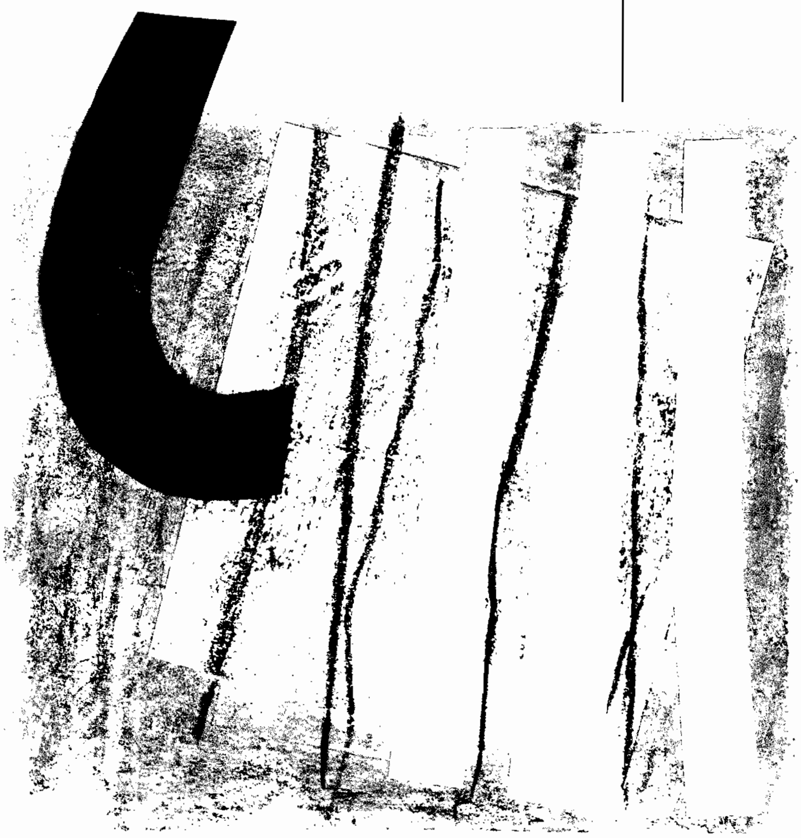
POOLING INFORMATION AND FORECASTING WITH DYNAMIC FACTOR ANALYSIS

Daniel Peña and Pilar Poncela

96-63



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 96-63
Statistics and Econometrics Series 26
November 1996

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

POOLING INFORMATIVON AND FORECASTING WITH DYNAMIC FACTOR ANALYSIS

Daniel Peña and Pilar Poncela

Abstract

In this paper, we present a generalized dynamic factor model for a vector of time series, which seems to provide a general framework to incorporate all the common information included in a collection of variables. The common dynamic structure is explained through a set of common factors, which may be stationary, or nonstationary as in the case of common trends. Also, it may exist a specific structure for each variable. Identification of the nonstationary factors is made through the common eigenstructure of the lagged covariance matrices. Estimation of the model is carried out in state space form with the EM algorithm, where the Kalman filter is used to estimate the factors or not observable variables. It is shown that this approach implies, as particular cases, many pooled forecasting procedures suggested in the literature. In particular, it offers an explanation to the empirical fact that the forecasting performance of a time series vector is improved when the overall mean is incorporated into the forecast equation for each component.

Key Words

Cointegration and common factors; generalized factor model; Kalman filter; pooling techniques; vector time series.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.

1 Introduction

The pooling of information by means of a common mean or "borrowing strenght from the average" appears in a natural way in empirical Bayes methods (e.g. Efron and Morris, 1973; Morris, 1983; Casella, 1983), hierarchical Bayesian models (Berger and Deely, 1988) and shrinkage estimators (James and Stein, 1961 and Green and Strawderman, 1991). In the time series literature, these ideas have been used by García Ferrer *et al* (1987) and Zellner and Hong (1989) among others to show that forecasting of the annual output growth rates of several countries can be improved in terms of an out-of-sample root-mean-squared criteria by introducing a median real stock return of all countries. This is a relevant example of how the forecast of a set of variables can be improved by adding the overall mean of the variables to the univariate ARMA models fitted. Other examples can be found in Clemen (1989), Ledolter and Lee (1993) and Min and Zellner (1993), among others.

This paper has two main contributions. The first one is a generalization of the dynamic factor model studied by Peña and Box (1987) to the nonstationary case. The second one is to show that the forecast generated from this model implies, as particular cases, many pooled forecasting procedures suggested in the literature. It is shown that the forecasts from the factor model incorporate a weighted average of all the components collected in the time series vector, with weights proportional to the inverse of the variances of the error terms of the series.

This article is organized as follows. Section 2 presents the generalized dynamic factor model and study its properties. Section 3 analyzes the problem of separating the nonstationary factors from the stationary ones and shows how this can be carried out by a generalization of a method proposed by Peña and Box (1987) for stationary factors. Also, this section summarizes the relationship between cointegration and factor models. Section 4 briefly reviews the estimation using the EM algorithm. Section 5 develops the forecasting equations and shows how a pooled forecasting procedure is obtained. Section 6 applys the model to four financial series of Spanish interbank interest rates: 1 day, 3 months, 6 months and 1 year. Two factors of different nature are found: the first factor is nonstationary and can be interpreted as a common trend driving all the series; the second factor is stationary and can be interpreted as a factor that differentiates between the short and the long run in the four series. Finally, a specific factor to each of the series is found, which explain the dynamic structure particular to each of the series. It is shown that the factor model provides better forecasts than a vector ARMA model.

2 The Factor Model

In the time domain, dynamic factor models have been studied by Engle and Watson (1981, 1983), Shumway and Stoffer (1982), Peña and Box (1987), Tiao and Tsay (1989), Gonzalo and Granger (1991) and Reinsel and Ahn (1992), among others. Let y_t be an m-dimensional vector of observable time series, generated by a set of not observable factors. We assume that each component of the vector of observed series, y_t, can be written as a linear combination of common and specific factors; that is

$$\underset{m\times 1}{y_t} = \underset{m\times r}{P}\ \underset{r\times 1}{f_t} + \underset{m\times 1}{n_t} \tag{1}$$

where f_t is the r-dimensional vector of **common factors**, P is the factor loading matrix, and n_t is the vector of **specific factors**. Therefore, all the common dynamic structure comes through the common factors, f_t, and the vector n_t explains the dynamics specific to each time series. If there is not any specific dynamic structure, n_t is reduced to white noise.

We suppose that the vector of common factors follows a VARMA(p,q) model

$$\Phi(B)f_t = \Theta(B)a_t, \tag{2}$$

where $\Phi(B) = I - \Phi(1)B-,\cdots,-\Phi(p)B^p$, and $\Theta(B) = I - \Theta(1)B-,\cdots,-\Theta(q)B^q$, are $r\times r$ polinomial matrices and B is the backshift operator. The sequence of vectors a_t are normally distributed, have zero mean and covariance matrix Σ_a, with full rank and are serially uncorrelated, that is $\mathrm{E}\ (a_t a'_{t-h}) = 0 \quad h \neq 0$.

