

# A DYNAMIC FACTOR MODEL FOR ECONOMIC TIME SERIES

**F. J. Fernández-Macho\***

Dpto. de Econometría y Estadística

& Instituto de Economía Pública,

*Universidad del País Vasco-Euskal Herriko Unibertsitatea.*

e-mail (internet): `etpfemaj@bs.ehu.es`

as published in *Kybernetika*, vol. 33, n<sup>o</sup> 6, 1997; pp. 583–606.

## Abstract

A dynamic factor model is introduced which may be viewed as an alternative to vector autoregressions in the treatment of cointegration. An obvious way of introducing dynamics in the standard factor analysis is to allow a realization of the common factors at a specific time interval to work its way through to the observed variables in several time periods. A problem arises however, when representing economic time series which generally are nonstationary. In this paper the dynamic factor model considered can handle nonstationarity rather trivially *via* unobserved factors with unit roots. The stochastic behaviour of these factors is explicitly modeled, and therefore the model is a member of the multivariate structural time series model class. A situation in which we might wish to entertain such a model is when considering two or more related economic variables which, as is often the case, appear to exhibit a common trend, and hence are cointegrated. The paper investigates the maximum likelihood estimation in the frequency domain and a scoring algorithm is provided. Also a generalization is considered in which independent common factors are made up of stochastic trends with stochastic common slopes and stochastic seasonals.

**Key Words:** Common Dynamic Factor, Common Trend, Cointegration, Maximum Likelihood Estimation, Frequency Domain, Structural Time Series Model, Unit Root, Unobserved Component.

---

\*This paper is based on research conducted while the author was research scholar in the Dept. of Stats. and Math. Sciences at the London School of Economics, for which financial support from the Dept. of Education of the Basque Government and the LSE's Suntory-Toyota foundation is acknowledged.

# 1 Introduction.

The analysis of cointegrated systems through vector autoregressions (VAR) has become a standard procedure in applied macroeconometrics following Johansen (1988, 1991) In many instances the main interest of the analysis consists in the extraction of dynamic common factors, such as common trends, and, although the vector moving-average (VMA) representation —determining the way in which nonstationarity is generated in the system— can be obtained from the VAR representation, it may be argued that if the main objective is the extraction of permanent components then possibly a better idea would be to formulate *directly* a model taking care of such permanent components. The corresponding VAR representation will in practice have a very high order (probably infinite) and clearly will not be appropriate for this purpose. As an alternative a dynamic factor model may be used.

The standard factor analysis (FA) was originally developed mainly to analyze intelligence tests so as to determine whether “intelligence” is made up of the combination of a few factors measuring attributes like “memory”, “mathematical ability”, “reading comprehension”, *etc.* In this sense, the basic idea of FA is, given observations on  $n$  variables, to assume a proper statistical model in which each observed variable is a linear function of  $k < n$  unobserved common components or factors plus a residual error term, *i.e.*

$$y_t = A \eta_t + e_t, \quad t = 0 \dots, T. \quad (1)$$

$(n \times 1)$        $(n \times k)$   $(k \times 1)$        $(n \times 1)$

Most applications of standard FA have been in the search for latent variables explaining psychological and sociological cross-section data. We note however that since dynamic effects are absent from the analysis, the technique is clearly inappropriate for analyzing time series data. An obvious way of introducing dynamics in (1) is to allow a realization of the common factors at a specific time period to work its way through to the observed variables in several time periods. In other words, we may assume a distributed-lag factor model,

$$y_t = A(L)\eta_t + e_t, \quad t = 0 \dots, T, \quad (2)$$

where  $A(L)$  is a polynomial matrix in the lag operator, *i.e.*  $A(L) = \sum_{r=0}^{\infty} A_r L^r$ , and the factors in  $\eta_t$  and the error terms in  $\varepsilon_t$  are generated by *stationary* random processes. For example Brillinger (1975, chapter 9) investigates the problem of representing a stationary series as a filtered version of a stationary signal series of reduced dimension plus an error series. Similar FA models have also been considered by Anderson (1963), Geweke (1977), Geweke & Singleton (1981), Engle & Watson (1981), Molenaar (1985) and others. Other data reduction techniques have also been considered by Priestley & Subba Rao (1973), Priestley, Subba Rao & Tong (1974), Subba Rao (1975), Box & Tiao (1977), Sims (1981)

and Velu, Reinsel & Wichern (1986). These techniques, unlike the dynamic FA model, pertain to the case in which observable input series are assumed to be given, *e.g.* lagged dependent variables. In particular, in the approach of Box & Tiao (1977) the original time series are assumed to follow a multiple stationary autoregressive model; then principal components of the one-step-ahead forecast error covariance matrix are extracted so as to obtain a transformed process whose components are ordered from least to most predictable.

As a first step towards identification of the structure (2) we will assume henceforth that  $A(L)$  is a geometric distributed lag, *i.e.*  $A_r = A\Phi^r$ , where  $\Phi$  is a  $(k \times k)$  matrix such that, in order to keep  $\{y_t\}$  stationary, its eigenvalues are less than one in absolute value. Thus

$$y_t = A \sum_{r=0}^{\infty} \Phi^r \eta_{t-r} + e_t, \quad t = 0 \dots, T, \quad (3)$$

which is equivalent to

$$y_t = A\mu_t + e_t, \quad t = 0 \dots, T, \quad (4)$$

$$\mu_t = \Phi\mu_{t-1} + \eta_t, \quad (5)$$

*c.f.* (1). This suggests a reinterpretation of the dynamic FA model in which  $\mu_t$ , rather than  $\eta_t$ , is the vector of common factors, being generated by a dynamic mechanism in the form of a transition equation. Thus (4)-(5) is a special case of the “state-space” model used in engineering to represent certain physical processes. Engle & Watson (1981) use a similar one-factor model (which they also described as “dynamic multiple indicator” model) to obtain estimates of the unobserved metropolitan wage rate for Los Angeles based on observations of sectorial wages. They use a time domain approach based on the Kalman filter (Pagan 1975, Harvey & Todd 1983) which may be computationally very demanding for multivariate time series. Later in this paper the time domain structure of the model is estimated from the spectral likelihood function as explained in Fernández-Macho (1990).

## 2 Unit roots and common trends.

Up to here we have considered the observed series (and hence the factors  $\mu_t$  in (4)-(5) to be stationary. This is certainly not very realistic if they are to represent economic time series. Yet previous techniques seem to run into trouble when attempting to tackle this problem. For example, if  $\{y_t\}$  is nonstationary in (2), the lag structure must be infinitely long thus rendering the analysis impossible (Molenaar 1985). In Box & Tiao (1977) the “most predictable” component will be nearly nonstationary representing the dynamic growth characteristic of economic series. However they note that the technique will break down in

the presence of strict nonstationarity. They also mention that differencing is of no help in this case since when analyzing multiple time series of this kind, it might be that the dynamic pattern in the data is caused by a small subset of nonstationary components, in which case differencing all the series could lead to complications in the analysis, particularly in the form of strict non-invertibility (which relates to the problem of cointegration treated below). It might also be worth mentioning that in Engle & Watson (1981) the unobserved component —metropolitan wage rate— appears to be nonstationary.

On the other hand the dynamic FA model (4)-(5) can handle nonstationary series rather trivially. Thus in the sequel we will consider a nonstationary version of (4)-(5) in which the transition equation has  $\Phi$  set to the identity matrix and a deterministic drift is also present. Further, it will be assumed for simplicity that the disturbance terms follow multivariate NID processes. (More generally they might be allowed to follow AR processes of low order, as in Fernández-Macho, Harvey & Stock (1987), but the statistical treatment is essentially the same). This simple dynamic factor model thus takes the form

$$\begin{aligned} \underset{(n \times 1)}{y_t} &= \underset{(n \times 1)}{\gamma} + \underset{(n \times k)}{A} \underset{(k \times 1)}{\mu_t} + \underset{(n \times 1)}{\varepsilon_t}, \quad t = 0 \dots, T, \\ \underset{(k \times 1)}{\mu_t} &= \underset{(k \times 1)}{\mu_{t-1}} + \underset{(k \times 1)}{\delta} + \underset{(k \times 1)}{\eta_t}, \end{aligned} \quad (6)$$

where  $\gamma$  and  $\delta$  are vectors of deterministic intercepts and slopes respectively and

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \text{NID} \left[ 0, \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \right].$$

That is, the common factors are specifically modeled as random-walk-cum-drift components and therefore they will be interpreted as common stochastic —or local— linear trends; *c.f.* the Multivariate exponential smoothing (MES) model in Fernández-Macho (1990).

Note that we assume  $0 < k < n$ . In the extreme cases  $k = 0$  or  $k = n$ , no common factors are present: the former collapses trivially to  $y_t = \varepsilon_t$  and the latter to the MES model with component  $\mu_t^\dagger = A\mu_t$ .

As it stands, model (6) is not identifiable. For example defining  $A^\dagger = A\Sigma_\eta^{-\frac{1}{2}}P'$ ,  $\mu_t^\dagger = P\Sigma_\eta^{-\frac{1}{2}}\mu_t$ ,  $\delta_t^\dagger = P\Sigma_\eta^{-\frac{1}{2}}\delta_t$ ,  $\eta_t^\dagger = P\Sigma_\eta^{-\frac{1}{2}}\eta_t$ , where  $P$  is an orthogonal matrix, we obtain an alternative dynamic FA model

$$\begin{aligned} y_t &= \gamma + A^\dagger \mu_t^\dagger + \varepsilon_t, \quad t = 0 \dots, T, \\ \mu_t^\dagger &= \mu_{t-1}^\dagger + \delta_t^\dagger + \eta_t^\dagger, \end{aligned}$$

in which the trend innovation covariance matrix is  $I_n$  for any choice of  $P$ , (note that the factors remain independent). In order to identify a structure we choose within each equivalence class that member satisfying the following restrictions on  $A$  and  $\Sigma_\eta$ :

$A$  is formed by the first  $k$  columns of a  $(n \times n)$  unit-lower-triangular matrix

$$A = [a_{ij}/a_{ij} = 0, \quad i < j; \quad a_{ii} = 1] \quad (7)$$

and  $\Sigma_\eta$  is a diagonal matrix

$$\Sigma_\eta = [\sigma_{\eta,ij}/\sigma_{\eta,ij} = 0, \quad i \neq j; \quad \sigma_{\eta,ii} = \sigma_{\eta,i}^2]. \quad (8)$$

Also it will be assumed that  $\Sigma_\varepsilon$  as well as  $\Sigma_\eta$  are of full rank so that the statistical treatment presented in section 4 does not break down.

