.7406
h=1	1	1.6567	1.6066	1.5701	1.5481	1.6604	1.5985	1.5337
h=5	1	2.5751	2.5420	2.5181	2.5034	2.5751	2.6319	2.4977
h=10	1	3.3195	3.2942	3.2760	3.2650	3.3192	3.5802	3.2668
h=20	1	4.5538	4.5362	4.5239	4.5168	4.5467	5.2523	4.5164
h=1	10	3.5331	3.4892	3.4581	3.4403	3.6629	3.6256	3.4566
h=5	10	7.3283	7.3056	7.2892	7.2793	7.3614	7.7708	7.2650
h=10	10	9.8499	9.8328	9.8206	9.8133	9.8801	10.8733	9.8297
h=20	10	14.2844	14.2722	14.2633	14.2578	14.3119	16.7690	14.2339

Table 1 Comparison of RMSE of prediction for the univariate ARIMA, pooled, vector ARMA and factor models for three different values of $\mu = .1, 1, 10$.

To give an idea of the results obtained, see for instance the central rows of the table, corresponding to the value $\mu = 1$. In this case 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 $\eta = .75$ 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 performs very similar to the factor model for all horizons. The factor model provides better forecasts than both VARMA models. The gains of the factor model with respect to the VARMA models are similar for all values of μ and sometimes are also considerable for the long periods of forecasting. For example, for $\mu = .1$ the gain for h=10 is 0.0470 and for h=20 is 0.0365.

In a second set of simulations we allow the precisions in estimating the factor associated to each of the series, μ_i, to be very different. We will assume the same set of parameters of the previous simulation, but for $\text{var}(\epsilon_t) = diag(.1^2, .5^2, 1^2, 3^2)$ and $\mathbf{P} = (1, .5, .2, .05)'$. The first column shows the horizon of prediction; the second one shows the results for the univariate ARI(3) model and columns 3 to 5 show the RMSE obtained with the pooling technique for the three values of η previously indicated; columns 6 and 7 show the RMSE of the VAR(3) for the series first differences and in levels, respectively, and finally, the last column shows the results for the factor model.

horiz.	UNIV	$\eta = .25$	$\eta = .50$	$\eta = .75$	VAR diff.	VAR lev.	FACT
h=1	1.8987	1.9819	2.3761	2.9595	1.9629	1.7949	1.7206
h=5	2.2157	2.2681	2.6030	3.1310	2.2462	2.0968	2.0569
h=10	2.5147	2.5373	2.8198	3.2962	2.5199	2.4925	2.3537
h=20	3.1215	3.1366	3.3663	3.7716	3.1290	3.3668	2.9916

Table 2 Comparison of RMSE of prediction for $\mu_1 = 100, \mu_2 = 2, \mu_3 = .2$ and $\mu_4 = .0056$.

The results of Table 2 correspond to the case in which each of the series behaves in a different way. In this case the pooling techniques are not so successful in forecasting, while the factor model keeps maintaining considerable gains over the remaining models. For instance, Table 2 shows that the factor model improves the univariate forecasts by $(1.8987 - 1.7206)/1.8987 = 0.0938$, for h=1, by 0.064 for h=10 and by 0.0416 for h=20. The pooling methods show the worst forecasting performance of all the models tested. The best forecasts with the pooling method correspond to $\eta = .25$, and as η increases the behavior of the pooling method worsens. The factor model provides better forecasts, as expected, than both VARMA models for all the horizons of prediction.

9 An economic application

The data we considered are annual observations of the real GNP, from 1949-1997, for some OECD countries. An extended data base, from 1948-1986, was analyzed by García-Ferrer et al (1987), who considered several alternatives to forecast the output growth rates defined as $g_t = \ln(\frac{O_t}{O_{t-1}})$, where O_t is real GNP, for several OECD countries. Forecasts were compared by the root mean square error of prediction (RMSE) for one step ahead forecasts. The problem was further studied by Zellner and Hong (1989, 1991), Mittnik (1990), Zellner et al (1991), Min and Zellner (1993), García-Ferrer et al (1996), Li and Dorfman (1996) and Zellner and Min (1998). García-Ferrer and Poncela (1999) based on empirical information and historical and geographical considerations made three groups of countries.

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

Figure 1 goes around here.

The difficulty in forecasting this data set is the presence of several turning points. We will show that the RMSE of prediction decreases in a dynamic factor model with respect to an ARIMA univariate model and pooled forecasts, when increasing the numbers 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 mean square error (MSE) of prediction for each country.

In order to achieve a systematic procedure for our comparison, and as in García-Ferrer et al (1987) and García-Ferrer and Poncela (1999), we fitted and AR(3) model for each growth rate.

$$g_{it} = \beta_{0i} + \beta_{1i} g_{it-1} + \beta_{2i} g_{it-2} + \beta_{3i} g_{it-3} + \varepsilon_{it}. \tag{54}$$

Then, we built a factor model. Applying the results in Peña and Poncela (1999), we found a common trend plus a common AR(1) stationary factor. Let $\mathbf{y}_t = (y_{1,t}, ..., y_{m,t})'$, $y_{i,t} = \log(O_{i,t})$, $i = 1, ..., m$, $t = 1, ..., T$, the factor model could be written as

$$\underset{m \times 1}{\mathbf{y}_t} = \underset{m \times r}{\mathbf{P}} \; \underset{r \times 1}{\mathbf{f}_t} + \underset{m \times 1}{\mathbf{n}_t}, \tag{55}$$

