

**IDENTIFICATION AND ESTIMATION OF CAUSAL FACTOR MODELS OF STATIONARY TIME  
SERIES**

*Chris Heaton, Dept. of Economics, Macquarie University and School of  
Economics, University of New South Wales;  
Victor Solo, School of Electrical Engineering and Telecommunications, University  
of New South Wales.*

**Abstract**

This paper makes three contributions to the literature on dynamic factor analysis. Firstly, we investigate the identification problem for a general class of causal dynamic factor model and provide conditions under which the model is identified. Secondly, we present an analytical expression for the information matrix of an autoregressive factor model which can be computed far more efficiently than the standard numerical expression, and thirdly we propose an accelerated EM algorithm which has the same convergence properties as the traditional scoring algorithm but has the same storage and CPU-time requirements per iteration as the standard EM algorithm. We illustrate the very significant computational gains over the standard approach with simulations.

## 1. Introduction

The theory of static factor analysis is well developed and has a long history. The early work on the single factor model by Pearson (1901) and Spearman (1904) was generalised to multiple factors by Thurstone (1947), identification theorems were developed by Anderson and Rubin (1956) and Reiersøl (1950), inference results established by Anderson and Rubin (1956) and computationally efficient estimation algorithms developed by Jöreskog (1967) and Lawley and Maxwell (1971). With the subsequent development of cheap computing power, factor analysis became a standard analytical tool in a wide range of physical and social sciences. In the time series context however, the theory of factor analysis remains in a primitive state. The consequence of this is particularly stark in economics where, despite the prevalence of latent variables in theory and the availability of large multivariate data sets, the fact that most economic variables are time series has meant that applications of factor analysis are rare in comparison to other disciplines.

Geweke (1977) proposed a model in which an observable vector is a two-sided filter of a vector of unobservable time series factors, plus coloured noise. This has been applied to an analysis of business cycles by Sargent and Sims (1977), and has been used by Singleton (1980) to model the term structure of interest rates. However, a limitation of this model is that the factors are not restricted to be causal to the observable variables. The most popular approach to time series factor analysis has been to specify the observable vector as the sum of a one-sided filter of a vector of autoregressive factors plus autoregressive noise, write the model in state space form, and base estimation on the Kalman filter. This was originally proposed by Engle and Watson (1981) and Watson and Engle (1983). In their 1981 paper they derive an expression for the elements of the information matrix in terms of only the first derivatives of the filtered state and its covariance. These are evaluated numerically and a scoring algorithm constructed. In their 1983 paper they construct an EM algorithm and compare its performance with the scoring procedure. In independent research, Shumway and Stoffer (1982) also proposed the EM algorithm. The economics literature contains many applications of this model including work on business cycles by Gregory, Head and Raynauld (1997), Gregory and Head (1999), Norrbin and Schlagenhauf (1996), and Lebow (1993), the estimation of business cycle indicators by Stock and Watson (1991) and Stock and Watson (1989), and the estimation of country risk indexes by Melvin and Schlagenhauf (1986). Subsequent research has proposed estimation algorithms for dynamic factor models with different factor specifications. In particular, Kim (1994), Kim and Yoo (1995), Chauvet (1998), Kim and Nelson (1998) and Harris and Martin (1998) have estimated models in which the factor follows Hamilton's (1989) Markov-switching process, and Dungey, Martin and Pagan (2000) have estimated a model in which the factor is autoregressive with GARCH disturbances. However, as far as we know, there has been no work done to further develop the Watson and Engle model. If the goal is to develop a complete theory of factor analysis for time series variables, then most of the work is yet to be done.

In our view there are two main weaknesses of the Watson and Engle approach as it stands. Firstly, and most importantly, little is known about identification in the model. It