Constraining  $\Sigma_\eta$  to be diagonal also ensures that the common factors are independent as is customary in standard FA. An interesting point here is that the factors themselves might be given an economic interpretation. In such a case it is sometimes useful to consider a rotation of the estimated factors. An appropriate choice of matrix  $P$  above may be used to redefine the common factors so as to give the desired interpretation.

### 3 Cointegration.

A situation in which we might wish to set down such a model as (6) is when considering two or more related economic variables which, as it is often the case, evolve in time in such a fashion that they do not diverge from each other. In other words, they appear to exhibit a common trend. An obvious example might be the prices of the same merchandise at different locations or the prices of close substitutes in the same market. In this case the common trend could be interpreted as the “latent” market price so that observed divergences are attributed to specific factors. Other typical examples are interest rates of different terms, national income and consumption, *etc.* Although individually all these economic series need differencing, it has already been mentioned that differencing a multiple time series is not appropriate if common trends are suspected: in general more unit roots than necessary will be imposed and only relationships between changes will be investigated while important relationships between the levels of the variables will be lost. Assuming that there are  $k < n$  common trends, only  $k$  unit roots should be imposed.

Let us consider the dynamic factor model (6). Since the factor loading matrix  $A$  is of full column rank  $k$  its orthogonal complement in a basis of  $\mathfrak{R}^n$ , say matrix  $B$ , will have full column rank  $n - k$  and its columns will be orthogonal to those of  $A$ , *i.e.*  $B'A = 0$ . Thus in (6)

$$B'y_t = \gamma + \xi_t, \quad t = 0 \dots, T, \quad (9)$$

where  $\xi_t \sim \text{NID}(0, B'\Sigma_\varepsilon B)$ . This means that there exists  $n - k$  linear combinations of the observed variables for which the trend components cancel out so that the vector of such linear combinations follow a stationary vector process even though each of the observed variables is best described by an integrated ARMA process. Time series which together exhibit this feature are called cointegrated (CI) series. Granger & Engle (1985) gave the following general definition:

“If each element of a vector of time series  $y_t$  must be differenced  $d$  times to achieve stationarity, but linear combinations  $B'y_t$  need be differenced only  $d - b$  times, the time series  $y_t$  are said to be CI of order  $(d, b)$  with cointegrating matrix  $B$ .”

In our case it follows that the series  $y_t$  in the dynamic factor model (6) are CI of order  $(1, 1)$ .

The converse is also true. As can be seen in Stock & Watson (1988), data generated by a CI process with  $n - k$  linearly independent vectors can be represented as linear combinations of  $k$  random-walk “trend” variables plus  $n - k$  “transitory” variables.

As an example, it is easy to see that in the typical bivariate case

$$\begin{aligned} y_{1t} &= y_{1,t-1} + a_{1t}, \\ y_{2t} &= \delta y_{1t} + a_{2t}, \end{aligned}$$

in Box & Tiao (1977, sec 4.4) or Hillmer & Tiao (1979) illustrating the problems which can arise with differencing (*e.g.* strict noninvertibility) there must be a common trend since the observed series are CI. The cointegrating vector is  $b = (-\delta, 1)'$  and  $\{b'y_t\}$  is stationary.

Back to our model we can see from (9) that  $B'y_t$  will wander around its mean. For that reason  $B'\bar{y}_t = \gamma$ , where  $\{\bar{y}_t\}$  represents the mean course of  $\{y_t\}$ , can be interpreted as a long run equilibrium towards which the observed variables are pushed back by economic forces whenever they drift apart. Also, at a particular time  $t = \tau$ ,  $\xi_\tau = B'(y_\tau - \bar{y}_\tau)$  is a measure of current disequilibrium.

Finally, since  $y_t \sim \text{CI}(1, 1)$ , it follows from theorem 1 in Engle & Granger (1987) that there exists an error-correction representation of the dynamic FA model (6), *i.e.* in such a representation a proportion of the disequilibrium  $\xi_\tau$  from one period  $\tau$  is corrected in the next period. That model (6) has error-correction representation is an interesting property because error-correction mechanisms are used very often in econometrics (*e.g.* Sargan (1981) and more recently Davidson, Hendry, Srba & Yeo (1978) and Salmon (1982) among others).

## 4 Maximum likelihood estimation of the FA model.

Apart of the intercept  $\gamma$ , which does not enter into the likelihood, there are  $\frac{1}{2}[k(3+2n-k) + n(n+1)]$  parameters to be estimated in the dynamic factor model (6) as follows:  $k$  parameters in the drift vector  $\delta$ ,  $nk - \frac{1}{2}k(k+1)$  in the factor loading matrix  $A$ ,  $k$  in the common trend innovation covariance matrix  $\Sigma_\eta$ , and  $\frac{1}{2}n(n+1)$  in the transitory term covariance matrix  $\Sigma_\varepsilon$ .

Before going any further, it is convenient that we define column vectors containing these unknown parameters, (*i.e.* eliminating those elements which are already set). Thus, let  $\alpha$  be a vector obtained from  $\text{vec } A$  by eliminating those elements above *and* on the main diagonal. Similarly, let  $\text{diag } \Sigma_\eta$  be the vector of diagonal elements of  $\Sigma_\eta$  and let  $\mathbf{v} \Sigma_\varepsilon$  be the vector of distinct elements of  $\Sigma_\varepsilon$  obtained from  $\text{vec } \Sigma_\varepsilon$  by eliminating all supradiagonal elements: *i.e.*

$$\alpha = [a_{ij} \in A / i > j], \quad \text{diag } \Sigma_\eta = [\sigma_{\eta,ij} \in \Sigma_\eta / i = j], \quad \mathbf{v} \Sigma_\varepsilon = [\sigma_{\varepsilon,ij} \in \Sigma_\varepsilon / i \geq j].$$

Let us further define 0-1 matrices mapping  $\alpha$ ,  $\text{diag } \Sigma_\eta$  and  $\mathbf{v} \Sigma_\varepsilon$  into  $\text{vec } A$ ,  $\text{vec } \Sigma_\eta$  and  $\text{vec } \Sigma_\varepsilon$  respectively, *i.e.*

$$D_\alpha \alpha + \text{vec} \begin{bmatrix} I_k \\ 0_{n-k} \end{bmatrix} = \text{vec } A, \quad H \text{diag } \Sigma_\eta = \text{vec } \Sigma_\eta, \quad D \mathbf{v} \Sigma_\varepsilon = \text{vec } \Sigma_\varepsilon,$$

so that

$$\frac{\partial \text{vec } A}{\partial \alpha'} = D_\alpha, \quad \frac{\partial \text{vec } \Sigma_\eta}{\partial (\text{diag } \Sigma_\eta)'} = H, \quad \frac{\partial \text{vec } \Sigma_\varepsilon}{\partial (\mathbf{v} \Sigma_\varepsilon)'} = D.$$

This matrices help to simplify the analysis since the derivatives with respect to the unknown parameters can be obtained from the derivatives with respect to the corresponding  $\text{vec}$  by straightforward application of the chain rule, *i.e.*  $\frac{\partial L}{\partial \alpha} = \left( \frac{\partial \text{vec } A}{\partial \alpha'} \right)' \frac{\partial L}{\partial \text{vec } A} = D_\alpha' \frac{\partial L}{\partial \text{vec } A}$ , and so on.

Since  $y_t \sim \text{CI}(1, 1)$  in model (6) taking first differences we obtain a stationary series

$$z_t = \Delta y_t - A\delta = A\eta_t + \Delta \varepsilon_t, \quad t = 1 \dots, T. \quad (10)$$

Let us consider the Fourier transform of  $\{z_t\}$

$$w_j = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T z_t e^{i\lambda_j t}, \quad \lambda_j = \frac{2\pi j}{T}, \quad j = 0 \dots, T-1. \quad (11)$$

This can be expressed more compactly as

$$\text{vec } W = \begin{pmatrix} U & \otimes & I_n \end{pmatrix} \text{vec } Z, \quad (12)$$

$\begin{matrix} (nT \times 1) & & (T \times T) & & (n \times n) & & (nT \times 1) \end{matrix}$

where  $W = (w_0 \dots, w_{T-1})$ ,  $Z = (z_1 \dots, z_1)$ ,  $U$  is the  $(T \times T)$  Fourier matrix whose  $(h, k)^{th}$  element is  $(2\pi T)^{-\frac{1}{2}} \exp(ik\lambda_{h-1})$ , and  $I_n$  denotes the identity matrix of order  $n$ . Note also that  $UU^* = (2\pi)^{-1} I_T$ , *i.e.*  $U$  is proportional to a unitary matrix by a factor of  $(2\pi)^{-\frac{1}{2}}$ .

Since  $\{z_t\}$  is a multivariate stationary nondeterministic gaussian process,  $\{w_j\}$  is asymptotically distributed as a normal independent zero-mean heteroscedastic process, *i.e.*

$$w_j \stackrel{a}{\sim} \mathcal{N}\left(0, \frac{1}{2\pi}G_{zj}\right), \quad j = 0 \dots, T-1,$$

where  $G_{zj}$  is the autocovariance matrix generating function (AMGF) of  $\{z_t\}$  evaluated at  $\lambda_j = 2\pi j/T$ . From (10) it is easy to see that

$$\begin{aligned} G_z(u) &= A\Sigma_\eta A' + (1-u)(1-u^{-1})\Sigma_\varepsilon, \\ G_{zj} &= G_z(e^{i\lambda_j}) = A\Sigma_\eta A' + c_j\Sigma_\varepsilon, \quad c_j = 2 - 2\cos\lambda_j. \end{aligned} \quad (13)$$

Since the  $w_j$ 's are independent the joint density function is simply the product of the individual densities whose logarithm, apart of the usual constant, is equal to

$$\ell_j = -\frac{1}{2}[\log \det G_{zj} + \text{tr} G_{zj}^{-1}(2\pi P_{zj})], \quad j > 0, \quad (14)$$

where  $\Sigma_\varepsilon > 0$  ensures that  $G_{zj} > 0$  for  $j > 0$  and  $P_{zj}$  denotes the *real* part of the periodogram matrix of  $\{z_t\}$  at frequency  $\lambda_j$ .

It is also clear that, for  $j=0$ ,  $G_{z0} = A\Sigma_\eta A'$  is of deficient rank  $k < n$ . Thus  $w_0$  has a singular (or degenerated) multivariate normal distribution, and no explicit determination of the density function is possible in  $\Re^n$ . However the density exists in a subspace and, according to Rao & Mitra (1971, p. 204), the logarithmic density of  $w_0$  in the hyperplane  $K'w_0 = 0$  (where  $K$  is a  $n \times (n-k)$  matrix of rank  $(n-k)$  such that  $K'G_{z0} = 0$  and  $K'K = I_{n-k}$ ) can be written as

$$\ell_0 = -\frac{1}{2}\log\varphi - \pi w_0^* G_{z0}^+ w_0,$$

where  $\varphi$  is the product of nonzero eigenvalues of  $G_{z0}$  and  $G_{z0}^+$  denotes the Moore-Penrose generalized inverse of  $G_{z0}$ .