The components of the vector of common factors, f_t, can be either stationary or nonstationary. The specific dynamic structure associated with each of the observable series is included in the vector of specific factors, n_t. Of course, some componentes of n_t can be white noise, while other ones can have dynamic structure and follow an ARMA model. In general,

$$\Phi_n(B)n_t = \Theta_n(B)e_t, \tag{3}$$

with Φ_n and Θ_n $m\times m$ diagonal matrices given by $\Phi_n(B) = I-\Phi_n(1)B-,\cdots,-\Phi_n(p)B^p$, and $\Theta_n(B) = I-\Theta_n(1)B-,\cdots,-\Theta_n(q)B^q$, and therefore each component follows an univariate ARMA(p_i,q_i), $i=1,2,\cdots,m$, being p=max(p_i) and q=max(q_i), $i=1,2,\cdots,m$. The sequence of vectors e_t are normally distributed, have zero mean and diagonal covariance matrix Σ_e. Recall that if all the dynamic structure comes through the common factors, the components of n_t are white noise, and $\Theta_n(B) = I$ and $\Phi_n(B) = I$. We assume that the noises from the two different set of factors, common and specific, are also uncorrelated for all lags,

$$\mathrm{E}\ (a_t e'_{t-h}) = 0 \quad \forall h. \tag{4}$$

The model as stated is not identified, because for any $r \times r$ non singular matrix H the observed series y_t can be expressed in terms of a new set of factors,

$$y_t = P^* f_t^* + n_t \tag{5}$$

$$\Phi^*(B) f_t^* = \Theta(B)^* a_t^* \tag{6}$$

with $P^{*\prime} P^* = (H^{-1})' P' P H^{-1}$, $f_t^* = H f_t$, $a_t^* = H a_t$, $\Phi^*(B) = H \Phi H^{-1}$, $\Theta^*(B) = H \Theta H^{-1}$, and $\Sigma_a^* = H \Sigma_a H'$. Models (1), (2) and (5), (6) are identical from the point of view of the available data.

To solve the identification problem, we follow the work by Hannan (1969, 1971, 1976) and Kohn (1979) which has been more recently extended to nonstationary state space models by Wall (1987), and look for parametrizations that are unique in their effect on first and second moments of the observed time series. The observational equivalence between two parameter structures gives a set of relation equations between the matrices from the two alternative parametrizations of the model, and restrictions should be imposed until the relation between the two structures is given by the identity matrix. In this case, both parametrizations are the same.

As the scale of the factors is irrelevant, the factors noise covariance matrix, Σ_a, may be chosen to be the identity matrix. Then, Σ_a^* will not be the identity unless H is orthogonal, and still the model is not identified to rotations. Other common solutions used to avoid this indeterminancy is to choose P, such that $P'P = I$. Some parameters of the processes followed by the factors may be also restricted by the nature of the processes. (For example, if there is a common trend orthogonal to some stationary factors, the matrix Φ has already some fixed paramenters.) When n_t is white noise and the factors are stationary model (1) and (2) is the factor model studied by Peña and Box (1987).

3 Stationary and nonstationary factors

For dynamic stationary factor models, Peña and Box (1987) developed a method of identyfing the number of common factors based in the common eigenstructure of the lagged covariances matrices of the vector of time series. Nevertheless, in many cases real time series vectors are nonstationary. Suppose that the vector of time series is I(1). In a general case, some common factors will be stationary, while others will be nonstationary. A factor can also be a common trend, in the sense of Stock and Watson (1988), driving all the series.

Suppose that the specific factors, if they exist, are stationary, and that there are some common I(1) factors. To identify this non-stationary common factors, we define the matrix of sample second moments,

$$A_y(k) = \frac{1}{T^2} \sum y_{t-k} y_t' \tag{7}$$

Notice that the sum of second moments is divided by T^2, so that only the submatrix of A_y associated with the I(1) factors will converge to a non zero random matrix. To see this, subtitute y_t, expressed as in (1), in the equation above (7)

$$\begin{aligned} A_y(k) = \frac{1}{T^2} \sum y_{t-k} y_t' &= P(\frac{1}{T^2} \sum f_{t-k} f_t')P' + P(\frac{1}{T^2} \sum f_{t-k} n_t') \\ &+ (\frac{1}{T^2} \sum n_{t-k} f_t')P' + \frac{1}{T^2} \sum n_{t-k} n_t') \end{aligned}$$

Let us call

$$\Gamma_f(k) = \text{ plim } \frac{1}{T^2} \sum f_{t-k} f_t'$$

and

$$\Gamma_{f_1}(k) = \text{ plim } \frac{1}{T^2} \sum f_{1,t-k} f_{1,t}'$$

Since the specific factors are I(0) and the first and second moments of n_t exist and are finite, it is straightforward to show (see appendix), that

$$\frac{1}{T^2} \sum n_{t-k} n_t' \xrightarrow{p} 0$$

For the common stationary factors, $f_{2,t}$, stationarity and the existence of finite first and second moments imply

$$\frac{1}{T^2} \sum f_{2,t-k} f_{2,t}' \xrightarrow{p} 0$$

From equation (4), both noise processes, the one associated with the common factors, a_t, and the one associated with the specific ones, e_t, are uncorrelated for all lags. In Appendix 1, it is also shown that

$$\frac{1}{T^2} \sum f_{t-k} n_t' \xrightarrow{p} 0$$

so,

$$\Gamma_y(k) = \text{ plim } \frac{1}{T^2}\sum y_{t-k}y_t' = P(\text{ plim } \frac{1}{T^2}\sum f_{t-k}f_t')P' = P\Gamma_f(k)P' \tag{8}$$

or more explicitly

$$\Gamma_y(k) = [P_1 P_2]\begin{bmatrix} \Gamma_{f_1}(k) & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} P_1' \\ P_2' \end{bmatrix} = P_1\Gamma_{f_1}(k)P_1' \tag{9}$$

Following the discussion in Peña and Box (1987) but applying it now to the nonstationary factor model, the random matrix, $\Gamma_y(k)$ has as non zero eigenvalues the diagonal term of $\Gamma_{f_1}(k)$ and as eigenvectors corresponding to these eigenvalues P_1. So the number of nonstationary common factors is the number of non-zero eigenvalues of $\Gamma_y(k)$, equals the number of eigenvalues of, $\Gamma_{f_1}(k)$

Common factors are very related to cointegration relationships. See Stock and Watson (1988), Johansen (1991), Gonzalo and Granger (1991), Reinsel and Ahn (1992) and Escribano and Peña (1994), among others. Let y_t be a vector of m variables, $y_{1,t}, y_{2,t}, \cdots, y_{m,t}$, that are I(1). It is said that the m I(1) variables are cointegrated of order 1 and rank $l, l < m$, if there are l linearly independent combinations of them, say $z_{1,t}, z_{2,t}, \cdots, z_{r_1,t}$, that are I(0) or in a short way, each component of y_t is I(1), but Ay_t are I(0), where $A = [a_{ij}]$ and rank$(A) = l < m$. The existence of cointegration relations in a set of time series variables is directly related to the existence of common I(1) factors as it can be seen in Escribano and Peña (1994), from where we take the following result. The vector of m time series y_t is cointegrated of order 1 and rank $l < m$ if, and only if, y_t has $m - l$ common factors that are I(1).