is well known that the static factor model, itself a special case of the Watson and Engle model, is unidentified in its unrestricted form. Indeed, the need to place strong restrictions on the model (e.g. zero restrictions) is often regarded as a weakness of factor analysis. It is not known whether generalising the model to include filtered dynamic factors and disturbances creates the need for further identifying restrictions, or whether the dynamic structure itself helps to identify the model. Until this issue is resolved, the interpretation of the results of any estimation procedure is problematic and the development of inference results cannot proceed. The second weakness of the Watson and Engle approach is that it involves a heavy computational burden. Each iteration of the scoring algorithm requires  $r+1$  runs of the Kalman filter and the storage of  $Tp(p+1)$  scalar values where  $T$  is the sample size,  $p$  the number of variables in the model and  $r$  the number of parameters. Thus, the algorithm is expensive in terms of CPU time, and the storage requirements are heavy for even quite small models. Since the storage requirements increase in the order of  $p^3$ , the number of variables that can be feasibly included in the model is restricted, even with the massive improvements in available computing power that have occurred since the publication of their paper. Alternative programming can reduce the storage requirements to  $2Tp(p+1)$ , but requires  $r^2$  passes of the Kalman filter. The iterations of the EM algorithm are not as severely affected by the size of the model, the storage requirements are far more modest, and only a single pass of the Kalman filter and a fixed-interval smoother is needed for each iteration. However, it is often the case that a very large number of iterations are necessary for convergence. From good starting values the EM algorithm will usually take longer to converge than scoring even though it computes each iteration far quicker. In any case, the computation of standard errors still requires the information matrix to be constructed and so the problems of the scoring algorithm are not entirely avoided by using the EM algorithm.

In this paper we make three contributions to the literature on time series factor analysis. Firstly, we discuss the identification problem in the time series setting and prove global identification under certain conditions. The focus of our attention is the Watson and Engle model, but the theorems are sufficiently general to also apply to a range of other factor specifications, including models with switching means and conditionally heteroscedastic factors. Secondly, we derive an analytical expression for the information matrix in the Watson and Engle model that generally requires much less CPU time and storage space to be computed. Thirdly, we propose an accelerated EM algorithm which has the same convergence properties as the scoring algorithm, but computes all but the first iteration in the same time as the standard EM algorithm. We illustrate the computational gains with simulations.

## 2. The Model

Modern economists have at their disposal an enormous and growing number of time series which measure various aspects of economic activity. However, economic theory suggests that these variables are driven by a relatively small, typically unobservable set of influences, e.g. aggregate demand, expectations various prices, willingness to pay, etc. Informal empirical analysis typically supports this view. Eigenvalue decompositions of autocovariance matrices often find a relatively small number of large eigenvalues and a

large number of eigenvalues which are relatively close to zero. This suggests that models with a large number of measured variables, but a relatively small number of important sources of variation might be useful. However, standard multivariate econometric techniques such as VAR models are not well-suited to anything other than small-dimensional data sets, necessitating significant aggregation with the possible loss of information.

We propose the following model.

$$y_t = \xi_t + \varepsilon_t, \quad t = 1, \dots, T \quad (2.1)$$

where  $y_t$  is a  $p$ -vector of observed stationary time series,  $\varepsilon_t$  is an unobserved Gaussian stationary process with diagonal spectral matrix  $S_\omega^\varepsilon$ ,  $\xi_t$  is an unobserved Gaussian stationary signal. Thus, the correlation structure of  $y_t$  is determined solely by  $\xi_t$ . A fairly general model would have (with  $L$  the lag or backshift operator)

$$\xi_t = T(L)\eta_t$$

where  $T(L) = [T_{ij}(L)]$  is a  $p \times k$  matrix of rational transfer functions and  $\eta_t$  is a Gaussian white noise of variance  $\sigma^2 I_k$ . In general we would have  $k=p$ , but in light of the previous discussion, it seems reasonable to allow  $k \ll p$ .

We write

$$T_{ij}(L) = \frac{b_{ij}(L)}{a_{ij}(L)}$$

where  $b_{ij}(L)$  and  $a_{ij}(L)$  (which is monic) are coprime (see Appendix 4) polynomials in  $L$ . In the control theory literature when a state-space representation of a strictly proper (see Appendix 4) transfer function is required, it is assumed that  $\text{degree}(b_{ij}(L)) < \text{degree}(a_{ij}(L))$ . However, such an assumption is not required in what follows so we place no restriction on the degrees of the polynomials other than finiteness. There are a number of routes one might attempt to follow to investigate the identifiability of (2.1). Even if  $\eta_t$  was observed, the identification theory of (2.1) is not straightforward (Hannan and Deistler, 1986). It can, for instance, be developed through minimal dimension state space theory. Because we are not yet able to develop a completely general theory, we take an intermediate route. For  $j = 1, \dots, k$ , let  $d_j(L)$  be the least common multiple of  $a_{ij}(L)$ ,  $i = 1, \dots, p$ . We can then write