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,

$$\begin{bmatrix} T_t \\ f_{2,t} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} T_{t-1} \\ f_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \tag{56}$$

where $\mathbf{a}_t = (a_{1,t} \quad a_{2,t})' \sim N_2(\mathbf{0}, \mathbf{\Sigma}_a)$, is serially uncorrelated, $\mathrm{E}(\mathbf{a}_t \mathbf{a}_{t-h}') = \mathbf{0}, \quad h \neq 0$. After extracting the common dynamic structure, we fitted a univariate AR(3) for each of the specific components in order to capture the remaining dynamic structure

$$n_{it} = \alpha_{0i} + \alpha_{1i} n_{it-1} + \alpha_{2i} n_{it-2} + \alpha_{3i} n_{it-3} + e_{it}. \tag{57}$$

The sequence of vectors $\mathbf{e}_t = (e_{1,t}, ..., e_{m,t})'$ 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 repeat the process discarding one country at the time. The series were discarded in a Spanish alphabetical order (Spain starts with ESP in Spanish). The results are shown in table 3. The first column has the number of series considered in each of the factor models. Columns 2 to 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.

m	BEL	FRA	SPA	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

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

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 4 shows a comparison of the factor model, ARI(3) model and pooled forecasts.

Model	BEL	FRA	SPA	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	1.64	1.69	1.82	1.79	1.55	1.79
P, $\eta = 0.50$	1.88	1.63	1.62	1.93	1.77	1.77	1.77
P, $\eta = 0.25$	1.90	1.63	1.57	2.06	1.77	1.79	1.77
ARI(3)	1.92	1.64	1.54	2.22	1.77	1.82	1.77

Table 4: RMSE of prediction for each country for the complete factor model, pooled forecasts for three values of η, 0.75,0.50 and 0.25 and an ARI(3) model.

It is clearly seen 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 a greater weight to the pooling term the mean and, therefore, the sum of RMSE of predictions decreases. The ARI(3) 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 4, discarding one series at the time. The series are discarded in Spanish alphabetical order. The results are shown on tables 5, 6 and 7.

Model	BEL	FRA	SPA	HOL	Mean	Median
FM	1.45	1.57	1.31	1.90	1.56	1.51
P, $\eta = 0.75$	1.90	1.66	1.71	1.85	1.78	1.78
P, $\eta = 0.50$	1.90	1.64	1.63	1.96	1.78	1.77
P, $\eta = 0.25$	1.90	1.64	1.58	2.08	1.80	1.78
ARI(3)	1.92	1.64	1.54	2.22	1.83	1.78

Table 5: RMSE of prediction for each country for the factor model with four series, pooled forecasts for three values of η, 0.75,0.50 and 0.25 and an ARI(3) model.

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

Table 6: RMSE of prediction for each country for the factor model with three series, pooled forecasts for three values of η, 0.75,0.50 and 0.25 and an ARI(3) model.

Model	BEL	FRA	Mean	Median
FM	1.82	2.55	2.18	2.18
P, $\eta = 0.75$	1.85	1.64	1.75	1.75
P, $\eta = 0.50$	1.87	1.64	1.76	1.76
P, $\eta = 0.25$	1.90	1.64	1.77	1.77
ARI(3)	1.92	1.64	1.78	1.78

Table 7: RMSE of prediction for each country for the factor model with two series, pooled forecasts for three values of η, 0.75,0.50 and 0.25 and an ARI(3) model.

Again, the factor model gives the best results and the ARI(3) model the worst ones. The effect of the number of series is consistent in the sample.

10 Conclusions

We have shown that, under certain restrictions, the forecasting equations of the factor model imply a pooling term. In particular, when the noise variance associated to the observed series is a diagonal matrix, the forecast incorporates a weighted sum of all the variables observed in t. For the one factor model, and under some further assumptions, the forecasting equation of the observed series imply the sample mean as the pooling term.

For the one factor model and assuming that the parameters are known, we have shown the gain in efficiency, in terms of the MSE of prediction, of the factor model with respect to univariate ARIMA and pooled forecasts. We have also shown that the pooled forecasts have a smaller MSE of prediction that the univariate forecasts, due to the introduction of some common information. On the steady state, we have shown that the factor model presents a similar MSE of prediction than the VARIMA model. Therefore, for the one factor model with known parameters, we have that

$$MSE_u > MSE_p > MSE_f = MSE_v,$$

where MSE_f and MSE_v are the trace of MSE of predictions of the factor and VARMA model, respectively, and MSE_u and MSE_p are the sum of MSE of prediction of the univariate ARIMA model and pooled forecasts. When the parameters are unknown we have shown by Monte Carlo that the factor model behaves better than the multivariate ARMA model.

11 Appendix

Proof of (21):

Starting from (18) and (11), and using (8) we have that

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

and calling $\mathbf{A}_2 = \tilde{\mathbf{P}}\mathbf{N}_t\tilde{\mathbf{P}}'\mathbf{\Sigma}_\epsilon^{-1}$, we can write $\mathbf{N}_t = \mathbf{G}^h\mathbf{V}_{t|t-1}(\mathbf{I}_s - \tilde{\mathbf{P}}'\mathbf{\Sigma}_\epsilon^{-1}\tilde{\mathbf{P}}(\tilde{\mathbf{P}}'\mathbf{\Sigma}_\epsilon^{-1}\tilde{\mathbf{P}} + \mathbf{V}_{t|t-1}^{-1})^{-1})$ and computing again the inverse of $(\tilde{\mathbf{P}}'\mathbf{\Sigma}_\epsilon^{-1}\tilde{\mathbf{P}} + \mathbf{V}_{t|t-1}^{-1})$, we obtain (21).