In our case the number of nonstationary factors or common trends is r_1. So, the rank of cointegration of the m variables collected in y_t is $m - r_1$. In other words, we can find r_1 common trends because there are $m - r_1$ cointegration relations. Remember that the number r_1 of common trends or non stationary factors is the number of non-zero eigenvalues of $\Gamma_y(k)$, (see 8) which can empirically estimated as the number of non zero eigenvalues of $A_y(k)$, (see 7) for T large.

4 Estimation

Model estimation is carried out by writing the model in state space form and using the EM algorithm. In the time series context, this was first done by Shumway an Stoffer (1982) while Stock and Watson (1983) suggested to use the EM algorithm with a final pass through the scoring to calculate an estimator of the information matrix. In this section, we follow their work, which we briefly review.

The model can be written in state space form as follows: the vector of observable time series y_t, is given by the **measurement equation**,

$$\underset{m\times 1}{y_t} = \underset{m\times s}{\tilde{P}} \underset{s\times 1}{z_t} + \underset{m\times 1}{\epsilon_t} \tag{10}$$

with $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{E}(\epsilon_t\epsilon_t') = \Sigma_\epsilon$ and $\mathrm{E}(\epsilon_t\epsilon_\tau') = 0$ if $t \neq \tau$. The vector of factors z_t is driven by the **transition equation**,

$$\underset{s\times 1}{z_t} = \underset{s\times s}{G} \underset{s\times 1}{z_{t-1}} + \underset{s\times 1}{u_t} \tag{11}$$

with $\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_tu_t') = \Sigma_u$ and $\mathrm{E}(u_tu_\tau') = 0$ if $t \neq \tau$. Both noises, ϵ_t and u_t, are also uncorrelated for all lags, $\mathrm{E}(\epsilon_tu_\tau') = 0$ for all t and τ. To write an ARMA(p,q) model in state space form, a state vector of dimension $\max(p,q+1)$ (e.g. Akaike, 1974; Gardner *et al* 1980; Ansley and Kohn, 1983) gives a minimal representation with uncorrelated errors in the transition and measurement equations. In this case, the ARMA model is a common factor, not observable, and the state vector, z_t, has to be increased to consider also the common trends and the specific factors. Its dimension is $r_1 + R + R^*$, where r_1 is the number of common trends, $R = \max(p,q+1)$ where (p,q) is the order of the VARMA proccess followed by the common factors and R^* is refered to the number of specific factors different from white noise and it is equal to $\sum_{i=1}^{m}\max(p_i,q_i+1)$ where (p_i,q_i) are the orders of the ARMA proccesses followed by the specific factors.

Once the model is written in state space form, estimation will be carried out by the EM algorithm (Dempster *et al*, 1977). In this case, two different set of unkowns should be distinguished: the parameters of the model (P, G, Σ_ϵ and Σ_u), usually known as hyperparameters and from now on denotated by α, and the state variables (z_t). The problem to be solved is to maximize the density of the observed data $Y = (y_1, y_2, \cdots, y_T)$, which is highly non linear function of these parameters. This problem is equivalent to consider the "complete data set" of the observed time series Y and the not observed state vectors $Z = (z_1, z_2, ..., z_T)$ and maximize the expectation of the joint density of all data, conditioned on the observed although incomplete data, evaluated at the estimation of the unknown parameters $\alpha^{(k)}$ available. This second problem is much easier to solve. Therefore $\hat{\alpha} = \arg\max\log f_Y(Y;\alpha)$ can be found as $\hat{\alpha} = \arg\max E_{Z|Y}\log f_{Y,Z}(Y,Z;\alpha)$. From (10) and (11)

$$\log f_{Y,Z}(Y,Z;\alpha) = \sum_{t=1}^{T}(\log f(y_t|z_t;\alpha) + \log f(z_t|z_{t-1};\alpha)) + \log f(z_0;\alpha) \tag{12}$$

When some componentes of the state vector are nonstationary there are several ways to handle this situation as (i) introducing proper prior information, (ii) estimating z_0 as a nuisance parameter, (iii) setting $\Sigma_0^{-1} = 0$, so the last term in the equation above just disappear. Using this last approach and taking expectations in (12) with respect to the distribution of $f_{Z|Y}$,

$$
\begin{aligned}
E_{Z|Y} \log f_{Y,Z}(Y,Z;\alpha) &= -\frac{Tm}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma_\epsilon| - \frac{1}{2}E_{Z|Y}\left[\sum_{t=1}^{T}(y_t - \tilde{P}z_t)'\Sigma_\epsilon^{-1}(y_t - \tilde{P}z_t)\right] \\
&- \frac{Ts}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma_u| \\
&- \frac{1}{2}\sum_{t=1}^{T}E_{Z|Y}\left[(z_t - Gz_{t-1})'\Sigma_u^{-1}(z_t - Gz_{t-1})\right] \qquad (13)
\end{aligned}
$$

This last expression is then maximized to estimate α and gives us the following estimators for the hyperparameters, $\hat{G} = [\sum E_{Z|Y}(z_{t-1}z_{t-1}')]^{-1}[\sum E_{Z|Y}(z_{t-1}z_t')]$, $\hat{\tilde{P}} = [\sum E_{Z|Y}(z_t z_t')]^{-1}[\sum E_{Z|Y}(z_t y_t')]$, $\hat{\Sigma}_\epsilon = 1/T\sum \hat{\epsilon}_t\hat{\epsilon}_t'$ with $\hat{\epsilon}_t = y_t - \hat{\tilde{P}}z_t$.