$$\left[ T_{1j}(L) \dots T_{pj}(L) \right]^T = \frac{c'_j(L)}{d_j(L)} = \frac{[c_{1j}(L) \dots c_{pj}(L)]}{d_j(L)} \quad (2.2)$$

where  $c_{ij}(L) = \sum_{r=1}^{q_{ij}} c_{ij}^{(r)} L^r$  and  $d_j(L) = 1 + \sum_{r=1}^{m_j} d_j^{(r)} L^r$  are coprime. Thus the factor model can be written as

$$\xi_t = \sum_{j=1}^k \frac{c_j(L)}{d_j(L)} \eta_t^{(j)} \quad (2.3)$$

$$= \begin{bmatrix} c_1(L) & \dots & c_k(L) \\ d_1(L) & \dots & d_k(L) \end{bmatrix} \eta_t = [c_1(L) \dots c_k(L)] \text{diag} \left[ \frac{1}{d_1(L)} \dots \frac{1}{d_k(L)} \right] \eta_t$$

$$= \beta(L) f_t \quad (2.4)$$

$$f_t = \begin{bmatrix} \frac{\eta_t^{(1)}}{d_1(L)} \dots \frac{\eta_t^{(k)}}{d_k(L)} \end{bmatrix}' = A^{-1}(L) \eta_t \quad (2.5)$$

Our model is then (2.1), (2.4) and (2.5). This specification is well-known in the engineering literature (Bennett 1980, Rosenbrock 1973) and, as is shown below, the construction of a state-space representation is straightforward.

In what follows we will refer to  $f_t$  as the factor vector,  $\beta(L)f_t$  as the common component of the model,  $\varepsilon_t$  as the disturbance vector, and to  $\eta_t$  as the factor innovation vector. Denoting  $n_j = \text{degree}(d_j(L))$ , the number of parameters in the common component of the model is  $\sum_{j=1}^k \left( m_j + \sum_{i=1}^p q_{ij} \right)$ .

### 3. Identification

Consider the causal dynamic factor model

$$M: \quad y_t = \beta(L) f_t + \varepsilon_t$$

where  $y_t$  is a  $p$ -vector of observable variables,  $\beta(L)$  is a  $p \times k$  finite-order one-sided or causal matrix polynomial in which  $L$  is the backshift operator,  $f_t$  is a  $k \times 1$  vector of unobservable factors and  $\varepsilon_t$  is a  $p \times 1$  vector of unobservable disturbances. The elements of  $f_t$  and  $\varepsilon_t$  are mutually independent but may be serially correlated. That is, the spectrum of  $f_t$  is diagonal and the spectrum of  $\varepsilon_t$  is diagonal. It is assumed that the autocorrelations of  $f_t$  and  $\varepsilon_t$  are absolutely summable so that their spectra are bounded. The spectrum of the observable variables may be written as

$$S_\omega^y = \beta_\omega S_\omega^f \beta_\omega^H + S_\omega^\varepsilon$$

where  $^H$  denotes the complex conjugate transpose and  $S_{\omega}^f$  and  $S_{\omega}^e$  are the spectra of  $f_t$  and  $\varepsilon_t$  respectively, which again are diagonal. We will consider second-order conditions only and accordingly make the following definition.

*Definition:* We define two factor models  $M$  and  $M^*$  to be observationally equivalent if their spectra  $S_{\omega}^y$  and  $S_{\omega}^{y^*}$  are equal for all frequencies  $\omega$ ,  
 $-\pi \leq \omega \leq \pi$ .

In the case where  $\beta(L)$  is of zero order and  $f_t$  and  $\varepsilon_t$  are Gaussian white noise it is well-known that a rotational indeterminacy exists in the model which may be eliminated by imposing fairly strong restrictions, such as zero-restrictions, on the model (see, for example, Howe, 1955). Two questions are of interest in the causal time series case. Firstly, can identifying zero-restrictions from the static case be generalised to the causal dynamic case, and secondly are there features of the causal dynamic model which allow for identification under weaker restrictions than are generally applied in the static case? Our objective is to find conditions under which the elements of the set of observationally equivalent models differ in only a trivial way such as sign changes or changes in the scaling or order of the factors.

Our first result, a dynamic extension of a static result in Anderson and Rubin (1956), gives conditions under which the process spectrum is identified.

Theorem 1: For the set of models  $M$ , if  $\beta(L)$  can be partitioned into two  $k \times k$  full-rank polynomial matrices, a  $1 \times k$  polynomial vector, and a  $(p-2k-1) \times k$  polynomial matrix, then the disturbance spectrum  $S_{\omega}^e$  is identified.