Proof of Lemma 1:

1. Equating variances in both sides of (33)

$$\widetilde{\sigma_j}^2 + \theta_j^2\widetilde{\sigma_j}^2 = p_j^2 + \sigma_j^2 + \phi^2\sigma_j{}^2, \tag{58}$$

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 (35), also solving for $\widetilde{\sigma_j}^2$ then $\widetilde{\sigma_j}^2 = p_j^2 + (1 + (\phi - \theta_j)\phi)\sigma_j{}^2$. Introducing (36) in the last equation $\widetilde{\sigma_j}^2 = \sigma_j^2 + \left(1 + \frac{\phi\theta_j}{1-\phi\theta_j}\right)p_j^2 = \sigma_j^2 + \frac{p_j^2}{1-\phi\theta_j}$.

2. To show that $\mathrm{E}(a_t v_{j,\tau}) = 0$, we solve for $v_{j,\tau}$ in (33) 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 (33) and introduce it in $\mathrm{E}(a_t v_{j,t}) = \mathrm{E}(a_t(p_j a_t + \epsilon_{j,t} - \phi\epsilon_{j,t-1} + \theta_j v_{j,t-1})) = p_j$, $\forall j = 1, ..., m$.

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

5. Applying (33) in time $j+h$ and by backward substitution $v_{j,t+h}$ h times, $\mathrm{E}(\epsilon_{j,t}v_{j,t+h}) = -\theta_j^{h-1}(\phi - \theta_j)\sigma_j^2$ $\forall j, i = 1, ..., m,\ j \neq i, \forall j = 1, ..., m, \forall h > 0$.

6. To show $\mathrm{E}(\epsilon_{j,t}v_{i,\tau}) = 0$, $\forall j, i = 1, ..., m;\ j \neq i, \forall\tau$ integer, is immediately from backward substitution of $v_{i,\tau}$ τ times from (33) and applying hypothesis (ii).

7. From (33), $\widetilde{\sigma_{ji}} = E(v_{j,t}, v_{i,t}) == E\left[(p_j a_t + \epsilon_{j,t} - \phi\epsilon_{j,t-1} + \theta_j v_{j,t-1})(p_i a_t + \epsilon_{i,t} - \phi\epsilon_{i,t-1} + \theta_i v_{i,t-1})\right] = p_j p_i + \theta_j\theta_i\widetilde{\sigma_{ji}}$. Solving for $\widetilde{\sigma_{ji}}$, we get $\widetilde{\sigma_{ji}} = \frac{p_i p_j}{1-\theta_j\theta_i}$ $\forall j, i = 1, ..., m;\ j \neq i$.

8. Applying recursively (33) h times and from parts 2 and 6 of this lemma, $E(v_{j,t+h}, v_{i,t}) = E\left[(p_j a_{t+h} + \epsilon_{j,t+h} - \phi\epsilon_{j,t+h-1} + \theta_j v_{j,t+h-1})v_{i,t}\right] = \theta_j E(v_{j,t+h-1}v_{i,t}) = \theta_j^h E(v_{j,t}v_{i,t}) = \theta_j^h\frac{p_i p_j}{1-\theta_j\theta_i}$.

Proof of Lemma 2:

The forecast of the observed series h steps ahead is given by

$$\hat{y}_{j,t+h|t} = \phi^h y_{j,t} - \phi^{h-1}\theta_j v_{j,t}, \tag{59}$$

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

$$y_{j,t+h} - \hat{y}_{j,t+h|t} = v_{j,t+h} + \sum_{i=1}^{h-1}\phi^{i-1}(\phi - \theta_j)v_{j,t+h-i}, \tag{60}$$

and the MSE of prediction

$$\begin{aligned} MSE(y_{j,t+h} - \hat{y}_{j,t+h|t}) &= \widetilde{\sigma_j}^2\left(1 + (\phi - \theta_j)^2\sum_{i=1}^{h-1}\phi^{2(i-1)}\right) \\ &= \left(\sigma_j^2 + \frac{p_j^2}{1-\phi\theta_j}\right)\left(1 + (\phi - \theta_j)^2\sum_{i=1}^{h-1}\phi^{2(i-1)}\right), \end{aligned} \tag{61}$$

where the last equality is obtained replacing $\widehat{\sigma_j}^2$ by its expression given in lemma 1.1. 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 (36) and after some straightforward algebra we finally have

$$U=\sum_{j=1}^{m}\left(\sigma_j^2+\frac{p_j^2\theta_j}{\phi(1-\phi\theta_j)}+\frac{\phi-\theta_j}{(1-\theta_j\phi)\phi}p_j^2\sum_{i=0}^{h-1}\phi^{2i}\right). \tag{62}$$

Proof of Theorem 3:

For the one factor model, substituting (13) in (15), with Σ_u=1 by the identification restriction and $\mathbf{G}=\phi$, we obtain that the trace of the MSE of predictions matrix is

$$tr(\mathbf{\Sigma}_{t+h|t})=\sum_{j=1}^{m}p_j^2\left(V_{t|t}\phi^{2h}+\sum_{i=0}^{h-1}\phi^{2i}\right)+\sum_{j=1}^{m}\sigma_j^2. \tag{63}$$

The difference between (37) and (63), after some straightforward algebra, is

$$\Delta_{u-f}=\sum_{j=1}^{m}p_j^2\left(\frac{\theta_j}{\phi(1-\phi\theta_j)}-V_{t|t}\phi^{2h}\right)-\sum_{j=1}^{m}\left(p_j^2\frac{\theta_j(1-\phi^2)}{(1-\theta_j\phi)\phi}\sum_{i=0}^{h-1}\phi^{2i}\right). \tag{64}$$