Each iteration of the EM algorithm takes two steps:(i) E or expectation step where the moment matrices involved in the estimation of the hyperparameters are calculated with $\alpha = \alpha^{(k)}$, where $\alpha^{(k)}$ denotes the parameter vector obtained at iteration k. These moment matrices are sufficient statistics of the parameters to be estimated. (ii) M or maximization step, where you obtain the unknown parameters of the model through the maximization of the above function,

$$\alpha^{(k+1)} = \arg\max E_{Z|Y}(\log L(YZ;\alpha|\alpha^{(k)})), \alpha \in \mathcal{P}$$

The procedure is implemented as follows:

1. Set up the factor model and the initial conditions for the model parameters, and for the state variable z_1 and its covariance var(z_1). This is done by (i) Set the number of I(1) factors as the number of nonzero eigenvalues of $\Gamma_y(1)$. (ii) Set P as the r first eigenvectors of $\Gamma_y(1)$, $\tilde{P}$ is a known function of P. (iii) Set G by writting the ARMA model as the transition equation of the state space model. (iv) Set $z_1 = \tilde{P}^- y_1 = (\tilde{P}'\tilde{P})^{-1}\tilde{P}'y_1$. (v) Set $\Sigma_\epsilon = I$ or any diagonal matrix.

2. Run the Kalman filter to estimate the state with the information available until time t, $z_{t|t} = E(z_t|y_t, ..., y_1)$. The state vector and its covariance matrix can be estimated through the well-known Kalman filter **forecasting equations**:

$$z_{t|t-1} = Gz_{t-1|t-1}, \qquad (14)$$

with associated covariance matrix,

$$V_{t|t-1} = GV_{t-1|t-1}G' + \Sigma_u, \qquad (15)$$

and also

$$y_{t|t-1} = \tilde{P} z_{t-1|t-1} \tag{16}$$

with covariance matrix given by,

$$\Sigma_{t|t-1} = \tilde{P} V_{t|t-1} \tilde{P}' + \Sigma_\epsilon. \tag{17}$$

Once a new observation of the time series is available the forecast for the state vector and its covariance matrix can be actualized trough the **updating equations** of the Kalman filter:

$$z_{t|t} = (I - K_t \tilde{P}) z_{t|t-1} + K_t y_t \tag{18}$$

$$V_{t|t} = V_{t|t-1} - V_{t|t-1} \tilde{P}' \Sigma_{t|t-1}^{-1} \tilde{P} V_{t|t-1} \tag{19}$$

where K_t is the *filter gain*, given by

$$K_t = V_{t|t-1} \tilde{P}' \Sigma_{t|t-1}^{-1} \tag{20}$$

3. The E step requires the computation of $z_{t|T} = E(z_t | y_T, ..., y_1)$. Any smoothing algorithm can be used at this point. The most widely used, and the one used here, is the *fixed interval smoother*, see Harvey (1989, p.154-155).

4. The maximization step of the EM algorithm gives a new estimation of the parameters of the model, that is we find $\hat{\alpha} = (\hat{P}, \hat{G}, \hat{\Sigma}_\epsilon, \hat{\Sigma}_u)$ such that maximizes the log likelihood function $\log f(\alpha|Y)$. This maximization is done in two steps. At iteration k, the estimation of the covariance matrices found in iteration $k-1$ is used to calculate the system matrices $\hat{G}$ and $\hat{\tilde{P}}$. These system matrices just estimated are then used to update the covariance matrix, Σ_ϵ. Recently, several algorithms have been developed, as the ECM algorithm (Meng and Rubin, 1993) or the ECME algorithm (Liu and Rubin, 1994) where the maximization step is replaced by several conditional maximizations.

5. And finally, repeat 2, 3, and 4 until convergence.

The equations of the EM algorithm allow the maximization of the log likelihood to obtain a new estimation of the parameters of the model, in a simpler way than the usual scoring algorithm. The procedure is repeated until convergence. It can be proved that under proper conditions at each iteration the value of the log likelihood never decreases and the algorithm converges to a stationary point: a local maximum, a global maximum or a saddle point. Therefore, in order to be sure that we reach the desired maximun one should try different set of initial values. The main characteristic of this algorithm is that it does not require at each iteraion the inverse of the information matrix as in the score algorithm. Another advantage is that it gives, by definition, positive definite estimation of the covariance matrices. Its main drawback is that it does not give standard errors of the parameter estimates. There are several ways to obtain them as the SEM algorithm (Meng and Rubin, 1991) or the SECM algorithm (van Dyk *et al*, 1995) if we are using conditional maximization at the M-step. Also one can calculate the information matrix, and this is the approach followed here. As the convergence of the algorithm is only linear, in some applications a large number of iterations are needed until the stationary point is reached, once we are in a neighbourhood of it, and lately several methods have been developed to accelerate the convergence rate. Nevertheless, for the factor model, convergence is obtained in few iterations. So, the standard EM algorithm was used.

5 Forecasting and pooling techniques

In this section it is shown that this approach implies as particular cases some of the pooled forecasting procedures suggested in the literature. This also will give a better insight of how common factors affect prediction. Forecast is made applying equations (14) through (17) to time period $t+h, h=1,2,\cdots$.To build the h-steps ahead forecast, first equation (14) is used h times to estimate $z_{t+h|t}$

$$z_{t+h|t} = G'^{h} z_{t|t}, \tag{21}$$

with associated covariance matrix obtained applying (15)

$$V_{t+h|t} = G^{h} V_{t|t} G'^{h} + \sum_{j=0}^{s-1} G^{j} \Sigma_u G'^{j-1}. \tag{22}$$

Then, the h-steps ahead forecast for the vector of observable time series is calculated applying equation (16) to $z_{t+h|t}$ and from equation (21)

$$\hat{y}_{t+h} = \tilde{P} z_{t+h|t} = \tilde{P} G'^{h} z_{t|t}, \tag{23}$$

with covariance matrix given by (17)

$$\Sigma_{t+h|t} = \tilde{P} V_{t+h|t} \tilde{P}' + \Sigma_\epsilon \tag{24}$$

Since $z_{t|t}$ is a linear combination of $z_{t|t-1}$ and y_t given by equation (18),

$$\hat{y}_{t+h} = A_1 z_{t|t-1} + \tilde{P} G'^{h} K_t y_t \tag{25}$$

with $A_1 = \tilde{P}G^h(I - K_t\tilde{P})$. The prediction of $\hat{y}_{t+h}$ is also a linear combination of the forecast of the state at time t with the information given at time $t-1$ and the vector of observations at time t. To find out how the information carried in y_t is incorporated into the forecast of $\hat{y}_{t+h}$, as K_t, the filter gain, is given in (20), so

$$\tilde{P}G^h K_t y_t = \tilde{P}G^h V_{t|t-1}\tilde{P}\Sigma_{t|t-1}^{-1} y_t \tag{26}$$

where $V_{t|t-1}$ and $\Sigma_{t|t-1}$ are given by (15) and (17) respectively. Applying to the inverse of (17), the well-known formula for the inverse of a sum of matrices $(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$ for A and C nonsingular matrices,

$$\Sigma_{t|t-1}^{-1} = \Sigma_\epsilon^{-1} - \Sigma_\epsilon^{-1}\tilde{P}(\tilde{P}'\Sigma_\epsilon^{-1}\tilde{P} + V_{t|t-1}^{-1})^{-1}\tilde{P}'\Sigma_\epsilon^{-1} \tag{27}$$