Proof: See Appendix 1.

Note that the theorem requires that  $p \geq 2k+1$ , providing us with a lower bound on the number of observable variables if our identification results are to apply. However, it is not sufficient to simply have a large number of variables relative to the number of factors; we also have restrictions on the linear independence of the filtering. For example, if it was the case that the data set consisted of time series observations of a panel of firms or individuals who have identical characteristics, then it may be the case that they all react to changes in the common factors in the same way, in which case, the rank of  $\beta(L)$  may be insufficient for the theorem to apply, even if there are a large number of variables. Thus, loosely speaking, theorem 1 says that it is insufficient to have a large number of variables; they must also be sufficiently diverse.

The next result shows that the number of factors is uniquely determined under the conditions of theorem 1.

Theorem 2: For the set of factor models  $M$ , if  $\beta(L)$  can be partitioned as in theorem 1, then the dimension of the factor vector is identified.

Proof: See Appendix 1.

With the disturbance spectrum and factor dimension identified, all that remains is to determine conditions under which  $\beta_\omega$  and  $S_\omega^f$  are uniquely determined by  $\beta_\omega S_\omega^f \beta_\omega^H$ . The following lemma, which is an extension of a static result by Reiersøl (1950), provides a useful representation of the set of observationally equivalent models which will be used subsequently.

Lemma 1: Under the restrictions of theorem 1, the set of observationally equivalent dynamic factor models has the spectral representation

$$S_\omega^y = \beta_\omega^* S_\omega^f \beta_\omega^{*H} + S_\omega^\varepsilon$$

where  $\beta_\omega^* = \beta_\omega M_\omega^{-1}$  and  $S_\omega^{f*} = M_\omega S_\omega^f M_\omega^H$

and  $M_\omega$  is a  $k \times k$  non-singular polynomial operator in  $e^{-i\omega}$ . Furthermore, if  $\beta(L)$  is irreducible (see Appendix 4), then  $M_\omega$  is unimodular.

Proof: See Appendix 1.

We now show that the factor spectrum and the filter  $\beta(L)$  are identified under a particular pattern of zero-restrictions on the factor loading matrix  $\beta(L)$ .

Theorem 3: If  $\beta(L)$  can be partitioned into a  $k \times k$  lower-triangular matrix, a  $k \times k$  full-rank matrix, a  $1 \times k$  row vector and a  $(p-2k-1) \times k$  matrix, then  $\beta(L)$  is identified and the factor spectrum  $S_\omega^f$  is identified up to a re-ordering of the factors, a rescaling of the factors, and a sign change on each factor.

Proof: See Appendix 1.

Note that Theorem 3 does not require  $\beta(L)$  to be irreducible. In fact its proof requires  $M_\omega$  to be non-singular, but not necessarily unimodular. Theorem 3 generalizes to the time series context the well-known result that an appropriate pattern of zero-restrictions identifies a static factor model. Thus, it tells us that identification is no more of a problem in the causal time-series setting than it is in the static case. It states that a  $k$ -factor model is identified if  $k-j$  variables are independent of  $j$  of the factors for  $j=1, \dots, k-1$ . While useful, the factor-exclusion assumptions that the theorem requires are strong and may not be satisfied in many applications. This is often considered a serious drawback of static factor analysis. In theorem 4 we show that under fairly general conditions, the results of theorem 3 hold without zero-restrictions.

Theorem 4: Write the factor-filter as  $\beta(L) = \sum_{r=1}^{\max q_{ij}} \beta_r L^r$ . If

1.  $\beta(L)$  may be partitioned as in theorem 1;
2.  $\text{rank}(\beta_r)=k, r=0, \dots, q$ ;
3.  $\beta(L)$  is irreducible (see Appendix 4);
4. the factor spectra are linearly independent functions, i.e.  
 $\lambda \text{diag}(S_\omega^f) = 0 \forall \omega \in \{0, \dots, \pi\} \Rightarrow \lambda = 0$  for any  $1 \times k$  vector  $\lambda$ ;

then  $\beta(L)$  is identified and the factor spectrum  $S_\omega^f$  is identified up to a re-ordering of the factors, a rescaling of the factors, and sign changes of the factors.

Proof: See Appendix 1.