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

$$\Delta_{u-f}=\sum_{j=1}^{m}p_j^2\left(\frac{\theta_j}{\phi(1-\phi\theta_j)}-V_{t|t}\right)\phi^{2h}. \tag{65}$$

We will compute Δ_{u-f} assuming that the Kalman filter has reached the steady state. Pre- and postmultiplying $\mathbf{\Sigma}_{t|t-1}^{-1}$ by $\mathbf{P}'\mathbf{P}$ and from (8) and using again the lemma for the inverse of the sum of two matrices, (Rao, 1973), $\mathbf{P}'\mathbf{\Sigma}_{t|t-1}^{-1}\mathbf{P}=\frac{m\overline{\mu}}{1+m\overline{\mu}V_{t|t-1}}$. Now, from (10), $V_{t|t}$ can be written as $V_{t|t}=\frac{V_{t|t-1}}{1+m\overline{\mu}V_{t|t-1}}$. Substituting $V_{t-1|t-1}$ as given by (10) in (6) and assuming that the filter reaches the steady state, $V_{t|t-1}=V_{t-1|t-2}=V$, we obtain the algebraic Riccatti equation

$$V(1-\phi^2)+V^2\phi^2\mathbf{P}'(V\mathbf{P}\mathbf{P}'+\mathbf{\Sigma}_\epsilon)^{-1}\mathbf{P}-1=0.$$

that can be written, using again the lemma for the inverse of the sum of two matrices (Rao, 1973), as

$$V(1-\phi^2)+\frac{V^2m\overline{\mu}}{Vm\overline{\mu}+1}\phi^2-1=0 \tag{66}$$

and the steady state of $V_{t|t}$ is

$$\frac{V}{Vm\overline{\mu}+1}=\frac{1}{Vm\overline{\mu}\phi^2}-\frac{1-\phi^2}{m\overline{\mu}\phi^2}. \tag{67}$$

Also from (66)

$$V^2 m\overline{\mu} - V(m\overline{\mu} + \phi^2 - 1) - 1 = 0. \tag{68}$$

>From (65) and since the steady state of $V_{t|t}$ is given by (67)

$$\Delta_{u-f} = \phi^{2(h-1)} \sum_{j=1}^{m} p_j^2 \left(\frac{\phi\theta_j}{(1-\phi\theta_j)} - \frac{1}{V m\overline{\mu}} + \frac{1-\phi^2}{m\overline{\mu}} \right). \tag{69}$$

Now from (36) and (35), and inserting in this equation V as the positive solution of (68) we have that

$$\begin{aligned} \Delta_{u-f} &= \phi^{2(h-1)} \times \\ & \sum_{j=1}^{m} p_j^2 \left(\frac{\phi^2 - 1 - \mu_j + \sqrt{(\mu_j + 1 + \phi^2)^2 - 4\phi^2}}{2\mu_j} + \frac{1-\phi^2}{m\overline{\mu}} - \frac{2}{m\overline{\mu} + \phi^2 - 1 + \sqrt{(m\overline{\mu} + \phi^2 - 1)^2 + 4m\overline{\mu}}} \right). \end{aligned} \tag{70}$$

where note that the first term is positive because $(\mu_j + 1 + \phi^2)^2 - 4\phi^2 = (\mu_j + 1 - \phi^2)^2 + 4\mu_j\phi^2$; and the last two terms go to 0 as $m \to \infty$.

Let us prove that in the nonstationary case if $m > 1$ then $\Delta_{u-f} > 0$. Inserting $\phi = 1$ in (70), we have

$$\Delta_{u-f} = \sum_{j=1}^{m} p_j^2 \left(\frac{\sqrt{\mu_j^2 + 4\mu_j} - \mu_j}{2\mu_j} - \frac{2}{m\overline{\mu} + \sqrt{m^2\overline{\mu}^2 + 4m\overline{\mu}}} \right);$$

we want to prove now that this difference is positive, that is, for $m > 1$,

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

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

For the stationary case, in the long run, when $h \to \infty$, there is no difference between the two models

$$\lim \Delta_{u-f} = \lim U - tr(\Sigma_{t+h|t}) = 0;$$

for finite h, a sufficient condition for $\Delta_{u-f} \geq 0$ is

$$\frac{\phi^2 - 1 - \mu_j + \sqrt{(\mu_j + 1 + \phi^2)^2 - 4\phi^2}}{2\mu_j} + \frac{1-\phi^2}{m\overline{\mu}} > \frac{2}{m\overline{\mu} + \phi^2 - 1 + \sqrt{(m\overline{\mu} - 1 + \phi^2)^2 + 4m\overline{\mu}}}. \tag{71}$$

Proof of Lemma 4:

(i) From (33),

$$\phi\epsilon_{j,t} - \theta v_{j,t} = pa_{t+1} + \epsilon_{j,t+1} - v_{j,t+1}; \tag{72}$$

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

$$\phi\overline{\epsilon}_t - \theta\overline{v}_t = pa_{t+1} + \overline{\epsilon}_{t+1} - \overline{v}_{t+1}, \tag{73}$$

where $\overline{\epsilon}_t = 1/m\sum_{i=1}^{m}\epsilon_{i,t}$ y $\overline{v}_t = 1/m\sum_{i=1}^{m}v_{i,t}$. Subtracting (73) from (72) we get the desired result.