As we know that for any matrix  $X$ ,  $XX^*$  and  $X^*X$  have the same nonzero eigenvalues writing  $X = A\Sigma_\eta^{\frac{1}{2}}$  we find that, since  $A'A$  is of full rank,  $\varphi = \det(A'A)\det\Sigma_\eta$ . Besides  $G_{z0}^+ = (A^+)'\Sigma_\eta^{-1}A^+$  where  $A^+ = (A'A)^{-1}A'$  since  $A$  is of full column rank.

Therefore the logarithm of the density function of  $w_0$  is

$$\ell_0 = -\frac{1}{2}\log\det(A'A) - \frac{1}{2}\log\det\Sigma_\eta - \frac{1}{2}\text{tr}\Sigma_\eta^{-1}[A^+(2\pi P_{z0})(A^+)]. \quad (15)$$

Finally from (12) we see that since  $U$  is  $(2\pi)^{-\frac{1}{2}}$  times a unitary matrix, the likelihood function—the density function of  $\text{vec} Z$  as a function of the sample—is  $(2\pi)^{-nT/2}$  times the density function of  $\text{vec} W$ . Thus the log-likelihood function can be written as

$$L = -\frac{nT}{2}\log(2\pi) + \sum_{j=0}^{T-1} \ell_j,$$



where the  $\ell_j$ 's are given by (14) and (15).

The periodogram of  $\{z_t\}$ , whose real part  $\{P_{zj}, j = 0 \dots, T-1\}$  is needed in (14), cannot be directly computed from the sample as it depends on the unknown  $\delta$ . From (10)-(11) and since

$$\sum_{t=1}^T e^{i\lambda_j t} = \begin{cases} T, & j = 0 \\ 0, & \text{otherwise} \end{cases},$$

we get

$$P_{zj} = \begin{cases} \frac{T}{2\pi} hh', & j = 0 \\ P_{\Delta y, j}, & \text{otherwise} \end{cases},$$

where  $h = T^{-1}(y_T - y_0) - A\delta$ . As  $\delta$  only appears in the log-likelihood *via*  $h$  in  $\ell_0$ , it can be easily concentrated out. First of all take into account that in (15)

$$A^+(2\pi P_{z0})(A^+)' = T(A^+h)(A^+h)', \quad (16)$$

being  $A^+h = T^{-1}A^+(y_T - y_0) - \delta$ , so that

$$-\frac{1}{2}\text{tr} \Sigma_\eta^{-1}[A^+(2\pi P_{z0})(A^+)] = -\frac{T}{2}(A^+h)' \Sigma_\eta^{-1}(A^+h). \quad (17)$$

Thus we can write

$$\frac{\partial L}{\partial \delta} = \frac{\partial \ell_0}{\partial \delta} = T \Sigma_\eta^{-1} A^+ h. \quad (18)$$

This is zero if and only if  $A^+h = 0$ ; therefore the ML estimator of  $\delta$  conditional on  $A$  is

$$\tilde{\delta}(A) = T^{-1}A^+(y_T - y_0). \quad (19)$$

From (18) we can also get the second derivatives involving  $\delta$  as follows

$$\begin{aligned} \frac{\partial^2 L}{\partial \delta \partial \delta'} &= T \Sigma_\eta^{-1} \frac{\partial(A^+h)}{\partial \delta'} = -T \Sigma_\eta^{-1}, \\ \frac{\partial^2 L}{\partial \delta \partial \alpha'} &= T \Sigma_\eta^{-1} \frac{\partial(A^+h)}{\partial(\text{vec } A)'} \frac{\partial \text{vec } A}{\partial \alpha'} = [(y_T - y_0)' \otimes \Sigma_\eta^{-1}] \frac{\partial \text{vec } A^+}{\partial(\text{vec } A)'} D_\alpha, \end{aligned}$$

where

$$\frac{\partial \text{vec } A^+}{\partial(\text{vec } A)'} = C_{nk}[(A'A)^{-1} \otimes (I_n - AA^+)] - [(A^+)' \otimes A^+], \quad (20)$$

being  $C_{nk}$  the commutation matrix such that for any matrix  $X$ ,  $\text{vec } X$  is converted into  $\text{vec } X'$ . Finally

$$\frac{\partial^2 L}{\partial \delta \partial(\text{diag } \Sigma_\eta)'} = T \frac{\partial(\Sigma_\eta^{-1} A^+ h)}{\partial(\text{vec } \Sigma_\eta)'} \frac{\partial \text{vec } \Sigma_\eta}{\partial(\text{diag } \Sigma_\eta)'} = -T[(A^+h)' \Sigma_\eta^{-1} \otimes \Sigma_\eta^{-1}] H,$$

but note however that  $E(\frac{\partial^2 L}{\partial \delta \partial(\text{diag } \Sigma_\eta)'}) = 0$  because  $E(A^+h) = 0$ . It is also obvious that

$$\frac{\partial^2 L}{\partial \delta \partial(\text{v } \Sigma_\varepsilon)'} = 0.$$

Substituting  $\delta$  by  $\tilde{\delta}$  in (16) we see that (17) (*i.e.* the last term in the expression (15) for  $\ell_0$ ) vanishes so that the concentrated log-likelihood can be written as

$$L^c = -\frac{nT}{2} \log(2\pi) - \frac{1}{2} \log \det(A'A) - \frac{1}{2} \log \det \Sigma_\eta + \sum_{j=1}^{T-1} \ell_j, \quad (21)$$

where the  $\ell_j$ 's are given by (14) with  $P_{zj} \equiv P_{\Delta y, j}$  for  $j = 1 \dots, T-1$ .

Let  $\vartheta$  be the vector of parameters to be estimated *via* maximisation of (21): *i.e.*  $\vartheta = \{\alpha', (\text{diag } \Sigma_\eta)', (\text{v } \Sigma_\varepsilon)'\}'$ . As the derivative vector  $dL^c(\vartheta)$  and the (concentrated) information matrix  $\Phi(\vartheta) = -E \frac{\partial^2 L^c}{\partial \vartheta \partial \vartheta'}$  are relatively easy to construct, a scoring algorithm is appropriate for ML estimation. This procedure involves finding a direction vector,  $p(\vartheta) = \Phi(\vartheta)^{-1} dL^c(\vartheta)$ , to obtain new estimates from the recursion  $\vartheta_{\kappa+1} = \vartheta_\kappa + \rho_\kappa p(\vartheta_\kappa)$  (with  $\rho_\kappa$  as a step length) and iterate until convergence. Analytic expressions for the derivatives are as follows.

From (21), since  $\frac{\partial \log \det(A'A)}{\partial \text{vec } A} = 2 \text{vec} [(A^+)']$ ,

$$\frac{\partial L^c}{\partial \text{vec } A} = -\text{vec} [(A^+)'] + \sum_{j=1}^{T-1} \left( \frac{\partial \ell_j}{\partial \text{vec } A} \right). \quad (22)$$

Also, since for any nonsingular square matrix  $X$  it is well known that  $d \log \det X = \text{tr} [X^{-1}(dX)] = (\text{vec } X^{-1})(\text{dvec } X)$ , we have

$$\frac{\partial L^c}{\partial \text{vec } \Sigma_\eta} = -\frac{1}{2} \text{vec } \Sigma_\eta^{-1} + \sum_{j=1}^{T-1} \left( \frac{\partial \ell_j}{\partial \text{vec } \Sigma_\eta} \right), \quad (23)$$

and finally

$$\frac{\partial L^c}{\partial \text{vec } \Sigma_\varepsilon} = \sum_{j=1}^{T-1} \left( \frac{\partial \ell_j}{\partial \text{vec } \Sigma_\varepsilon} \right). \quad (24)$$

Note that for positive definite  $G_{zj}$ , *i.e.* for  $j > 0$  here, it can be found that

$$\frac{\partial \ell_j}{\partial \psi} = \frac{1}{2} \left( \frac{\partial \text{vec } G_{zj}}{\partial \psi'} \right)' m_j, \quad \psi \in \{\text{vec } A, \text{vec } \Sigma_\eta, \text{vec } \Sigma_\varepsilon\},$$

where  $m_j = \text{vec} [G_{zj}^{-1}(2\pi P_{\Delta y, j})G_{zj}^{-1} - G_{zj}^{-1}]$ .

Differentiating (22)-(23)-(24) further and denoting  $\Phi_j(\psi_1, \psi_2) = -E \left( \frac{\partial^2 \ell_j}{\partial \psi_1 \partial \psi_2} \right)$ ,  $\psi_i \in \{\text{vec } A, \text{vec } \Sigma_\eta, \text{vec } \Sigma_\varepsilon\}$ ,  $i = 1, 2$ , we get that the diagonal blocks in the information matrix are

$$\Phi(\text{vec } A) = \frac{\partial \text{vec} [(A^+)']}{\partial (\text{vec } A)'} + \sum_{j=1}^{T-1} \Phi_j[\text{vec } A, (\text{vec } A)'],$$

where  $\frac{\partial \text{vec} [(A^+)']}{\partial (\text{vec } A)'} = [(A'A)^{-1} \otimes (I_n - AA^+)] - C_{kn}[(A^+)'] \otimes A^+$ ,

$$\Phi(\text{vec } \Sigma_\eta) = -\frac{1}{2} (\Sigma_\eta^{-1} \otimes \Sigma_\eta^{-1}) + \sum_{j=1}^{T-1} \Phi_j[\text{vec } \Sigma_\eta, (\text{vec } \Sigma_\eta)'],$$

$$\Phi(\text{vec } \Sigma_\varepsilon) = \sum_{j=1}^{T-1} \Phi_j[\text{vec } \Sigma_\varepsilon, (\text{vec } \Sigma_\varepsilon)'],$$

and the off-diagonal blocks

$$\Phi(\psi_1, \psi'_2) = \sum_{j=1}^{T-1} \Phi_j(\psi_1, \psi'_2), \quad \psi_1 \neq \psi_2.$$