Suppose that assumption 3 of theorem 4 was not satisfied. In this case, if  $N(L)$  is a  $k \times k$  full rank polynomial matrix common factor, the model  $y_t = \bar{\beta}(L) \bar{f}_t + \varepsilon_t$  where

$\beta(L) = \bar{\beta}(L)N(L)$  and  $\bar{f}_t = N(L)^{-1} f_t$  is observationally equivalent to  $y_t = \beta(L) f_t + \varepsilon_t$ .

Assumption 4 is automatically satisfied in some, but not all cases. The most simple case in which the factor spectra are not linearly independent functions is the static factor model in which all factors are unit-variance white noise. Indeed, any model in which the factors follow exactly the same processes including having the same parameter values, fails to satisfy assumption 4. Furthermore, if the factors are drawn from any set of processes which is closed under addition and scalar multiplication (e.g. ARMA processes), then assumption 4 could be violated even if the factors are all different. Note that we have also introduced an extra rank restriction on  $\beta(L)$ .

It should be noted that the above theorems provide conditions under which the spectra of the factor and disturbance processes are identified. However, our real interest is in the identification of the parameters of these processes. Such identification follows for any process for which the parameters are identified by their spectra. Thus, our theorems provide identification conditions for factor models in which the factors and disturbances follow invertible, irreducible, stationary ARMA processes. This includes the Watson and Engle model, which is the focus of interest in this paper, as a special case. However, in addition to this, our theorems apply to a range of other models. For example, models with factors which follow the Markov-switching process of Hamilton (1989), as estimated by Kim (1994), Kim and Yoo (1995), Chauvet (1998), Kim and Nelson (1998) and Harris and Martin (1998), are identified under the conditions set out above. Similarly, the multi-factor model of bond yields of Dungey, Martin and Pagan (2000), which features factors and disturbances which are autoregressive filters of GARCH processes is also identified under these conditions. However, these theorems do not guarantee identification for models in which the factors and/or the disturbances follow pure GARCH processes or pure stochastic volatility processes since the parameters of such processes are not identified by their unconditional second moments.

#### 4. State Space Representation

While it is possible to write state space representations of factor models with ARMA factors and restricted ARMA disturbances, this involves a much larger state dimension than is necessary for autoregressive factors and disturbances. Since the algorithm that we present in the next section involves a computational load that is related to the state dimension, there are significant computational advantages in restricting attention to models in which both the factors and the disturbances are purely autoregressive. We begin by considering the dynamic factor model with autoregressive factors and white noise disturbances.

White noise disturbances:

$$\begin{aligned} y_t &= \beta(L) f_t + \varepsilon_t \\ f_t &= a(L)^{-1} \eta_t \\ E(\varepsilon_t) &= 0, E(\eta_t) = 0, E(\eta_t \eta_{t-j}') = \delta_j I_k, E(\varepsilon_t \varepsilon_{t-j}') = \delta_j R \\ E(\eta_t \varepsilon_{t-j}') &= 0 \text{ for all } j \end{aligned}$$

where  $\beta(L)$  is a  $p \times k$  FIR operator of order  $q$ ,  $a(L)$  is a  $k \times k$  diagonal FIR operator of order  $m$ ,  $R$  is a  $p \times p$  diagonal matrix, and  $\delta_j$  is Kronecker's delta. A minimum-dimension state-space representation of a scalar autoregressive factor is, for  $i=1, \dots, k$

$$\begin{aligned} f_{it} &= e_1' \tilde{Z}_{it} \\ \tilde{Z}_{it} &= \tilde{A}_i \tilde{Z}_{it-1} + \eta_{it} e_1 \end{aligned}$$

where  $\tilde{Z}_{it} = \begin{pmatrix} z_{it} \\ z_{it-1} \\ \vdots \\ z_{it-m_i+1} \end{pmatrix}$  is the  $m_i \times 1$  state vector,  $\tilde{A}_i = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & \cdots & a_{im_i} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}$

is a so-called companion matrix (Kailath, 1980) and  $e_1$  is an  $m_i \times 1$  vector with 1 in the first element and zeros elsewhere.