(ii) Since $\mu_j = \mu, \forall j = 1, ..., m \Rightarrow \sigma_j^2 = \sigma^2, \forall j = 1, ..., m$, and $E(\epsilon_{j,t+1} - \overline{\epsilon}_{t+1})^2 = E\left(\frac{m-1}{m}\epsilon_{j,t+1} - \frac{1}{m}\sum_{i=1,i\neq j}^{m}\epsilon_{i,t+1}\right)^2 = \left(\frac{m-1}{m}\right)^2\sigma^2 + \frac{m-1}{m^2}\sigma^2 = \frac{m-1}{m}\sigma^2$.

(iii) $E(v_{j,t+1} - \overline{v}_{t+1})^2 = E\left(\frac{m-1}{m}v_{j,t+1} - \frac{1}{m}\sum_{i=1,i\neq j}^{m}v_{i,t+1}\right)^2$; expanding the square and taking into account $\mu_j = \mu, \forall j = 1, ..., m \Rightarrow \widetilde{\sigma}_j^2 = \widetilde{\sigma}^2, \forall j = 1, ..., m$, and that $\widetilde{\sigma}_{ij}$ is the same $\forall j, i = 1, ..., m; j \neq i$, ($\widetilde{\sigma}_{ij}$ has been defined in Lemma 1, part 7)

$$\begin{aligned} E(v_{j,t+1} - \overline{v}_{t+1})^2 &= \left(\frac{m-1}{m}\right)^2 E(v_{j,t+1}^2) + \frac{1}{m^2}E\left(\sum_{i=1,i\neq j}^{m} v_{i,t+1}\right)^2 - 2\frac{m-1}{m^2}E\left(v_{j,t+1}\sum_{i=1,i\neq j}^{m} v_{i,t+1}\right) \\ &= \left(\frac{m-1}{m}\right)^2\widetilde{\sigma}^2 + \frac{m-1}{m^2}\widetilde{\sigma}^2 + \frac{(m-1)(m-2)}{m^2}\widetilde{\sigma}_{ij} - \left(\frac{m-1}{m}\right)^2\widetilde{\sigma}_{ij} \\ &= \frac{m-1}{m}\left(\widetilde{\sigma}^2 - \widetilde{\sigma}_{ij}\right). \end{aligned}$$

By lemma 1, parts 1 and 7, $E(v_{j,t+1} - \overline{v}_{t+1})^2 = \frac{m-1}{m}\left(\sigma^2 + p^2\left(\frac{1}{1-\phi\theta} - \frac{1}{1-\theta^2}\right)\right) = \frac{m-1}{m}\sigma^2\left(1 + \frac{(\phi-\theta)^2}{1-\theta^2}\right)$.

(iv) $E[(\epsilon_{j,t+1} - \overline{\epsilon}_{t+1})(v_{j,t+1} - \overline{v}_{t+1})] = E(\epsilon_{j,t+1}v_{j,t+1} - \overline{\epsilon}_{t+1}v_{j,t+1} + \epsilon_{j,t+1}\overline{v}_{t+1} + \overline{\epsilon}_{t+1}\overline{v}_{t+1}) = \sigma^2 - \frac{\sigma^2}{m} + \frac{\sigma^2}{m} + \frac{\sigma^2}{m} = \frac{m-1}{m}\sigma^2$.

Proof of lemma 5:

First, we will calculate the difference between the univariate forecast and the mean of the univariate forecasts. From (59), and substituting $y_{j,t}$ by its expression as a function of the common factor

$$\begin{aligned} \hat{y}_{j,t+h|t} - \widehat{y}_t(h) &= \phi^h y_{j,t} - \phi^{h-1}\theta_j v_{j,t} - (\phi^h\overline{y}_t - \phi^{h-1}\theta\overline{v}_t) \\ &= \phi^h(pf_t + \epsilon_{j,t}) - \phi^{h-1}\theta_j v_{j,t} - \phi^h(pf_t + \overline{\epsilon}_t) + \phi^{h-1}\theta\overline{v}_t \\ &= \phi^h(\epsilon_{j,t} - \overline{\epsilon}_t) - \phi^{h-1}\theta_j(v_{j,t} - \overline{v}_t) \\ &= \phi^{h-1}(\epsilon_{j,t+1} - \overline{\epsilon}_{t+1} - (v_{j,t+1} - \overline{v}_{t+1})) \end{aligned} \tag{74}$$

where the last equality comes from part (i) of lemma 4. Take the expectation of the square of this expression

$$\begin{aligned} E(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))^2 &= \phi^{2(h-1)}[E(\epsilon_{j,t+1} - \overline{\epsilon}_{t+1})^2 + E(v_{j,t+1} - \overline{v}_{t+1})^2 \\ &\quad -2E(\epsilon_{j,t+1} - \overline{\epsilon}_{t+1})(v_{j,t+1} - \overline{v}_{t+1})]. \end{aligned}$$

Parts (ii), (iii) and (iv) of lemma 5 allows to write

$$
\begin{aligned}
E(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))^2 &= \phi^{2(h-1)}\left(\frac{m-1}{m}\sigma^2 + \frac{m-1}{m}\sigma^2\left(1+\frac{(\phi-\theta)^2}{1-\theta^2}\right) - 2\frac{m-1}{m}\sigma^2\right) \\
&= \phi^{2(h-1)}\frac{m-1}{m}\sigma^2\frac{(\phi-\theta)^2}{1-\theta^2}. \qquad (75)
\end{aligned}
$$

Proof of Lemma 6

To proof (a), applying parts 6 and 7 of lemma 1

$$
\begin{aligned}
E[(v_{j,t+\tau}(\epsilon_{j,t} - \overline{\epsilon}_t)] &= E\left(\frac{m}{m-1}v_{j,t+\tau}\epsilon_{j,t} - \frac{1}{m}v_{j,t+\tau}\sum_{i=1,i\neq j}^{m}\epsilon_{i,t}\right) \\
&= E\left(\frac{m}{m-1}v_{j,t+\tau}\epsilon_{j,t}\right) = -\frac{m}{m-1}\theta^{\tau-1}(\phi-\theta)\sigma^2,
\end{aligned}
$$

we get the desired result.