As shown elsewhere (Fernández-Macho 1990), we have that for  $j > 0$

$$\Phi_j(\psi_1, \psi'_2) = \frac{1}{2} \left( \frac{\partial \text{vec } G_{zj}}{\partial \psi'_1} \right)' M_j \left( \frac{\partial \text{vec } G_{zj}}{\partial \psi'_2} \right),$$

with  $M_j = (G_{zj}^{-1} \otimes G_{zj}^{-1})$ , and it only remains to find expressions for the derivatives of  $\text{vec } G_{zj}$ . From (13) they are as follows

$$\frac{\partial \text{vec } G_{zj}}{\partial (\text{vec } A)'} = \frac{\partial \text{vec } (A \Sigma_\eta A')}{\partial (\text{vec } A)'} = 2N_n(A \Sigma_\eta \otimes I_n),$$

where  $N_n = \frac{1}{2}(I_{n^2} + C_{nn})$  is the ‘‘symmetrification’’ matrix such that for any square matrix  $X$  it returns the  $\text{vec}$  of  $\frac{1}{2}(X + X')$ , as defined in Magnus & Neudecker (1988),

$$\frac{\partial \text{vec } G_{zj}}{\partial (\text{vec } \Sigma_\eta)'} = \frac{\partial \text{vec } (A \Sigma_\eta A')}{\partial (\text{vec } \Sigma_\eta)'} = (A \otimes A),$$

and

$$\frac{\partial \text{vec } G_{zj}}{\partial (\text{vec } \Sigma_\varepsilon)'} = c_j I_{n^2}.$$

Summarizing

$$dL^c(\vartheta) = \begin{bmatrix} D'_\alpha \{ -\text{vec} [(A^+)'] + (\Sigma_\eta A' \otimes I_n) \sum_{j=1}^{T-1} m_j \} \\ H' [ -\frac{1}{2} \text{vec } \Sigma_\eta^{-1} + \frac{1}{2} (A' \otimes A') \sum_{j=1}^{T-1} m_j ] \\ D' (\frac{1}{2} \sum_{j=1}^{T-1} c_j m_j) \end{bmatrix},$$

$$\Phi(\vartheta) = \begin{bmatrix} \Phi_{11}(\vartheta) & \Phi_{12}(\vartheta)' & \Phi_{13}(\vartheta)' \\ \Phi_{12}(\vartheta) & \Phi_{22}(\vartheta) & \Phi_{23}(\vartheta)' \\ \Phi_{13}(\vartheta) & \Phi_{23}(\vartheta) & \Phi_{33}(\vartheta) \end{bmatrix},$$

with

$$\Phi_{11}(\vartheta) = D'_\alpha \{ [(A' A)^{-1} \otimes (I_n - A A^+)] - C_{nk} [(A^+)'] \otimes A^+ + 2(\Sigma_\eta A' \otimes I_n) \left( \sum_{j=1}^{T-1} M_j \right) N_n (A \Sigma_\eta \otimes I_n) \} D_\alpha$$

$$\Phi_{12}(\vartheta) = H' [(A' \otimes A') \left( \sum_{j=1}^{T-1} M_j \right) (A \Sigma_\eta \otimes I_n)] D_\alpha$$

$$\Phi_{22}(\vartheta) = H' [ -\frac{1}{2} (\Sigma_\eta^{-1} \otimes \Sigma_\eta^{-1}) + \frac{1}{2} (A' \otimes A') \left( \sum_{j=1}^{T-1} M_j \right) (A \otimes A) ] H$$

$$\begin{aligned}\Phi_{13}(\vartheta) &= D'[(\sum_{j=1}^{T-1} c_j M_j)(A\Sigma_\eta \otimes I_n)]D_\alpha \\ \Phi_{23}(\vartheta) &= D'[(\sum_{j=1}^{T-1} c_j M_j)(A \otimes A)]H \\ \Phi_{33}(\vartheta) &= D'[(\sum_{j=1}^{T-1} c_j^2 M_j)]D\end{aligned}$$

From these expressions a scoring algorithm can be constructed for ML estimation of the parameters in  $\vartheta$ . Once these parameters have been estimated,  $\delta$  is estimated from (19) and  $\gamma$  from

$$\hat{\gamma}' = [\underbrace{0 \dots 0}_k, \frac{1}{T} \sum (y_{2t} - \hat{A}_2 \hat{A}_1^{-1} y_{1t})']$$

where  $(y_t, A)$  has been partitioned so that  $(y_{1t}, A_1)$  contains the first  $k$  rows and  $(y_{2t}, A_2)$  the remaining  $n - k$  rows.

This scoring algorithm should be initialized with a consistent estimator which can be constructed from the moment estimator as explained in the appendix. Once starting consistent estimates of  $\Sigma_\varepsilon$ ,  $A$  and  $\Sigma_\eta$  have been calculated efficient ML estimates can be obtained by iterating the scoring algorithm until convergence. Unfortunately the limiting properties of the ML estimator are not easily established under the present circumstances. We can apply some suitable LLN (*e.g.* Dunsmuir & Hannan (1976)) to assess almost sure convergence to the true values but, as the spectral density matrix happens to be singular at  $\lambda = 0$ , the CLT in Dunsmuir (1979) breaks down. This is of course the usual unit root problem arising here because of the presence of common trends. There are  $k$  common trends in (6), therefore there are only  $k < n$  unit roots in the AR part of the corresponding ARIMA representation but differencing, as in (10), imposes  $n$  unit roots, that is,  $n - k$  more than needed. The effect is similar to that of overdifferencing a univariate series in that  $n - k$  unit roots appear in the MA part and this is reflected in the spectra as  $F(0)$  being of deficient rank  $k < n$ .

Let us review this more formally. From (10) we see that the difference vector series follow an MA(1) process, *i.e.*  $z_t = (I_n - \Theta)\nu_t$ , where  $\nu_t \sim \text{NID}(0, \Sigma_\nu)$ . Thus the spectrum matrix can be expressed in two alternative forms:

$$2\pi F_z(\lambda) = A\Sigma_\eta A' + c(\lambda)\Sigma_\varepsilon = (I_n - \Theta e^{i\lambda})\Sigma_\nu(I_n - \Theta' e^{-i\lambda}),$$

where  $c(\lambda) = 2 - 2\cos\lambda$ . For  $\lambda > 0$ ,  $\Sigma_\varepsilon > 0$  implies  $F_z(\lambda) > 0$  which in turn implies  $\Sigma_\nu > 0$ . But for  $\lambda = 0$

$$2\pi F_z(0) = A\Sigma_\eta A' = (I_n - \Theta)\Sigma_\nu(I_n - \Theta')$$

is of deficient rank  $k$  which means that  $\Theta$  must have  $n - k$  unit roots. This situation has been discussed in Sargan & Bhargava (1983) for the univariate MA(1) process where the ML estimator is shown to be not asymptotically normally distributed. By inference, we are then to expect a similar result in our present multivariate case.

If the limiting distribution of the ML estimator is not normal the classical procedures, *e.g.* likelihood ratio test, Wald test and LM test, all run into difficulties and cannot be applied for testing the presence and number of common trends. The problem has received some attention in the literature. In the univariate case it is associated with testing for a unit root in autoregressive processes. Among the earliest contributions see in particular Fuller (1976, sec 8.5), Dickey & Fuller (1979), Dickey & Fuller (1981) and Fuller (1985). In the multivariate case Fountis & Dickey (1983), Granger & Engle (1985) and Stock & Watson (1988) have proposed and compared a variety of tests for common trends.

## 5 An extension of the model.

The dynamic factor model (6) can be easily extended in a number of ways in order to handle more general kinds of data, thus providing an attractive alternative to the VAR context advocated by Johansen (1988), Johansen (1991) where such extensions are not so obvious. In this section a generalization is considered in which independent common factors are made up of stochastic trends with stochastic common slopes and stochastic seasonals, *i.e.*

$$\begin{aligned}
\underset{(n \times 1)}{y_t} &= \underset{(n \times 1)}{\gamma} + \underset{(n \times k)}{A} \underset{(k \times 1)}{f_t} + \underset{(n \times 1)}{\varepsilon_t}, \quad t = 1 - s \dots, 0 \dots, T, \quad (25) \\
\underset{(k \times 1)}{f_t} &= \underset{(k \times 1)}{\mu_t} + \underset{(k \times 1)}{\xi_t}, \\
\underset{(k \times 1)}{\Delta \mu_t} &= \underset{(k \times k_\delta)}{A_\delta} \underset{(k_\delta \times 1)}{\delta_{t-1}} + \underset{(k \times 1)}{\eta_t}, \\
\underset{(k_\delta \times 1)}{\Delta \delta_t} &= \underset{(k_\delta \times 1)}{\zeta_t}, \\
\underset{(k \times 1)}{S(L)\xi_t} &= \underset{(k \times 1)}{\omega_t}, \\
\begin{pmatrix} \varepsilon_t \\ \eta_t \\ \zeta_t \\ \omega_t \end{pmatrix} &\sim \text{NID} \left[ 0, \begin{pmatrix} \Sigma_\varepsilon & & & \\ & \Sigma_\eta & & \\ & & \Sigma_\zeta & \\ & & & \Sigma_\omega \end{pmatrix} \right],
\end{aligned}$$

where  $0 \leq k_\delta \leq k \leq n$  (strict inequalities in nontrivial cases) and  $S(L)$  represents the seasonal sum operator.

Since the factors  $f_t$  are assumed to be independent we must have that the covariance matrices  $\Sigma_\nu$ ,  $\nu \in \{\eta, \zeta, \omega\}$  are diagonal. Also, for identification, the loading matrices  $A$  and

$A_\delta$  are restricted to be respectively the first  $k$  and the first  $k_\delta$  columns of unit-lower-triangular matrices.

Note that in general the trend components  $\mu_t$  are themselves cointegrated in the sense that  $\mu_t$  has to be differenced twice to achieve stationarity but there exists  $B_\delta$  such that  $B'_\delta A_\delta = 0$  so that  $B'_\delta \mu_t$  is stationary after differencing only once. In the terminology of Engle & Granger (1987) we write  $\mu_t \sim \text{CI}(2, 1)$ .

As for the observed variables we will see below that  $\Delta \Delta_s y_t$  is stationary but it is clear that there exists  $B$  such that  $B'A = 0$  so that  $B'y_t$  is stationary. We may write  $y_t \sim \text{CI}(1, 1) \times (1, 1)_s$ —in an obvious notation extending that of Engle & Granger (1987) to allow for seasonality—which is a sort of *complete cointegration* as mentioned in Lee (1992).