Let  $Z_{it} = \begin{pmatrix} z_{it} \\ z_{it-1} \\ \vdots \\ z_{it-\max(q+1, m)} \end{pmatrix}$

If  $q_i+1 < m_i$ , define  $A_i = \tilde{A}_i$  and  $B_i = (B_{i0} \ \cdots \ B_{iq} \ 0 \ \cdots \ 0)$ . Otherwise, let  $B_i = (B_{i0} \ \cdots \ B_{iq})$  and let  $A_i$  be the  $(q_i+1) \times (q_i+1)$  matrix defined as

$$A_i = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{im_i} & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & & 0 & 1 & 0 & & 0 \\ \vdots & & & \vdots & \ddots & & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & 1 & 0 \end{pmatrix}$$

If the vector of observed variables  $y_t$  is modelled in terms of  $k$  independent autoregressive factors, then we may write the model as

$$(3.1) \quad y_t = BZ_t + \varepsilon_t$$

$$(3.2) \quad Z_t = AZ_{t-1} + \eta_t$$

$$\text{where } Z_t = \begin{pmatrix} Z_{1t} \\ \vdots \\ Z_{kt} \end{pmatrix}, A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & A_k \end{pmatrix}, B = (B_1 \quad \cdots \quad B_k), \eta_t = \begin{pmatrix} \eta_{1t}e_1 \\ \vdots \\ \eta_{kt}e_1 \end{pmatrix}.$$

Autoregressive disturbances:

Models with autoregressive disturbances are accommodated by quasi-differencing. Consider the model

$$\begin{aligned} x_t &= \beta(L)f_t + \varepsilon_t \\ \varepsilon_t &= (\varepsilon_{1t} \quad \cdots \quad \varepsilon_{pt})' \\ \varepsilon_{it} &= g_i(L)^{-1} \delta_t \\ f_t &= Af_{t-1} + \eta_t \end{aligned}$$

where  $x_t$  is now the  $p \times 1$  observable vector,  $g_i(L)$  is a monic scalar polynomial of degree  $v_i$ , and  $\delta_t$  is a white noise sequence. Construct the matrices  $P_j = \text{diag}(g_{1j} \quad \cdots \quad g_{pj})$  for  $j = 1, \dots, \max_i(v_i)$  where  $g_{ij}$  is the  $j^{\text{th}}$  coefficient of the polynomial  $g_i(L)$  if  $i \leq v_i$ , or is zero if  $i > v_i$ . Let  $P = [P_1 w_p, \dots, P_v w_p]$  where  $w_p$  is a  $p$ -vector of ones. If we define

$$y_t = x_t - \sum_{j=1}^v P_j x_{t-j} \quad \text{and}$$

$$B = \left[ \left[ \beta, 0_{p \times v} \right] - P_1 \left[ 0_{p \times 1}, \beta, 0_{p \times (v-1)} \right] - P_2 \left[ 0_{p \times 2}, \beta, 0_{p \times (v-2)} \right] - \dots - P_v \left[ 0_{p \times v}, \beta \right] \right]$$

then the model may be written in the above state-space form (3.1) and (3.2). Note that the state dimension of the model is  $\sum_{j=1}^k \max_i q_{ij} + \sum_{i=1}^p v_i$  which is independent of the number of observable variables in the model.

## 5. Parameter Estimation

In their 1983 paper, Watson and Engle suggest estimation of the model by either the EM algorithm or scoring based on numerical differentiation. Both algorithms have attractive and unattractive features. The EM algorithm is robust to poor starting values and, since each iteration requires only a single run of the Kalman filter and a fixed-interval smoother, the iterations can be computed relatively quickly and with modest storage requirements. However, convergence can be extremely slow once the algorithm has moved into the neighbourhood of the maximum. Furthermore, the EM algorithm does not easily produce standard error estimates<sup>1</sup>. On the other hand, the scoring algorithm converges rapidly from good starting values and produces standard error estimates. For this reason, Watson and Engle (1983) wisely recommend that the EM algorithm be used initially to provide starting values for the scoring algorithm. However, the most pressing problem with the scoring algorithm is that it requires the Kalman filter to be executed

$\sum_{j=1}^k \left( m_j + \sum_{i=1}^p q_{ij} \right) + \sum_{i=1}^p v_i$  times in each iteration of the algorithm and for the output to be stored. The computational load of the algorithm, both in terms of CPU time and storage space, is therefore high. Furthermore, the computational cost is increasing in  $p$ , placing severe practical limitations on the number of variables that may be included in the analysis. We propose that this problem be avoided by replacing the scoring algorithm with an accelerated version of the EM algorithm.