To proof (b), after some straightforward algebra and applying part 8 of lemma 1

$$
\begin{aligned}
E[(v_{j,t+\tau}(v_{j,t} - \overline{v}_t)] &= E\left(\frac{m}{m-1}v_{j,t+\tau}v_{j,t} - \frac{1}{m}v_{j,t+\tau}\sum_{i=1,i\neq j}^{m}v_{i,t}\right) \\
&= -\frac{1}{m}\sum_{i=1,i\neq j}^{m}E\left(v_{j,t+\tau}v_{i,t}\right) = -\frac{m-1}{m}\theta^{\tau}\frac{p^2}{1-\theta^2}.
\end{aligned}
$$

>From (36), the previous expression can be written as

$$
E[(v_{j,t+\tau}(v_{j,t} - \overline{v}_t)] = -\frac{m-1}{m}\theta^{\tau-1}\frac{(\phi-\theta)(1-\phi\theta)}{1-\theta^2}.
$$

Once (a) and (b) have been proven, it is easy to show (c) $E\left[v_{j,t+\tau}(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))\right] = -\frac{m-1}{m}\theta^{\tau-1}\phi^{h-1}\frac{(\phi-\theta)^2}{1-\theta^2}, \forall\tau > 0$. From (74), taking into account (a) and (b), and after some straightforward algebra

$$
\begin{aligned}
E\left[v_{j,t+\tau}(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))\right] &= \phi^{h-1}\left(\phi E(v_{j,t+\tau}(\epsilon_{j,t} - \overline{\epsilon}_t)) - \theta E(v_{j,t+\tau}(v_{j,t} - \overline{v}_t))\right) \\
&= -\phi^{h-1}\frac{m-1}{m}\theta^{\tau-1}(\phi-\theta)\sigma^2\left(\phi - \frac{\phi(1-\phi\theta)}{1-\theta^2}\right) \\
&= -\phi^{h-1}\frac{m-1}{m}\theta^{\tau-1}\frac{(\phi-\theta)^2}{1-\theta^2}\sigma^2.
\end{aligned}
$$

Proof of theorem 7:

The proof is structured in two parts. First, from lemma 6 we will show that $E\left[(y_{j,t+h} - \hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))\right] < 0, \forall h > 0$ and will give its expression. And second, we will show that the pooling forecasts methods produces a MSE of prediction smaller than the one obtained with the ARIMA univariate forecasts.

(i) Proof of $E\left[(y_{j,t+h} - \hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))\right] < 0$, $\forall h > 0$. From (60) and part (c) of lemma 6,

$$\begin{aligned} E\left[(y_{j,t+h} - \hat{y}_{j,t+h|t})(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))\right] &= \\ &= E\left[\left(v_{j,t+h} + \sum_{i=1}^{h-1} \phi^{i-1}(\phi-\theta)v_{j,t+h-i}\right)\left(\hat{y}_{j,t+h|t} - \widehat{y}_t(h)\right)\right] \\ &= -\frac{m-1}{m}\frac{(\phi-\theta)^2}{1-\theta^2}\phi^{h-1}\sigma^2 \times \left(\theta^{h-1} + (\phi-\theta)\sum_{i=1}^{h-1}\phi^{i-1}\theta^{h-i-1}\right). \end{aligned}$$

To solve the parenthesis, take ϕ^{h-2} as common factor in the summation, $\sum_{i=1}^{h-1}\phi^{i-1}\theta^{h-i-1} = \phi^{h-2}\sum_{i=0}^{h-2}\left(\frac{\theta}{\phi}\right)^i$; where now the sum is the geometric series of general term $\frac{\theta}{\phi}$. Since $\frac{\theta}{\phi}$ is always smaller than 1, (but for the trivial case $\phi = 0$, which we do not consider here), the general term of the geometric series $\left(\frac{\theta}{\phi}\right)^i$ converges to zero as $i \to \infty$. Therefore, it is straightforward to show that $\theta^{h-1} + (\phi-\theta)\sum_{i=1}^{h-1}\phi^{i-1}\theta^{h-1-i} = \phi^{h-1}$. Introducing this last expression in (76)

$$E((y_{j,t+s} - \hat{y}_{j,t+s|t})(\hat{y}_{j,t+s|t} - \widehat{y}_t)) = -\frac{m-1}{m}\phi^{2(h-1)}\sigma^2\frac{(\phi-\theta)^2}{1-\theta^2}, \tag{76}$$

which obviously is negative as we have expected.