Replacing the expression above in equation (26),

$$\tilde{P}G^h K_t y_t \quad = \quad \tilde{P}A_2\tilde{P}'\Sigma_\epsilon^{-1} y_t = D\Sigma_\epsilon^{-1} y_t \tag{28}$$

with $D = \tilde{P}A_2\tilde{P}'$ and $A_2 = G^h V_{t|t-1}(I - \tilde{P}'\Sigma_\epsilon^{-1}\tilde{P}(\tilde{P}'\Sigma_\epsilon^{-1}\tilde{P} + V_{t|t-1}^{-1})^{-1})$, an $s \times s$ and $m \times m$ matrices. If the filter reaches an steady state, $V_{t|t-1}$ and A_2 can be considered as time invariant matrices respectively. The h-steps ahead forecast of the time series vector $\hat{y}_{t+h}$ is given then by the following linear combination

$$\hat{y}_{t+h} = A_1 z_{t|t-1} + D\Sigma_\epsilon^{-1} y_t. \tag{29}$$

and as Σ_ϵ^{-1} is a diaginal matrix, the j-th component of the vector of time series predicted can be written as

$$\hat{y}_{j,t+h} \quad = \quad (A_1 z_{t|t-1})_j + (\frac{d_{j1}}{\sigma_1^2} y_{1,t} + \frac{d_{j2}}{\sigma_2^2} y_{2,t} + \cdots + \frac{d_{jm}}{\sigma_m^2} y_{m,t}). \tag{30}$$

where $(Y)_j$ represents the j-th component of vector Y and d_{ji} is the (j,i) element of matrix D. This equation shows that the forecast for each component of the vector y_t incorporates a pooling term which is a weighted sum of all the individual series with weights inversely proportional to the noise variance of each serie.

5.1 The single factor model

A single factor model can be of special interest. From the macroeconomic point of view it can be seen as an unobserved variable describing the state of the economy, approach related to the theory of the bussiness cycle. In many research areas it can represent the situation where there are several measures available of the same unobserved dynamic variable. Besides this practical interest, from the theoretical point of view it is also interesting since it clarifies the nature (permanent or transitory) of the pooling term of the forecasting equation of the vector of time series. For special cases it has deeper implications as it will be pointed out.

When the vector of time series is generated by one common factor the model is

$$\underset{m\times 1}{y_t} = \underset{m\times 1}{P}\ \underset{1\times 1}{f_t} + \underset{m\times 1}{\epsilon_t} \tag{31}$$

with $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{E}(\epsilon_t\epsilon_t') = \Sigma_\epsilon$, Σ_ϵ diagonal and $\mathrm{E}(\epsilon_t\epsilon_\tau') = 0$ if $\tau \neq t$. First, it will be analyzed the case where the common factor is given by

$$\underset{1\times 1}{f_t} = \underset{1\times 1}{\phi}\ \underset{1\times 1}{f_{t-1}} + \underset{1\times 1}{a_t} \tag{32}$$

with $\mathrm{E}(a_t) = 0$, $\mathrm{var}(a_t) = \sigma_a^2$, $\mathrm{cov}(a_t a_\tau) = 0$ if $\tau \neq t$ and $|\phi| \leq 1$. Notice that this specification implies AR(1) stationary factors when $|\phi| < 1$, as well as common trends for $\phi = 1$. The model is already in state space form with $\tilde{P} = P$, $z_t = f_t$, $r = s = 1$, $u_t = a_t$ and $G = \phi$.

For a general factor loading matrix $\tilde{P} = (p_1, p_2, \cdots, p_m)'$, the pooling term implied in forecasting equation (29) $D\Sigma_\epsilon^{-1}y_t = \tilde{P}A_2\tilde{P}'\Sigma_\epsilon^{-1}y_t$ and A_2 is now an scalar given by $A_2 = \phi^h\sigma^2_{f,t|t-1}(1 - c_t)$ with $c_t = \sum_{i=1}^m \frac{p_i^2}{\sigma_i^2}(\sum_{i=1}^m \frac{p_i^2}{\sigma_i^2} + \frac{1}{\sigma^2_{f,t|t-1}})^{-1}$ and $\sigma^2_{f,t|t-1}$ the variance of the single factor at time t with the information given at time $t-1$. The forecast of $\hat{y}_{t+h}$ is, from (29)

$$\hat{y}_{t+h} = A_1 z_{t|t-1} + \phi^h\sigma^2_{f,t|t-1}(1 - c_t)\tilde{P}(\frac{p_1}{\sigma_1^2}, \frac{p_2}{\sigma_2^2}, \cdots, \frac{p_m}{\sigma_m^2})y_t. \tag{33}$$

The j-th component of the vector of time series is predicted as

$$\hat{y}_{j,t+h} = (A_1 z_{t|t-1})_j + \phi^h\sigma^2_{t|t-1}(1 - c_t)(\frac{p_1p_j}{\sigma_1^2}y_{1,t} + \frac{p_2p_j}{\sigma_2^2}y_{2,t} + \cdots + \frac{p_mp_j}{\sigma_m^2}y_{m,t}). \tag{34}$$

This equation shows that the forecast for each component of the vector y_t incorporates a pooling term which is a weighted sum of all the individual series, with weights inversely proportional to the noise variance of each serie and directly proportional to the product of the factor loading of the serie with the remaining factor loadings. For a stationary common factor $|\phi| < 1$ and $\phi^h \to 0$ when h gets larger. So this pooling term exponentially decay towards zero, which means that for long term predicitons this pooling effect disappears or has a transitory nature. For a common trend $\phi^h = 1$ and the nature of the pooling term is permanent, as it should be expected.

Another interesting conclusion can be drawn when the common factor, stationary or not, affects to all the series in a similar way. In this case the loading matrix $\tilde{P}$ can be an $m \times 1$ vector of ones, $\mathbf{1} = (1, 1, \cdots, 1)'$. The forecast of the vector $\hat{y}_{t+h}$ is just inversely proportional to the noise variance of each of the series.

$$\hat{y}_{t+h} = A_1 z_{t|t-1} + \phi^h \sigma^2_{f,t|t-1}(1 - c_t)\mathbf{1}(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \cdots, \frac{1}{\sigma_m^2})y_t \tag{35}$$

with $c_t = \sum_{i=1}^m \frac{1}{\sigma_i^2}(\sum_{i=1}^m \frac{1}{\sigma_i^2} + \frac{1}{\sigma^2_{f,t|t-1}})^{-1}$ and for the j-th component of the vector of time series,

$$\hat{y}_{j,t+h} = (Az_{t|t-1})_j + \phi^h \sigma^2_{f,t|t-1}(1 - c_t)(\frac{1}{\sigma_1^2}y_{1,t} + \frac{1}{\sigma_2^2}y_{2,t} + \cdots + \frac{1}{\sigma_m^2}y_{m,t}) \tag{36}$$

Of course for the nonstationary case $\phi^h = 1$ while for the stationary one the pooling term disappears in the long run.