Let  $Y = (y_1 \dots y_T)$  and  $Z = (Z_1 \dots Z_T)$ .  $\theta, \theta_0 \in \Theta$  where  $\Theta$  is the parameter space. Following the literature on the EM algorithm (e.g. Dempster, Laird and Rubin, 1977) we define

$$Q(\theta, \theta_0) = E_Z [\ln f(Y, Z | \theta) | Y, \theta_0]$$

where  $\ln f(Y, Z | \theta) = -\frac{T}{2} \ln R - \frac{1}{2} \sum_{t=1}^T (y_t - Z_t)' R^{-1} (y_t - Z_t) - \frac{1}{2} \sum_{j=1}^k \sum_{t=1}^T (e_2' Z_t - a_j Z_{t-1})^2$

and  $a_j$  is a row vector containing the first row of  $A_j$  and  $e_2$  is a  $k \cdot \max(q+1, m_j)$ -vector containing zeros with a 1 in the first entry. Note that  $f(Y, Z | \theta)$  is the likelihood function

<sup>1</sup> In general, estimates of the information matrix can be derived from the EM algorithm. For example, Oakes (1999) shows how the Hessian may be derived from the complete data likelihood. However, in the current application, this would involve differentiation of the Kalman filter and the smoother – a heavier computational burden than the Watson and Engle procedure.

for a model with the same structure as the dynamic factor model except that the factors are assumed to be observed. Note also that the expectation is taken over  $Z$  given  $\theta_0$ , i.e.

$$Q(\theta, \theta_0) = \int_{Z \in \Psi} \ln f(y, Z | \theta) h(Z | Y, \theta_0) dZ$$

where  $h(Z|Y, \theta_0)$  is the likelihood of the state and  $\Psi$  is the state space. Denoting the likelihood of  $Y$  as  $\ell(Y | \theta)$  we have

$$\begin{aligned} \left. \frac{\partial Q(\theta, \theta_0)}{\partial \theta} \right|_{\theta=\theta_0} &= \int_{Z \in \Psi} \left. \frac{\partial f(Y, Z | \theta)}{\partial \theta} \right|_{\theta=\theta_0} \frac{h(Z | Y, \theta_0)}{f(Y, Z | \theta_0)} dZ = \int_{Z \in \Psi} \left. \frac{\partial f(Y, Z | \theta)}{\partial \theta} \right|_{\theta=\theta_0} dZ \frac{1}{\ell(Y | \theta_0)} \\ &= \left. \frac{\partial \ell(Y | \theta)}{\partial \theta} \right|_{\theta=\theta_0} \frac{1}{\ell(Y | \theta_0)} = \left. \frac{\partial \ln \ell(Y | \theta)}{\partial \theta} \right|_{\theta=\theta_0} \end{aligned}$$

Thus, rather than computing the gradient numerically, as in Watson and Engle (1981) we derive the following analytical expressions

$$(5.1) \quad \frac{\partial \ell}{\partial \phi(A)} = \sum_{t=1}^T \text{vec} \left[ E(Z_t Z_t' | y_t, \theta_0) - A E(Z_{t-1} Z_{t-1}' | y_t, \theta_0) \right]$$

$$(5.2) \quad \frac{\partial \ell}{\partial \phi(B)} = \sum_{t=1}^T \text{vec} \left[ R^{-1} y_t E(Z_t | y_t, \theta_0) - R^{-1} B E(Z_t Z_t' | y_t, \theta_0) \right]$$

$$(5.3)$$

$$\frac{\partial \ell}{\partial \text{diag}(R)'} = \frac{1}{2} \sum_{t=1}^T \text{diag} \left[ R^{-1} (y_t y_t' - B y_t E(Z_t | y_t, \theta_0) - E(Z_t | y_t, \theta_0) y_t' B' - B E(Z_t Z_t' | y_t, \theta_0) B') R^{-1} - \text{TR}^{-1} \right]$$

where  $\phi(\cdot)$  is a function which stacks the free parameters of its matrix argument into a column vector and the necessary expectations are computed using a fixed-interval smoother such as that presented in Shumway (1988) and Shumway and Stoffer (1982). This approach to computing gradients has been noted by Fisher (1925), Louis (1982), Meilijson (1989), Ruud (1991) and Oakes (1999). Koopman and Shephard (1992) have also derived this result in the context of a state-space model.

Since the computational load of the smoother is related to the state dimension, whereas the computational load of numerical differentiation is related to the number of parameters in the model, this approach offers significant computational gains. All that remains is to choose a computationally efficient method to approximate the Hessian.