Finally the expression of the MSE of prediction when we use pooled forecasts is obtained substituting (75) and (76) in (42),

$$\begin{aligned} MSE(y_{j,t+h} - \widetilde{y}_{j,t+h}) &= MSE(y_{j,t+h} - \hat{y}_{j,t+h|t}) + \eta^2 E(\hat{y}_{j,t+h|t} - \widehat{y}_t(h))^2 \\ &\quad + 2\eta E((y_{j,t+s} - \hat{y}_{j,t+s|t})(\hat{y}_{j,t+s|t} - \widehat{y}_t)) \\ &= MSE(y_{j,t+h} - \hat{y}_{j,t+h|t}) - \eta(2-\eta)\frac{m-1}{m}\phi^{2(h-1)}\sigma^2\frac{(\phi-\theta)^2}{1-\theta^2}. \end{aligned}$$

Since η is between 0 and 1, the second term of the previous equality is always positive and since it is subtracting, the MSE of prediction obtained with pooling forecasts is always smaller to the one obtained only from the ARIMA univariate forecasts, since we always add a negative amount to this last one. Summing up for all $j = 1, ..., m$ we finally have (43).

Proof of the results of section 7:

The forecast of the observed series h steps ahead is $\hat{\mathbf{y}}_{t+h|t} = \mathbf{\Phi}^h\mathbf{y}_t - \mathbf{\Phi}^{h-1}\boldsymbol{\theta}\,\mathbf{v}_t$. The true value $\mathbf{y}_{t+h}$ can be written as $\mathbf{y}_{t+h} = \mathbf{\Phi}^h\mathbf{y}_t + \mathbf{v}_{t+h} + \sum_{i=1}^{h-1}\mathbf{\Phi}^{i-1}(\mathbf{\Phi}-\boldsymbol{\theta})\mathbf{v}_{t+h-i} - \mathbf{\Phi}^{h-1}\boldsymbol{\theta}\mathbf{v}_t$, so the forecast error is

$$\begin{aligned} \mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t} &= \mathbf{v}_{t+h} + \sum_{i=1}^{h-1}\mathbf{\Phi}^{i-1}(\mathbf{\Phi}-\boldsymbol{\theta})\mathbf{v}_{t+h-i} \\ &= \mathbf{v}_{t+h} + \sum_{i=1}^{h-1}\mathbf{\Phi}^{i-1}\frac{\lambda}{1+\lambda m\overline{\mu}}\mathbf{P}\mathbf{P}'\mathbf{\Sigma}_\epsilon^{-1}\mathbf{v}_{t+h-i}, \end{aligned} \tag{77}$$

where the last equality is obtained form (53). From (51) and since $\mathbf{P}'\mathbf{\Sigma}_\epsilon^{-1}\mathbf{\Sigma}_v\mathbf{\Sigma}_\epsilon^{-1}\mathbf{P} = m\overline{\mu}(1+\lambda m\overline{\mu})$, the matrix of MSE of predictions $\mathbf{D}$ is given by $\mathbf{D} = MSE(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h|t})$ and

$$\begin{aligned}\mathbf{D} &= \mathbf{\Sigma}_v + \left(\frac{\lambda\phi}{1+\lambda m\overline{\mu}}\right)^2 \sum_{i=1}^{h-1} \mathbf{\Phi}^{i-1}\mathbf{P}\mathbf{P}'\mathbf{\Sigma}_\epsilon^{-1}\mathbf{\Sigma}_v\mathbf{\Sigma}_\epsilon^{-1}\mathbf{P}\mathbf{P}'(\mathbf{\Phi}^{i-1})' \\ &= \lambda\mathbf{P}\mathbf{P}' + \mathbf{\Sigma}_\epsilon + \frac{\phi^2\lambda^2 m\overline{\mu}}{1+\lambda m\overline{\mu}} \sum_{i=1}^{h-1} \mathbf{\Phi}^{i-1}\mathbf{P}\mathbf{P}'(\mathbf{\Phi}^{i-1})'.\end{aligned}$$

If $\mathbf{C} = \mathbf{I}_m$ in (49), $\mathbf{\Phi} = \phi\mathbf{I}$, so the MSE of prediction matrix is $\mathbf{D} = \mathbf{\Sigma}_\epsilon + \left(1 + \frac{\lambda m\overline{\mu}}{1+\lambda m\overline{\mu}} \sum_{i=1}^{h-1}\phi^{2i}\right)\lambda\mathbf{P}\mathbf{P}'$ and the trace of the MSE of prediction matrix

$$tr(\mathbf{D}) = \sum_{j=1}^{m} \sigma_j^2 + \lambda p_j^2 \left(1 + \frac{\lambda m\overline{\mu}}{1+\lambda m\overline{\mu}} \sum_{i=1}^{h-1}\phi^{2i}\right). \tag{78}$$

and from (78), (15) and (13)

$$\Delta_{v-f} = \sum_{i=1}^{m} p_i^2 \left(\lambda + \left(\frac{\lambda^2 m\overline{\mu}}{1+\lambda m\overline{\mu}} - 1\right)\sum_{j=0}^{h-1}\phi^{2j} - \frac{\lambda^2 m\overline{\mu}}{1+\lambda m\overline{\mu}} - V_{t|t}\phi^{2h}\right). \tag{79}$$