Obviously, if the series have a similar variability the optimal forecast from this model is obtained by shrinking each serie towards the common mean. That is, if $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2 = \sigma^2$, then

$$\hat{y}_{j,t+h} = (Az_{t|t-1})_j + \phi^h \frac{\sigma^2_{f,t|t-1}}{\sigma^2}(1 - c_t)(y_{1,t} + y_{2,t} + \cdots + y_{m,t}) \tag{37}$$

This can be an explanation to the fact that the incorporation of the mean of a vector of time series improves the forecasting performance of a model as it was empirically found by García Ferrer *et al* (1987) and Zellner and Hong (1989).

For a MA(q) proccess; the dimension of the state vector is $s = R = q + 1$ and the transition matrix is Toeplitz of the form

$$G = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

It is straightforward to show that $G^h = 0$ for $h > q$, for any q positive integer. So the pooling term is transitory and has a cut-off for $h > q$.

An AR(p) proccess can be written as an infinite MA proccess so in the infinite horizont of prediction the pooling term disappear. From the practical point of view, for h large enough it can be considered that it is vanished. For smaller values of h, the forecast for each component of the vector y_t incorporates a pooling term which is a weighted sum of all the individual series with weights inversely proportional to the noise variance of each serie. The consecuences for an ARMA(p, q) model can be derived from the above results for AR and MA proccesses.

6 An application to Spanish Interbank Interest Rate

The data consist of 164 observations, from June 84 until January 96, of four time series of Spanish interbank interest rates: one day, r_1, three months, r_{90}, six months, r_{180}, and one year, r_{365}. Figure 1 shows a graph of the series.

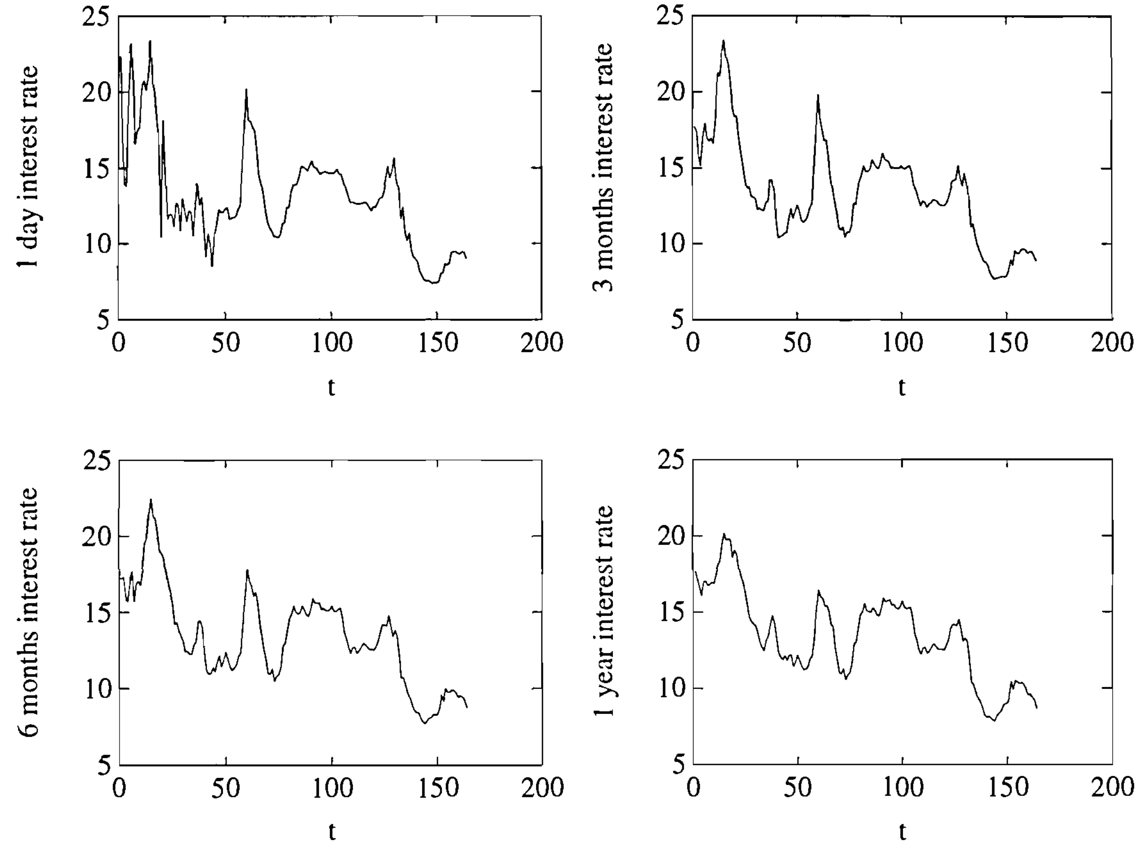


Figure 1 Graphs of the four series of Spanish Interbank Interest Rate

The sample was divided in two periods. The first includes 134 observations (from June 84 until June 93) and was used for estimation, the second include 30 (from July 93 until January 96) points and was reserved to evaluate and compare the models in terms of forecasting. Predictions 1,2,..., 30 steps ahead were calculated for the four series, and their mean square error evaluated. First, an VMA(1) model was fitted to the differenced series. A log transformation was tried first, but it gave worst results in forecasting than the originals series because of the smoothness of the 1 year interest rate serie. The VARMA model fitted to the series is given in Table 1.

Table 1: VMA(1) model fitted to ∇y_t, where $y_t' = (r_1, r_{90}, r_{180}, r_{365})$

$$\theta = \begin{bmatrix} .34 & -.67 & .47 & -.96 \\ -.09 & 0 & 0 & -.46 \\ 0 & -.62 & .88 & -.73 \\ 0 & 0 & -.18 & .23 \end{bmatrix} \Sigma_\epsilon = \begin{bmatrix} 2.16 & .63 & .39 & .20 \\ .63 & .56 & .46 & .31 \\ .39 & .46 & .42 & .29 \\ .20 & .31 & .29 & .25 \end{bmatrix}$$

Estimation was carried out with SCA software. Notice the large variance of the first serie, as can be seen in Figure 1.

The first thing required for fitting a factor model is to find the number of common factors. This is done through the eigenstructure of the second moment matrices of the series. Table 2 shows the largest and second largest eigenvalues of $A_y(k) = \frac{1}{T^2} \sum \tilde{y}_t \tilde{y}_{t-k}'$ where $\tilde{y}_t = y_t - c$, and their corresponding eigenvectors.