The dynamic FA model (25) can be analyzed much in the same way as model (6) above. Since  $\Delta_s = \text{S}(L)\Delta$  we have that

$$\begin{aligned}\Delta \Delta_s \mu_t &= A_\delta \Delta_s \delta_{t-1} + \Delta_s \eta_t = A_\delta \text{S}(L) \zeta_{t-1} + \Delta_s \eta_t, \\ \Delta \Delta_s \xi_t &= \Delta^2 \text{S}(L) \xi_t = \Delta^2 \omega_t.\end{aligned}$$

Therefore

$$\begin{aligned}z_t &= \Delta \Delta_s y_t \\ &= A[\Delta_s \eta_t + A_\delta \text{S}(L) \zeta_{t-1} + \Delta^2 \omega_t] + \Delta \Delta_s \varepsilon_t, \quad t = 1, \dots, T,\end{aligned}$$

is stationary with AMGF given by

$$\begin{aligned}G_z(u) &= (1 - u^s)(1 - u^{-s})A\Sigma_\eta A' \\ &\quad + \text{S}(u)\text{S}(u^{-1})AA_\delta \Sigma_\zeta A'_\delta A' \\ &\quad + [(1 - u)(1 - u^{-1})]^2 A\Sigma_\omega A' \\ &\quad + (1 - u)(1 - u^{-1})(1 - u^s)(1 - u^{-s})\Sigma_\varepsilon.\end{aligned}$$

Thus setting  $u = \exp(i\lambda_j)$ ,  $\lambda_j = 2\pi j/T$ , we have, for  $j > 0$ ,

$$G_{zj} = G_z(e^{i\lambda_j}) = c_{sj}A\Sigma_\eta A' + (c_{sj}/c_{1j})AA_\delta \Sigma_\zeta A'_\delta A' + c_{1j}^2 A\Sigma_\omega A' + (c_{1j}c_{sj})\Sigma_\varepsilon, \quad (26)$$

where  $c_{rj} = 2 - 2\cos(r\lambda_j)$ ,  $r \in \{1, s\}$ , and for  $\lambda = 0$

$$G_{z0} = G_z(1) = s^2 AA_\delta \Sigma_\zeta A'_\delta A'.$$

As before the Fourier transform of  $\{z_t\}$ , say  $\{w_j\}$ , is distributed as  $\text{NID}(0, (2\pi)^{-1}G_{zj})$ ,  $j = 0 \dots, T - 1$ , asymptotically and in particular,  $w_0$  has a singular distribution. Following the same arguments as in the previous section the logarithm of the density of  $w_0$  in the

appropriate hyperplane of dimension  $(n - k_\delta)$  is  $\ell_0 = -\frac{1}{2} \log \varphi - \pi w_0^* G_{z_0}^+ w_0$  where  $\varphi$  is the product of nonzero eigenvalues of  $G_{z_0}$ , *i.e.*  $\varphi = s^{(2k_\delta)} \det(A'_\delta A' A A_\delta) \det \Sigma_\zeta$ , and  $G_{z_0}^+$  is the Moore-Penrose generalized inverse, *i.e.*  $G_{z_0}^+ = s^{-2} (A A_\delta)^+ \Sigma_\zeta (A A_\delta)^+$ . Therefore

$$\ell_0 = -k_\delta \log(s) - \frac{1}{2} \log \det(A'_\delta A' A A_\delta) - \log \det \Sigma_\zeta - (\pi s^{-2}) \text{tr} [\Sigma_\zeta^{-1} (A A_\delta)^+ P_{z_0} (A A_\delta)^+].$$

For  $j > 0$ ,  $\Sigma_\varepsilon > 0$  ensures that  $G_{z_j} > 0$  so that the logarithm of the density function of  $w_j$  is

$$\ell_j = -\frac{1}{2} \{ \log \det G_{z_j} + \text{tr} [G_{z_j}^{-1} (2\pi P_{z_j})] \}, \quad j > 0.$$

Let  $\vartheta$  be the unknown parameter vector in the model, *i.e.*  $\vartheta = \{\alpha', \alpha'_\delta, (\text{diag } \Sigma_\eta)'\}, (\text{diag } \Sigma_\zeta)', (\text{diag } \Sigma_\omega)', (\text{v } \Sigma_\varepsilon)'\}'$ , where  $\alpha$  and the operators  $\text{v}(\cdot)$ ,  $\text{diag}(\cdot)$  are as defined in the previous section and  $\alpha_\delta$  is as  $\alpha$  but with reference to  $A_\delta$ . Since the  $w_j$ 's are independent the log-likelihood is

$$L(\vartheta) = -\frac{nT}{2} \log(2\pi) + \sum_{j=0}^{T-1} \ell_j(\vartheta).$$

This is to be maximized with respect to the elements of  $\vartheta$ , perhaps using a quasi-Newton method like the Gill-Murray-Pitfield algorithm which does not need the explicit evaluation of derivatives since the construction of the Hessian, and hence of the scoring algorithm, is rather cumbersome in the present case. To give an example, the scoring algorithm will be shown only for the special case in which  $k_\delta = k$  and  $A_\delta \Sigma_\zeta A'_\delta = I_k$  so that, in addition to (7)-(8), the common factors remain independent.

Firstly, after some algebra we get

$$\begin{aligned} \frac{\partial \ell_0}{\partial \text{vec } A} &= -\text{vec}(A^+) - s^{-2} \text{vec} [(I_n - A A^+) (2\pi P_{z_0}) (A^+) (A' A)^{-1} - A^+ (2\pi P_{z_0}) (A^+) A^+], \\ \Phi_0(\text{vec } A) &= C_{nk} [I_k \otimes (A^+) A^+] + (2I_{nk} - C_{kn}) [(A^+) \otimes A^+] - C_{nk} [(A' A)^{-1} \otimes (I_n - A A^+)]. \end{aligned}$$

Secondly, for  $j > 0$ , we know that

$$\begin{aligned} \frac{\partial \ell_j}{\partial \vartheta} &= \frac{1}{2} \left( \frac{\partial \text{vec } G_{z_j}}{\partial \vartheta'} \right)' m_j, \\ \Phi_j(\vartheta) &= \frac{1}{2} \left( \frac{\partial \text{vec } G_{z_j}}{\partial \vartheta'} \right)' M_j \left( \frac{\partial \text{vec } G_{z_j}}{\partial \vartheta'} \right), \end{aligned}$$

where  $m_j = \text{vec} [G_{z_j}^{-1} (2\pi P_{z_j}) G_{z_j}^{-1} - G_{z_j}^{-1}]$ ,  $M_j = (G_{z_j}^{-1} \otimes G_{z_j}^{-1})$  and from (26)  $\text{dvec } G_{z_j} = c_{sj} \text{dvec} (A \Sigma_\eta A') + (c_{sj}/c_{1j}) \text{dvec} (A A') + c_{1j}^2 \text{dvec} (A \Sigma_\omega A') + (c_{1j} c_{sj}) \text{dvec} \Sigma_\varepsilon$ .

Since, in general, for *any* matrices  $B$ ,  $\Sigma_\nu$

$$\begin{aligned} \text{dvec} (B \Sigma_\nu B') &= \text{vec} [(dB) \Sigma_\nu B'] + \text{vec} [B \Sigma_\nu (dB)'] + \text{vec} [B (d\Sigma_\nu) B'] \\ &= 2N_n \text{vec} [(dB) \Sigma_\nu B'] + \text{vec} [B (d\Sigma_\nu) B'] \\ &= 2N_n (B \Sigma_\nu \otimes I_n) (\text{dvec } B) + (B \otimes B) (\text{dvec } \Sigma_\nu), \end{aligned}$$

we have that

$$\frac{\partial \text{vec } G_{zj}}{\partial (\text{vec } A)'} = 2N_n(AQ_j \otimes I_n),$$

where  $Q_j = c_{sj}\Sigma_\eta + c_{1j}^2\Sigma_\omega + (c_{sj}/c_{1j})I_n$ .

Also

$$\frac{\partial \text{vec } G_{zj}}{\partial (\text{vec } \Sigma_\eta)'} = c_{sj}(A \otimes A), \quad \frac{\partial \text{vec } G_{zj}}{\partial (\text{vec } \Sigma_\omega)'} = c_{1j}^2(A \otimes A), \quad \frac{\partial \text{vec } G_{zj}}{\partial (\text{vec } \Sigma_\eta)'} = (c_{sj}/c_{1j})I_{n^2},$$

from where we can easily construct the expressions for the first derivative vector  $dL(\vartheta)$  and the Hessian  $\Phi(\vartheta)$  of the log-likelihood function with respect to the parameters in  $\vartheta$ .

## 6 Factor extraction

Once the model parameters have been estimated, an interesting question is how to obtain estimates of the  $k$  unobserved, or hidden, factors. It goes without saying that since the dynamic FA model can be written in state-space form, the optimal factor estimators should be obtained applying the Kalman filter (forwards) and a smoother (backwards) Fernández-Macho (1990).

On the other hand multiplying the measurement equation of the FA model in (6) or in (25) by the Moore-Penrose generalized inverse of  $A$  we obtain

$$\mu_t = A^+(y_t - \gamma - \varepsilon_t), \quad t = 0 \dots, T,$$

and, therefore, a moment estimator can be easily obtained as linear combinations of the  $n$  observed series as follows:

$$\hat{\mu}_t = \hat{A}^+(y_t - \hat{\gamma})$$

which in contrast with the filtered optimal estimate, is contaminated with the errors  $\{\varepsilon_t\}$  and hence will not be optimal in general; *c.f.* Gonzalo & Granger (1995) whose estimator is essentially the same as  $\hat{\mu}_t$  and hence will not be optimal.

For example in an  $n$ -variate one-factor model (*i.e.*  $k = 1$ ) with  $A = (1, a_2 \dots, a_n)'$  this “quick” estimate of the common factor will be given by

$$\hat{\mu}_t = [y_{1t} + \hat{a}_2(y_{2t} - \hat{\gamma}_2) + \dots + \hat{a}_n(y_{nt} - \hat{\gamma}_n)] / (1 + \hat{a}_2^2 + \dots + \hat{a}_n^2) \quad (27)$$



## 7 Examples

### 7.1 Simulated data

Figure 1 (thin lines) plots 180 observations of two artificial series generated by

$$\begin{aligned} y_{1t} &= \mu_t + \varepsilon_{1t}, \\ y_{2t} &= 0.8\mu_t + \varepsilon_{2t}, \\ \mu_t &= \mu_{t-1} + 0.02 + \eta_t, \quad \mu_0 = 100, \\ \text{Var} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \eta_t \end{pmatrix} &= \begin{bmatrix} 0.008 & -0.004 & 0 \\ -0.004 & 0.02 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}, \end{aligned}$$

together with their underlying common factor (thick line).

Maximizing the log-likelihood (21) for a bivariate dynamic factor model, *i.e.* assuming a common drifted random-walk factor, we obtained the following results:

$$\begin{aligned} \text{factor loads:} & \hat{A} = (1 \quad 0.8401)' \\ \text{intercept vector:} & \hat{\gamma} = (0 \quad -4.1038)' \\ \text{factor innovation variance:} & \hat{\Sigma}_\eta = 0.0100 \\ \text{factor drift:} & \hat{\delta} = 0.0245 \\ \text{error covariance matrix:} & \hat{\Sigma}_\varepsilon = \begin{pmatrix} 0.0110 & \\ -0.0059 & 0.0159 \end{pmatrix} \end{aligned}$$

which are very close to the true ones used to generate the data.