We will analyze the previous formula for the two cases of interest considered in the paper. Starting with the nonstationary common factor, $\phi = 1$ and $\mathbf{C} = \mathbf{I}_m$, all the series will have the same autoregressive structure, $\mathbf{\Phi} = \phi\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}' + \mathbf{C}(\mathbf{I} - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}') = \mathbf{I}_m$, which indicates that the series follow a VARIMA (0,1,1) with $\boldsymbol{\theta} = \mathbf{I}_m - \frac{\lambda}{1+\lambda m\overline{\mu}}\mathbf{P}\mathbf{P}'\mathbf{\Sigma}_\epsilon^{-1}$ and $\mathbf{\Sigma}_v = \lambda\mathbf{P}\mathbf{P}' + \mathbf{\Sigma}_\epsilon$, and λ verifies

$$m\overline{\mu}\lambda^2 - \lambda m\overline{\mu} - 1 = 0. \tag{80}$$

The trace of the difference between the MSE of predictions obtained through a VARIMA model and a factor model is given by (79) when $\phi = 1$, therefore and after some straightforward algebra

$$\Delta_{v-f} = \sum_{i=1}^{m} p_i^2 \left(\frac{\lambda}{1+\lambda m\overline{\mu}} - V_{t|t}\right)$$

since λ satisfies (80). The previous equation states that the difference between both methods is due to the difference between $V_{t|t}$, the variance of the estimation of the state in t, with information up to time t and the ratio $\frac{\lambda}{1+\lambda m\overline{\mu}}$. In the steady state, $V_{t|t}$ is given by (67) and there are no differences between both methods since λ and V satisfy the same equation: equation (68) of the appendix for V and (80) for λ.

For stationary factors, the trace of the difference of MSE of prediction between both methods is since $\phi^2 < 1$

$$\begin{aligned}\Delta_{v-f} &= \sum_{i=1}^{m} p_i^2 \left(\lambda - \frac{\lambda^2 m\overline{\mu}}{1+\lambda m\overline{\mu}} + \left(\frac{\lambda^2 m\overline{\mu}}{1+\lambda m\overline{\mu}} - 1\right)\frac{1-\phi^{2h}}{1-\phi^2} - V_{t|t}\phi^{2h}\right) &(81)\\ &= \sum_{i=1}^{m} p_i^2\phi^{2h}\left(\frac{\lambda}{1+\lambda m\overline{\mu}} - V_{t|t}\right) &(82)\end{aligned}$$

where the last equality is obtained by (52). Again in the steady state $V_{t|t}$ is given by (67) and there is no difference between the two methods since λ and V must satisfy the same equation: (68) for V and (52) for λ.

>From both cases, we can conclude that there are no differences, once the steady state has been reached (which usually takes place in a few iterations) between the factor model and the VARMA model.

12 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.

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

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

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

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

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

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

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

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

García-Ferrer, A., Hoyo, J. del, Novales, A. and Young, P. C., 1996, Recursive Identification, Estimation and Forecasting of Nonstationary Economic Time Series with Applications to GNP International Data, in J. Berry, K. Chaloner and J. Geweke (eds.) Bayesian Analysis in Statistics and Econometrics:Essays in Honour of Arnold Zellner, New York: John Wiley, 1996,.15-27.

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

García-Ferrer, A. and P. Poncela, 1999, Forecasting international GNP data through common factor models and other procedures, Mimeo. Universidad Autónoma de Madrid.

Geweke, J., 1977, The dynamic factor analysis of economic time series models, in D.J. Aigner y A.S. Goldberger (eds) Latent variables in: socio-economic models (North Holland, New York).

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

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

Gourieroux, C., A. Monfort, and E. Renault, 1991, A general framework for factor models, Document de Travail 9107, INSEE.

Harvey, A.C., 1989, Forecasting structural time series models and the Kalman filter, (Cambridge University Press, Cambridge).

Lin, J.L. and R. Tsay, 1996, Co-integration constraint and forecasting. An empirical examination, Journal of Applied Econometrics 11, 519-538.

Min, C. and A. Zellner, 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. and G. Box, 1987, Identifying a simplifying structure in time series, Journal of the American Statistical Association 82, 836-843.

Peña, D. and P. Poncela, 1999, Nonstationary dynamic factor models, Universidad Carlos III, Mimeo.

Rao, C.R., 1973, Linear statistical inference and its applications (Wiley, New York).

Reinsel, G.C. and S.K. Ahn, 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. and M.W. Watson, 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097-1107.

Stock, J.H. and M.W. Watson, 1999, Diffusion indexes, Manuscript.

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

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

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

Zellner, A., and C. Hong, 1991, Bayesian methods for forecasting turning points in economic time series: sensitivity of forecasts to asymmetry of loss structures, in: H. Lahiri and G. Moore, eds., Leading economic indicators: new approaches and forecasting records, (Cambridge University Press, Cambridge) 129-140.

Zellner, A., C. Hong, and C. Min, 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.

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

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

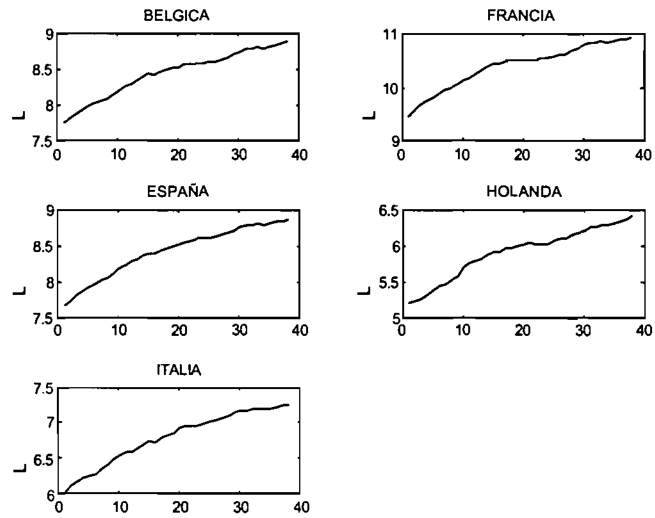

Figure 1: Graph 1: Logs of real GNP of Belgium, France, Spain, Holland and Italy.