Table 2: First and second eigenvalue and eigenvector $A_y(k)$

lag	1	2	3	4	5
10λ	2.37	2.31	2.11	1.98	1.80
	.52	.52	.52	.52	.52
	.53	.53	.53	.52	.52
	.50	.50	.50	.49	.49
	.45	.45	.45	.45	.46

lag	1	2	3	4	5
10λ	.072	.055	.053	.051	.046
	.79	.79	.82	.85	.87
	-.02	-.01	-.04	-.12	-.16
	-.29	-.28	-.31	-.34	-.34
	-.53	-.54	-.48	-.38	-.30

The stability of the eigenstructure through the lagged covariance matrices suggested two common factors. The first factor is a common trend since there is one "large" stable eigenvalue in the lagged covariance matrices. Notice the stability of the eigenvector associated with this largest eigenvalue. The second common factor is a sationary one. An specific factor to each of the interbank interest rate series was also found. The final estimated factor model is given in Table 3.

Table 3: Matrices of the factor model, $\Sigma_a = I$

$$P = \begin{bmatrix} .64 & .49 \\ .66 & .27 \\ .60 & .08 \\ .51 & -.14 \end{bmatrix} \phi = \begin{bmatrix} 1 & 0 \\ 0 & .92 \end{bmatrix} \phi_n = \begin{bmatrix} .25 & 0 & 0 & 0 \\ 0 & .41 & 0 & 0 \\ 0 & 0 & .50 & 0 \\ 0 & 0 & 0 & .46 \end{bmatrix} \Sigma_e = \begin{bmatrix} 1.15 & 0 & 0 & 0 \\ 0 & .677 & 0 & 0 \\ 0 & 0 & .577 & 0 \\ 0 & 0 & 0 & .491 \end{bmatrix}$$

The elements of the first column of the loading matrix are all positive and more or less of similar magnitude. Then the first factor can be interpreted as a mean of all the series or a common trend driving all the series. In the second column of matrix P, the first and second elements are positive, the third is close to zero and the fourth one is negative. The magnitude of the first element (.49) is almost twice the one of the second coefficient (.27). This second factor can be interpreted as opposing the short and the long run in the series.

Once the models were estimated, we made forecast 1, 2, ..., 30 steps ahead and calculate the mean square error of the first 10, 20 and 30 forecasts. Results for 10 periods ahead were a little better for the VARMA model. Results for 20 and 30 steps ahead are given below, in Table 4.

Table 4: Forecast comparison between the VARMA and the factor model

	VMA(1) model			Factor model	
	20	30		20	30
r_1	1.40	14.46	r_1	1.91	1.45
r_{90}	1.95	8.05	r_{90}	1.22	.93
r_{180}	2.52	4.51	r_{180}	1.17	.85
r_{365}	3.40	2.21	r_{365}	1.35	1.06

Notice that although the VARMA model is able to forecast in the medium run, in the long run it fails. The generalized factor model seems to be able to capture the long run dynamics; that is the reason of its better performance.

Appendix

In this appendix, it is shown that for n_t, f_t defined as in section 2,
(a)

$$\frac{1}{T^2}\sum n_{t-k}n_t' \xrightarrow{p} 0$$

and
(b)

$$\frac{1}{T^2}\sum f_{t-k}n_t' =\xrightarrow{p} 0$$

(a) Let n_t be an $m \times 1$ vector of specific stationary factors, then

$$\frac{1}{T^2}\sum n_{t-k}n_t' \xrightarrow{p} 0$$

This is inmediatly, since under the stationary assumption and the assumptions made on the errors, (p. 3),

$$\frac{1}{T}\sum n_{t-k}n_t' \xrightarrow{p} E(n_{t-k}n_t')$$

Since $E(n_{t-k}n_t')$ exists and is finite,

$$\frac{1}{T^2}\sum n_{t-k}n_t' \xrightarrow{p} 0$$

(b) Let f_t be an $r \times 1$ vector of common factors and suppose that r_1 of them are common trends, while r_2 are stationary, $r = r_1 + r_2$, then

$$\frac{1}{T^2}\sum f_{t-k}n_t' = \frac{1}{T^2}\sum \begin{bmatrix} f_{1,t-k}n_t' \\ f_{2,t-k}n_t' \end{bmatrix}$$

(b1) First, it will be shown that for the stationary common factors,

$$\frac{1}{T^2}\sum f_{2,t-k}n_t' \xrightarrow{p} 0$$

that is, we will prove that

$$\lim_{T\to\infty} P[\|\frac{1}{T^2}\sum f_{2,t-k}n_t'\| > \epsilon] = 0$$

Let $f_{2,t-k} = \sum_l A_l a_{2,t-k-l}$ and $n_t = \sum_i C_i e_{t-i}$. Then for $r > 0$ and by the Markov inequality, $\forall\delta > 0$,

$$P[\|\frac{1}{T^2}\sum f_{2,t-k}n_t'\| > \delta] \leq \frac{E(\|\frac{1}{T^2}\sum f_{2,t-k}n_t'\|)}{\delta}$$

and

$$\begin{aligned}
E(\|\frac{1}{T^2}\sum f_{2,t-k}n_t'\|) &= E(\|\frac{1}{T^2}\sum_t\sum_l A_l a_{2,t-k-l}\sum_i e_{t-i}'C_i'\|) \\
&= E(\|\frac{1}{T^2}\sum_t\sum_l\sum_i A_l a_{2,t-k-l}e_{t-i}'C_i'\|) \\
&\leq \|\frac{1}{T^2}\sum_t\sum_l\sum_i A_l E(a_{2,t-k-l}e_{t-i}')C_i'\| = 0
\end{aligned}$$

since by hypothesis the noise sequences a_t and e_τ are uncorrelated for all t and τ. So,

$$P[\|\frac{1}{T^2}\sum f_{2,t-k}n_t'\| > \delta] \leq \frac{E(\|\frac{1}{T^2}\sum f_{2,t-k}n_t'\|)}{\delta} \leq 0.$$

(b2) Now, for the term associated with the nonstationary common factors,

$$\frac{1}{T^2}\sum f_{1,t-k}n_t' \xrightarrow{p} 0$$

Let $f_{1,t-k} = \sum_{l=1}^{t-k} w_l + f_0$, where w_s is a zero mean stationary proccess and f_0 is finite, for example $f_0 = 0$, so substituing $f_{1,t-k}$ for the former expression

$$\frac{1}{T^2}\sum f_{1,t-k}n_t' \;=\; \frac{1}{T^2}\sum_{t=1}^{T-k}\sum_{l=1}^{t-k} w_l n_t' + \frac{1}{T^2}\sum_{t=1}^{T-k} f_0 n_t'.$$