Thus in the present one-factor bivariate case the common factor “quick” estimate given by (27) is

$$\hat{\mu}_t = 2.0212 + 0.5862y_{1t} + 0.4925y_{2t}. \quad (28)$$

Figure 2 compares this quick estimate of the common factor with both observed series (left axis:  $\{y_1, \hat{\mu}\}$ , right axis:  $\{y_2, \hat{\mu}\}$ ).

Notwithstanding the possibility of factor rotation already mentioned at the end of section 2, it must be said that the observed *systematic deviation* of the “quick” estimate in figure 2 from the underlying true factor in figure 1 is just consequence of estimation error and it can be expressed as

$$d\hat{\mu}_t = d\hat{A}^+(y_t - \hat{\gamma}) \quad (29)$$

Figure 1: Simulated data: drifted random walk plus errors

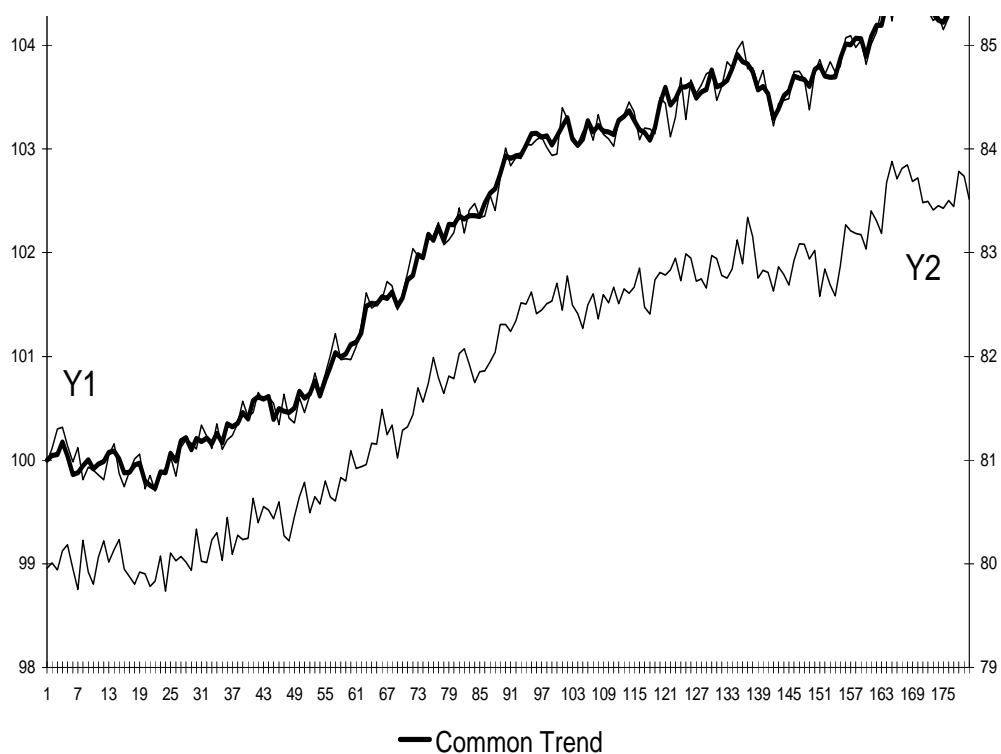


Figure 2: Simulated data: quick estimate of common factor

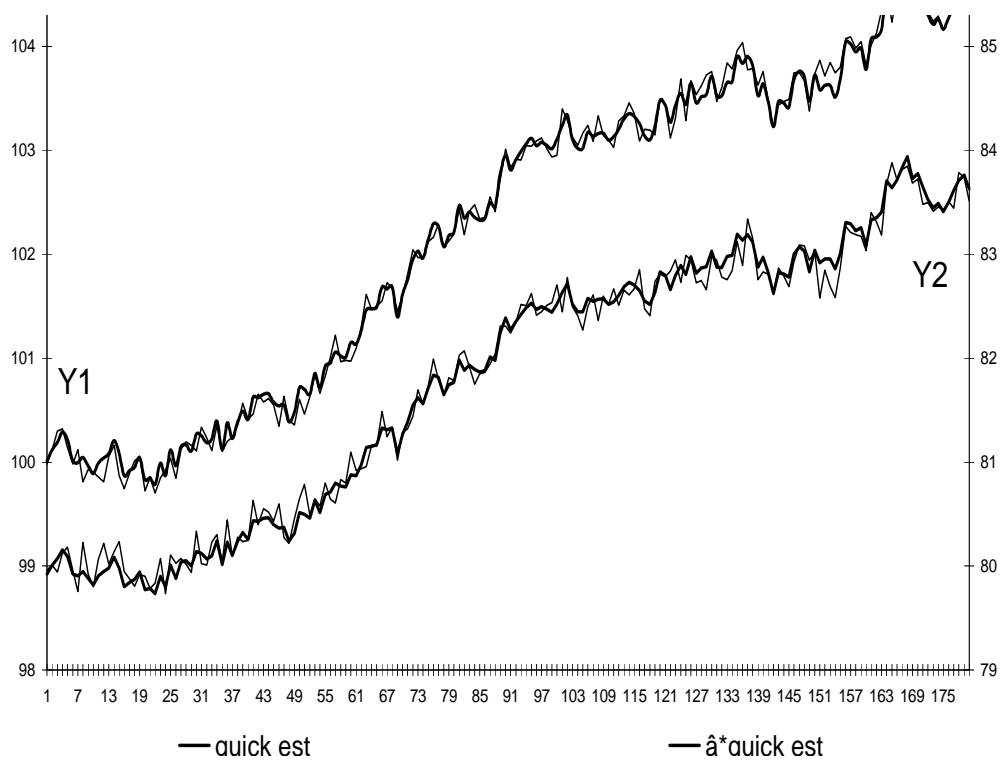
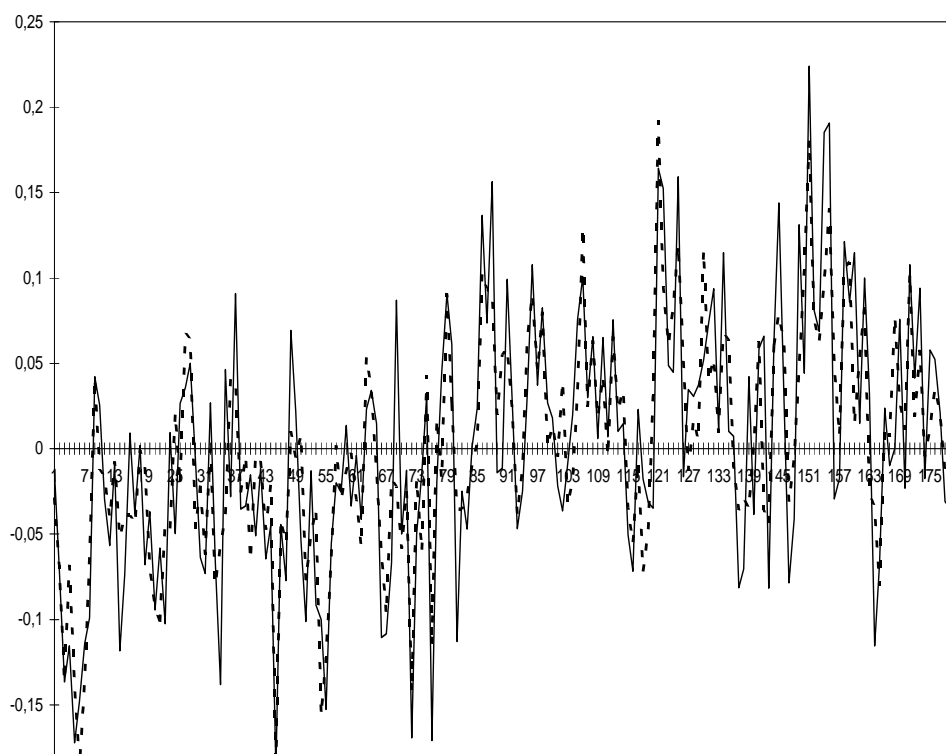


Figure 3: Common factor: estimation errors



where  $dA^+$  is obtained from (20). In the present bivariate example we have then that the quick estimate of the common factor will have a systematic deviation of

$$d\hat{\mu}_t = [(1 - \hat{a}^2)(y_{2t} - \hat{\gamma}_2) - 2\hat{a}y_{1t}]da / (1 + \hat{a}^2)^2$$

where  $da$  is the estimation error of the factor load  $a$ . The following table gives expressions for some chosen values of  $\hat{a}$ .

$\hat{a} = 0$	$d\hat{\mu}_t = da y_{2t}$
$\hat{a} = \pm 1$	$d\hat{\mu}_t = \mp \frac{da}{2} y_{1t}$
$\hat{a} = \pm 2$	$d\hat{\mu}_t = (-\frac{3}{25}y_{2t} \mp \frac{4}{25}y_{1t})da$
$\hat{a} = \pm 0.5$	$d\hat{\mu}_t = (\frac{75}{64}y_{2t} \mp \frac{25}{16}y_{1t})da$

In particular, since  $\hat{a} = 0.8401$  and  $\hat{\gamma} = -4.1038$  in our case, the systematic deviation of our factor “quick” estimate is given by

$$d\hat{\mu}_t = (0.4148 + 0.1011y_{2t} - 0.5774y_{1t})da.$$

This, for an estimation error of  $da = 0.0401$ , gives

$$d\hat{\mu}_t = 0.0166 + 0.0040y_{2t} - 0.0231y_{1t}.$$

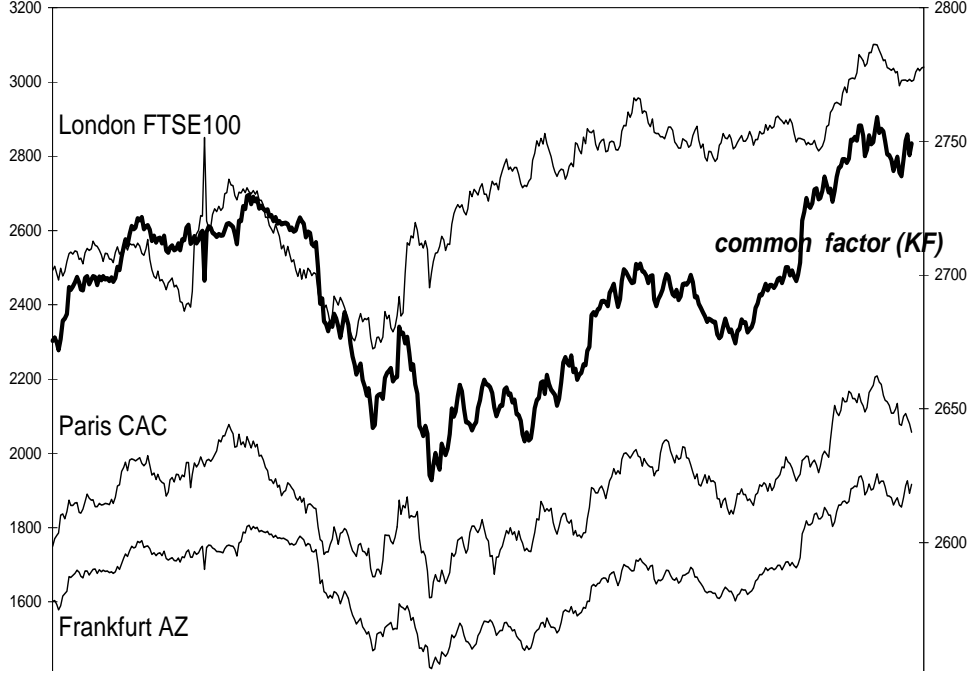
Figure 3 compares these estimation errors for the quick estimate (28) (solid line) with those obtained for the optimal —Kalman filter plus smoother— estimate (dotted line). In both cases they are relatively small: their respective sum of squares being 1.0555 for the quick estimate and 0.7494 for the smoothed one.