The last term of the right hand side converges to zero in probability since it can be written as

$$\frac{1}{T^2}\sum f_0 n_t' = \frac{f_0}{T}\frac{1}{T}\sum n_t' \xrightarrow{p} 0$$

and the second part of the right hand side goes to the expectation of n_t as T gets larger, finite, and the first part goes to zero, so the product goes to zero. Now the first term can be written as

$$\frac{1}{T^2}\sum_{t=1}^{T-k}\sum_{l=1}^{t-k} w_l n_t' = \frac{1}{T^2}\sum_{t=1}^{T-k} w_t n_t' + \frac{1}{T^2}\sum_{t=2}^{T-k}\sum_{l=1}^{t-1} w_l n_t. \tag{38}$$

Since $v_t = w_t n_t'$ has finite first ($\mathrm{E}(v_t) = 0$) and second moments,

$$\frac{1}{T^2}\sum_{t=1}^{T-k} w_t n_t' \xrightarrow{p} 0.$$

Applying now the Markov inequality to the second term of (38)

$$P[\|\frac{1}{T^2}\sum_{t=2}^{T-k}\sum_{l=1}^{t-1} w_l n_t'\| > \delta] \le \frac{E(\|\frac{1}{T^2}\sum_{t=2}^{T-k}\sum_{l=1}^{t-1} w_l n_t'\|)}{\delta}$$

for $\delta > 0$. And applying the law of iterated expectations

$$E_l E_{t|l}(\|\frac{1}{T^2}\sum_{t=2}^{T-k}\sum_{l=1}^{t-1} w_l n_t'\|) = E_l E_{t|l}(\|\frac{1}{T^2}\sum_{l=1}^{T-k-1}\sum_{t=l+1}^{T-k} w_l n_t'\|) \le 0,$$

since n_t is a zero mean stationary proccess and

$$\frac{1}{T^2}\sum f_{1,t-k}n_t' \xrightarrow{p} 0$$

From (b1) and (b2),

$$\frac{1}{T^2}\sum f_{t-k}n_t' = \frac{1}{T^2}\sum \begin{bmatrix} f_{1,t-k} \\ f_{2,t-k} \end{bmatrix} n_t' = \frac{1}{T^2}\sum \begin{bmatrix} f_{1,t-k}n_t' \\ f_{2,t-k}n_t' \end{bmatrix} \xrightarrow{p} 0$$

References

AKAIKE, H. (1974) Markovian representation of stochastic proccesses and its application to the analysis of autoregressive moving average proccesses. *Annals of Institute of Statistical Mathematics* 26, 363-87.

ANSLEY, C. F. and KOHN, R. (1983) Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. *Biometrika* 70, 275-8.

BERGER, J. O. and DEELY, J. (1988) A Bayesian approach to ranking and selection of related means with alternatives to Analysis-of-Variance methodology. *Journal of the American Statistical Association* 83, 364-73.

CLEMEN, R. T. (1989) Combining forecasts: a review with annotated bibliography. *Journal of Forecasting* 5, 559-84.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* B 39, 1-38.

VAN DYK, D. A., MENG, X. L. and RUBIN, D. B. (1995) Maximun likelihood estimation via the ECM algorithm: computing the asymptotic variance. *Statistica Sinica* 5, 55-75.

EFRON, B. and MORRIS, C. N. (1973) Stein's estimation rule and its competitors–An parametrical Bayes approach. *Journal of the American Statistical Association* 68, 117-30.

ENGLE, R. F. and WATSON, M.W. (1981) A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association* 76, 774-81.

ESCRIBANO, A. and PEÑA, D. (1994) Cointegration and common factors. *Journal of Time Series Analysis* 15, 577-86.

GARCIA-FERRER, A., HIGHFIELD, R. A., PALM, F. and ZELLNER, A. (1986) Macroeconomic forecasting using pooled international data. *Journal of Business and Economic Statistics* 5, 53-67.

GARDNER, G., HARVEY, A. C. and PHILLIPS. G. D. A. (1980) An algorithm for exact maximun likelihood estimation of autogressive-moving average models by means of Kalman filtering. *Applied Statistics* 29, 311-22.

GONZALO, J. and GRANGER, C. W. J. (1995) Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics* 13, 27-36.

GREEN, E. J. and STRAWDERMAN, W. E. (1991) A James-Stein type estimator for combining unbiased and possibly biased estimators. *Journal of the American Sta-*

tistical Association 86, 1001-6.

HANNAN, E. J. (1969) The identification of vector mixed autoregressive- moving average systems. *Biometrika* 56, 223-5.

HANNAN, E. J. (1971) The identification problem for multiple equation systems with moving average errors. *Econometrica* 39, 751-65.

HANNAN, E. J. (1976) The identification and paramatrization of ARMAX and state space forms. *Econometrica* 44, 713-23.

HARVEY, A. (1989) *Forecasting Structural Time Series Models and the Kalman Filter* (2nd edn). Cambridge: Cambridge University Press.

JAMES, W. and STEIN, C. (1961) Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (vol 1). Berkeley, CA: University of California Press, 361-80.

KOHN, R. (1978) Local and global identification and strong consistency in time series models. *Journal of Econometrics* 8, 269-73.

LEDOLTER, J. and LEE, C. (1993) Analysis of many short time series sequences: forecast improvements achieved by shrinkage. *Journal of Forecasting* 12, 1-11.

LIU, C. and RUBIN, D. (1994) The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81, 633-48.

MENG, X. L. and RUBIN, D. B. (1991) Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association* 86, 899-909.

MENG, X. L. and RUBIN, D. B. (1993) Maximun likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80, 267-78.

MIN, C. and ZELLNER A. (1993) Bayesian and non-Bayesian methods for combining models and forecast with applications to forecasting international growth rates. *Journal of Econometrics* 56, 89-118.

MORRIS, C. N. (1983) Parametrical empirical Bayes inference: theory and applications, with discussion. *Journal of the American Statistical Association* 78, 47-66.

PEÑA, D. and BOX, G. (1987) Identifying a simplifying structure in time series. *Journal of the American Statistical Association* 82, 836-43.

REINSEL, G. C. and AHN, S. K. (1992) Vector autoregresive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting. *Journal of Time Series Analysis* 13, 353-75.

SHUMWAY, R. H. and STOFFER, D. S. (1982) An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* 3, 253-64.

STOCK, J. H. and WATSON, M. W. (1988) Testing for common trends. *Journal of the*

American Statistical Association 83, 1097-107.

TIAO, G. C. and TSAY, R. S. (1989) Model specification in multivariate time series. *Journal of the Royal Statistical Society* Serie B 51, 157-213.

WATSON, M.W. and ENGLE, R. F. (1983) Alternative algorithms for the estimation of dynamic, mimic and varying coefficient regression models. *Journal of Econometrics* 23, 385-400.

WALL, K. D. (1987) Identification theory for varying coefficient regression models. *Journal of Time Series Analysis* 8, 359-71.

ZELLNER, A., and HONG, C. (1989) Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics* 40, 183-202.