## 7.2 European stock exchange data

Figure 4 (thin lines) presents 431 observations of three European stock exchange indices from January 92 to September 93 (excluding nontrading days). Looking at the graph it not only appears that all series are nonstationary, but also that they tend to move together in the long run. The usual unit root tests applied to the logs of the three series come in support of this notion as shown in the following table (figures in **bold** type mean rejection of unit root at the 5% significance level):

	DF test:		Engle-Granger test:
	1st and 2nd unit roots		1 residual unit root
London	-1.124	<b>-22.26</b>	-2.382
Paris	-1.633	<b>-23.18</b>	<b>-4.279</b>
Frankfurt	-0.648	<b>-17.87</b>	<b>-3.919</b>

Figure 4: **Three European stock exchanges: observed indices and common factor**



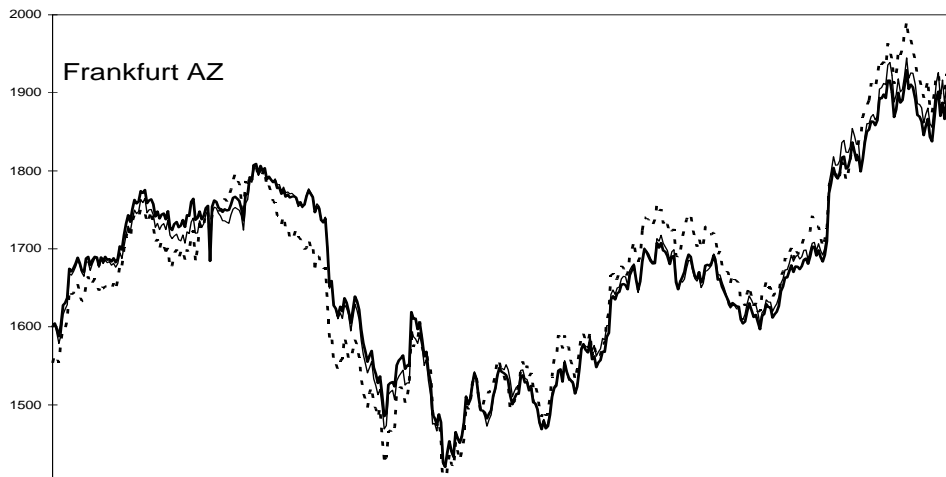
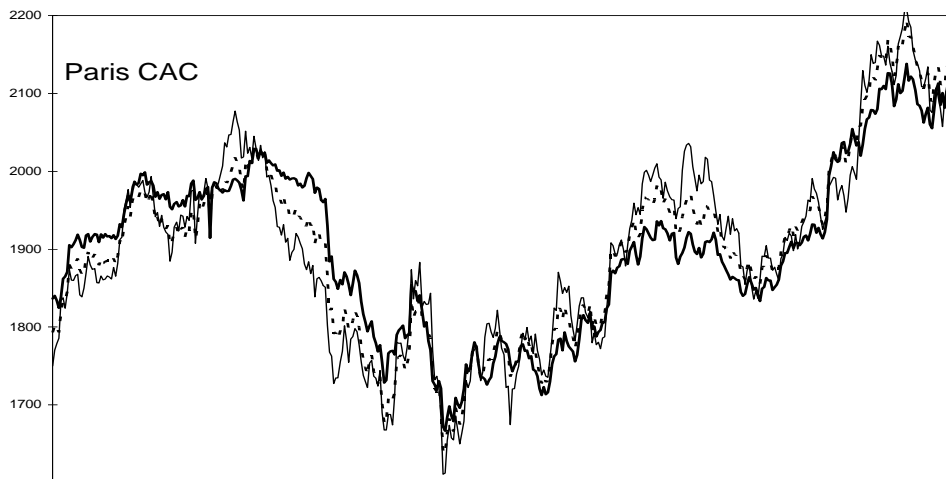
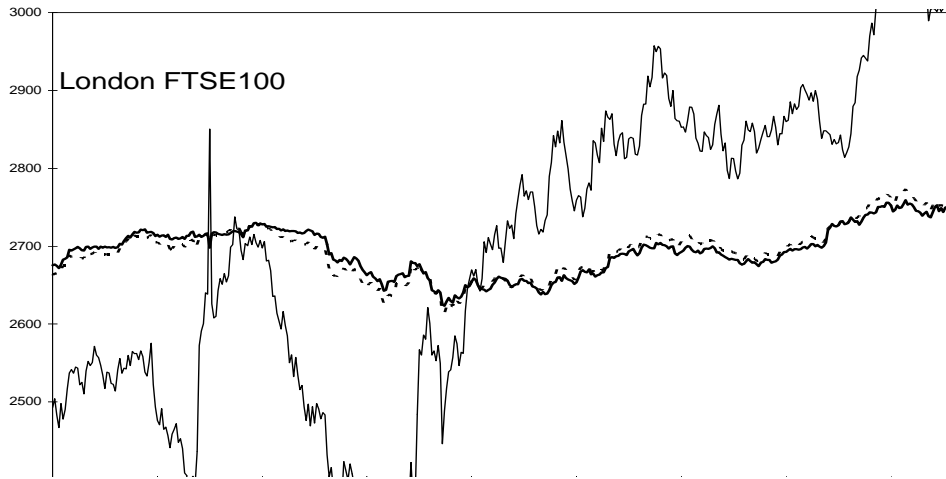
In summary, we may conclude that all three series are  $I(1)$ , but there appears to be two  $I(0)$  linear combinations between them, thus implying the existence of one common factor. Therefore a dynamic factor model (6) with  $n = 3$  and  $k = 1$  was adjusted to the stock exchange data. Maximization of the spectral log-likelihood (21) produced the following results:

$$\begin{aligned}
 y_{1t} \equiv \log(\text{London}_t) &= \mu_t + \varepsilon_{L,t}, \\
 y_{2t} \equiv \log(\text{Paris}_t) &= 4.9468\mu_t - 31.5239 + \varepsilon_{P,t}, \\
 y_{3t} \equiv \log(\text{Frankfurt}_t) &= 6.0779\mu_t - 40.5879 + \varepsilon_{F,t}, \\
 \Delta\mu_t &= 0.8236 \times 10^{-4} + \eta_t,
 \end{aligned} \tag{30}$$

$$\text{Var} \begin{pmatrix} \varepsilon_{L,t} \\ \varepsilon_{P,t} \\ \varepsilon_{F,t} \\ \eta_t \end{pmatrix} = \begin{bmatrix} 17.7022 & 6.0080 & 1.0017 & 0 \\ 6.0080 & 5.9712 & -0.4607 & 0 \\ 1.0017 & -0.4607 & 0.2197 & 0 \\ 0 & 0 & 0 & 0.0208 \end{bmatrix} \times 10^{-4}.$$

The thick line in figure 4 shows the common factor extracted by Kalman filtering and smoothing. On the other hand, for the stock exchange data, the common factor “quick”

Figure 5: European stock exchanges: common factor estimates



estimate given by (27) is

$$\hat{\mu}_t = 6.4512 + 0.0160y_{1t} + 0.0793y_{2t} + 0.0974y_{3t}.$$

Figure 5 compares the observed series with both the optimal smoothed estimate and this quick estimate of the common factor (both scaled in accordance with (30)). Notwithstanding the possibility of factor rotation, we have that the systematic deviation (29) of the “quick” estimate in figure 5 is

$$\begin{aligned} d\hat{\mu}_t = & \frac{(1 - \hat{a}_2^2 + \hat{a}_3^2)(y_{2t} - \hat{\gamma}_2) - 2\hat{a}_2(y_{1t} + \hat{a}_3)(y_{3t} - \hat{\gamma}_3)da_2}{(1 + \hat{a}_2^2 + \hat{a}_3^2)^2} \\ & + \frac{(1 - \hat{a}_3^2 + \hat{a}_2^2)(y_{3t} - \hat{\gamma}_3) - 2\hat{a}_3(y_{1t} + \hat{a}_2)(y_{2t} - \hat{\gamma}_2)da_3}{(1 + \hat{a}_2^2 + \hat{a}_3^2)^2} \end{aligned}$$

where  $da_2, da_3$  are the estimation errors of the factor loads  $a_2, a_3$  respectively. In particular, since  $\hat{a}_2 = 4.9468, \hat{a}_3 = 6.0779$  and  $\hat{\gamma}_2 = -31.5239, \hat{\gamma}_3 = -40.5879$  in our case, the systematic deviation of our “quick” estimate of the common factor is given by

$$\begin{aligned} d\hat{\mu}_t = & (-0.5176 - 0.0025y_{1t} + 0.0034y_{2t} - 0.0154y_{3t})da_2 \\ & + (-0.6062 - 0.0031y_{1t} - 0.0154y_{2t} - 0.0029y_{3t})da_3. \end{aligned}$$

Therefore, since the estimation errors will probably be small due to the length of the sample, we expect the quick estimate not to be far from the optimal one in this case.

## References

- Anderson, T. (1963), ‘The use of factor analysis in the statistical analysis of multiple time series’, *Psychometrika* **28**, 1–25.
- Box, G. & Tiao, G. (1977), ‘A canonical analysis of multiple time series’, *Biometrika* **64**, 355–65. sec. 4.4.
- Brillinger, D. R. (1975), *Time Series: Data Analysis and Theory*, Holt, Rinehart and Wilson, New York.
- Davidson, J., Hendry, D., Srba, F. & Yeo, S. (1978), ‘Econometric modelling of the aggregate time-series relationship between consumers’ expenditure and income in the United Kingdom’, *Economic Journal* **86**, 661–692.
- Dickey, D. A. & Fuller, W. A. (1979), ‘Distribution of the estimators for autoregressive time series with a unit root’, *Journal of the American Statistical Society* **74**, 427–31.

- Dickey, D. A. & Fuller, W. A. (1981), ‘Likelihood ratio statistics for autoregressive time series with a unit root’, *Econometrica* **49**, 1057–72.
- Dunsmuir, W. (1979), ‘A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise’, *The Annals of Statistics* **7**, 490–506.
- Dunsmuir, W. & Hannan, E. (1976), ‘Vector linear time series models’, *Advances Applied Probability* **8**, 339–364.
- Engle, R. & Watson, M. (1981), ‘A one-factor multivariate time series model of metropolitan wage rates’, *Journal of the American Statistical Society* **76**, 774–781.
- Engle, R. F. & Granger, C. W. J. (1987), ‘Cointegration and error correction: representation, estimation and testing’, *Econometrica* **55**, 251–276.
- Fernández-Macho, F., Harvey, A. C. & Stock, J. H. (1987), ‘Forecasting and interpolation using vector autoregressions with common trends’, *Annales d’Economie et de Statistique* **6-7**, 279–287.
- Fernández-Macho, F. J. (1990), ‘Estimation and testing of a multivariate exponential smoothing model’, *Journal of Time Series Analysis* **11**, 89–105.
- Fountis, N. & Dickey, D. (1983), Testing for a unit root nonstationarity in multivariate autoregressive time series, Paper presented at, Statistics: An Appraisal, International Conference to Mark the 50th Anniversary of the Iowa State University Statistical Laboratory.
- Fuller, W. (1985), Nonstationary autoregressive time series, in E. Hannan, P. Krishnaiah & M. Rao, eds, ‘Handbook of Statistics 5: Time Series in the Time Domain’, North-Holland, Amsterdam, pp. 1–23.
- Fuller, W. A. (1976), *Introduction to Statistical Time Series*, John Wiley.
- Geweke, J. (1977), The dynamic factor analysis of economic time-series models, in D. Aigner & A. Goldberger, eds, ‘Latent Variables in Socio-Economic Models’, North-Holland, New York.
- Geweke, J. & Singleton, K. (1981), ‘Maximum likelihood confirmatory factor analysis of economic time series’, *International Economic Review* **22**, 37–54.
- Gonzalo, J. & Granger, C. W. J. (1995), ‘Estimation and common long-memory components in cointegrated systems’, *Journal of Business and Economic Statistics* **13**, 17–35.



- Granger, C. & Engle, R. (1985), Dynamic model specification with equilibrium constraints: Co-integration and error-correction, Discussion Paper 85-18, University of California, San Diego.
- Harvey, A. C. & Todd, P. H. (1983), 'Forecasting economic time series with structural and Box-Jenkins models (with comments)', *Journal of Business and Economic Statistics* **1**, 299–315.
- Hillmer, S. & Tiao, G. (1979), 'Likelihood function of stationary multiple autoregressive moving average models', *Journal of the American Statistical Society* **74**, 652–660.
- Johansen, S. (1988), 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamics and Control* **12**, 231–254.
- Johansen, S. (1991), 'Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models', *Econometrica* **59**(6), 1551–1580.
- Lee, H. S. (1992), 'Maximum likelihood inference on cointegration and seasonal cointegration', *Journal of Econometrics* **54**, 1–47.
- Magnus, J. & Neudecker, H. (1988), *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley, New York.
- Molenaar, P. (1985), 'A dynamic factor model for the analysis of multivariate time series', *Psychometrika* **50**, 181–202.
- Pagan, A. (1975), 'A note on the extraction of components from time series', *Econometrica* **43**, 163–168.
- Priestley, M. & Subba Rao, T. and Tong, H. (1973), Identification of the structure of multivariable stochastic systems, in P. Krishnaiah, ed., 'Multivariate Analysis III', Academic Press, New York, pp. 351–368.
- Priestley, M., Subba Rao, T. & Tong, H. (1974), 'Applications of principal component analysis and factor analysis in the identification of multivariate systems', *IEEE: Transactions in Automatic Control* **19**, 730–734.
- Rao, C. & Mitra, S. (1971), *Generalized Inverse of Matrices and its Applications*, John Wiley and Sons, New York.
- Salmon, M. (1982), 'Error correction mechanisms', *The Economic Journal* **92**, 615–629.

- Sargan, J. (1981), Wages and prices in the UK: a study in econometric methodology, *in* P. Hart, G. Mills & J. Whittaker, eds, ‘Econometric Analysis for National Economic Planning’, Butterworths, London.
- Sargan, J. & Bhargava, A. (1983), ‘Maximum likelihood estimation of regression models with MA(1) errors when the root lies on the unit circle’, *Econometrica* **51**, 799–820.
- Sims, C. (1981), An autoregressive index model for the U.S., 1948-1975, *in* J. Kmenta & J. Ramsey, eds, ‘Large-Scale Macro-Econometric Models’, North Holland, Amsterdam, pp. 283–327.
- Stock, J. (1987), ‘Asymptotic properties of least-squares estimators of cointegrating vectors’, *Econometrica* **55**, 1035–56.
- Stock, J. H. & Watson, M. W. (1988), ‘Testing for common trends’, *Journal of the American Statistical Society* **83**, 1097–1107.
- Subba Rao, T. (1975), ‘An innovation approach to the reduction of the dimensions in a multivariate stochastic system’, *International Journal of Control* **21**, 673–680.
- Velu, R., Reinsel, G. & Wichern, D. (1986), ‘Reduced rank models for multiple time series’, *Biometrika* **73**, 105–118.

## Appendix: Initial estimates.

The autocovariance matrices of the differenced series (10) are

$$\Gamma_z(0) = A\Sigma_\eta A' + 2\Sigma_\varepsilon, \quad \Gamma_z(\pm 1) = -\Sigma_\varepsilon, \quad \Gamma_z(\pm r) = 0, \quad |r| > 1.$$

Thus by the method of moments we obtain

$$\hat{\Sigma}_\varepsilon = -\frac{1}{2}[\hat{\Gamma}_z(1) + \hat{\Gamma}'_z(1)], \quad \hat{\Sigma}_\eta^\dagger = \hat{\Gamma}_z(0) + [\hat{\Gamma}_z(1) + \hat{\Gamma}'_z(1)],$$

where  $\hat{\Sigma}_\eta^\dagger$  will *just approximate*  $\hat{A}\hat{\Sigma}_\eta\hat{A}'$  because in general it will not be of reduced rank  $k < n$ . The problem is then how to obtain  $\hat{A}$ ,  $\hat{\Sigma}_\eta$  satisfying the identification conditions (7)-(8) so that  $\hat{A}\hat{\Sigma}_\eta\hat{A}'$  is of reduced rank  $k < n$ . Since Stock (1987) shows that the least squares estimator of the CI matrix, *i.e.*  $B$  in (9), is consistent we can construct an idempotent matrix  $M = I_n - \hat{B}\hat{B}^+$  which has the property  $M\hat{A} = \hat{A}$ . Then by constructing  $\mathbb{Z} = M\hat{\Sigma}_\eta^\dagger M$ , a symmetric matrix of appropriate reduced rank  $k < n$  is obtained such that

$$\mathbb{Z} = \hat{A}\hat{\Sigma}_\eta\hat{A}'. \tag{31}$$

As, for identifiability,  $A$  and  $\Sigma_\eta$  are restricted so that the former be a truncated unit-lower-triangular matrix and the latter be a diagonal matrix, (31) can be interpreted as a *rank-deficient* Cholesky decomposition.

Strictly speaking a Cholesky decomposition only exists for strict positive definite matrices but, notwithstanding this, an approximate decomposition can be obtained in the following way. Let  $C\Lambda C'$  be the spectral decomposition of matrix  $\mathcal{Z}$ , *i.e.*  $\mathcal{Z} = C\Lambda C'$  where  $\Lambda$  denotes the diagonal matrix of eigenvalues in descending order and  $C$  the matrix of corresponding eigenvectors. Since  $\text{rank } \mathcal{Z} = k < n$  the last  $(n - k)$  eigenvalues must be zero. Let us substitute the zero eigenvalues in  $\Lambda$  by a small but positive number  $h$  and let us call  $\Lambda_h$  the resulting diagonal matrix so that

$$\mathcal{Z}_h = C\Lambda_h C' \quad (32)$$

is a positive definite matrix for which a standard Cholesky decomposition exists. Let

$$\begin{aligned} \mathcal{Z}_h &= L_h D_h L_h' \\ &= \begin{matrix} (k) \\ (n-k) \end{matrix} \begin{bmatrix} L_1 & 0 \\ L_2 & L_3 \end{bmatrix} \begin{matrix} (k) & 0 \\ (n-k) & D_2 \end{matrix} \begin{bmatrix} L_1' & L_2' \\ 0 & L_3' \end{bmatrix} \end{aligned} \quad (33)$$

be such decomposition (note that the eigenvalues in  $D_h$  are in descending order). It is easy to see from (32) that the smaller  $h > 0$  is the closer  $\mathcal{Z}_h$  gets to  $\mathcal{Z}$ , *i.e.*

$$\lim_{h \rightarrow 0} \mathcal{Z}_h = \mathcal{Z}. \quad (34)$$

Similarly in (33) it must be that  $\lim_{h \rightarrow 0} D_2 = 0$ , implying that

$$\lim_{h \rightarrow 0} \mathcal{Z}_h = \lim_{h \rightarrow 0} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} D_1 \begin{bmatrix} L_1' & L_2' \end{bmatrix}. \quad (35)$$

Combining (34) and (35) we have

$$\mathcal{Z} = \lim_{h \rightarrow 0} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} D_1 \begin{bmatrix} L_1' & L_2' \end{bmatrix}.$$

This suggests, by comparison with (31), that  $\hat{A}$  and  $\hat{\Sigma}_\eta$  such that  $\mathcal{Z} = \hat{A}\hat{\Sigma}_\eta\hat{A}'$  can be approximated by the  $(n \times k)$  matrix of  $k$  first columns of  $L_h$  and the  $(k \times k)$  diagonal matrix of  $k$  first columns and rows of  $D_h$ . The approximation depends on the choice of  $h > 0$  and therefore is as accurate as desired (or rather as permitted by the machine accuracy).