Engle and Watson (1981) propose an element-by-element construction of the information matrix and show that this requires only first derivatives of the state vector and state-covariance matrix which may be carried out numerically. While effective, this approach involves a heavy computational burden, which increases cubically as the number of variables in the model, is increased. Large improvements in computational efficiency may be had if an alternative procedure may be found. Since we have an efficient procedure for constructing the gradient vector, the obvious approach is to use a quasi-Newton method in which the Hessian is approximated by a function of the gradient. After some experimentation, our experience is that the best approach is to start the algorithm by computing the information matrix at the initial parameter vector and to then use the rank-two update of the Davidson-Fletcher-Powell algorithm to approximate the Hessian in subsequent iterations. This requires the information matrix to be computed only twice during the algorithm – once to start the algorithm, and once to compute standard errors – rather than once every iteration.

The information matrix may be computed from the Kalman filter outputs as

$$E \left[ \frac{\partial^2 L}{\partial \phi(\theta_i) \partial \phi(\theta_j)'} \right] = - \sum_{t=1}^T \left\{ \frac{1}{2} \left[ \frac{\partial \text{vec} \Omega_t}{\partial \phi(\theta_i)'} \right]' (\Omega_t^{-1} \otimes \Omega_t^{-1}) \left[ \frac{\partial \text{vec} \Omega_t}{\partial \phi(\theta_j)'} \right] + \left[ \frac{\partial \xi_t}{\partial \phi(\theta_i)'} \right]' \Omega_t^{-1} \left[ \frac{\partial \xi_t}{\partial \phi(\theta_j)'} \right] \right\}$$

where  $\xi_t = y_t - BZ_t^{t-1}$ ,  $\Omega_t = BP_{tt}^{t-1}B' + R$ ,  $Z_t^{t-1} = E_{t-1}(Z_t)$ ,  $P_{tt}^{t-1} = E_{t-1}(Z_t - Z_t^{t-1})(Z_t - Z_t^{t-1})'$ ,  $\theta_i$  and  $\theta_j$  are B, A or R,  $Z_t$  is the state variable, and  $\phi(\cdot)$  is a function which creates a column-vector from the free parameters of its matrix argument. This is a matrix version of the element-by-element expression for the information matrix given as equation (3.12) in Engle and Watson (1981). The term with the Kronecker product in the above expression has dimension  $p^2 \times p^2$ . Thus, for anything other than small values of  $p$  the storage requirements of the algorithm will be high. This problem may be avoided and the information matrix for models with quite large numbers of variables may be feasibly computed by differentiating  $\Omega_t$  with respect to each of the coefficient matrices and multiplying out the terms in the information matrix. The resulting expressions are in appendix 2. While inelegant, they are straightforward to program. Of particular importance is the fact that the necessary derivatives are all of the state dimension, rather than the model dimension, and that the largest Kronecker product terms have dimension of order  $p$  rather than  $p^2$ . The derivative terms could be approximated by numerically differentiating the Kalman filter, however it is usually computationally more efficient to differentiate the Kalman filter analytically to produce derivative filters. These are presented in appendix 3.

Thus, our algorithm is as follows.

1. Estimate the model using the EM algorithm with a coarse convergence criterion to provide starting values for the parameter vector  $\theta_r$ . This will usually require only a handful of iterations.
2. At  $\theta_r$  execute the Kalman filter and the smoother and store the outputs.

3. Compute and invert the information matrix using the expressions in Appendices 2 and 3. Denote the inverted information matrix  $H_r$ .
4. Iterate the following steps:
  - Using the output from the smoother, compute  $E(Z_t Z_t' | Y, \theta_r)$ ,  $E(Z_t Z_{t-1}' | Y, \theta_r)$ , and  $E(Z_t | Y, \theta_r)$  for  $t=1, \dots, T$ ;
  - Evaluate the gradient vector  $g_m$  using expressions (5.1), (5.2) and (5.3);
  - Compute  $\theta_{r+1} = \theta_r + \lambda_r H_r g_r$  for step size  $\lambda_r$ ;
  - Compute the Davidson-Fletcher-Powell update of the Hessian approximation:
 
$$H_{r+1} = H_r + \frac{(\theta_{r+1} - \theta_r)(\theta_{r+1} - \theta_r)'}{(\theta_{r+1} - \theta_r)'(g_{r+1} - g_r)} - \frac{H_r(g_{r+1} - g_r)(g_{r+1} - g_r)'H_r}{(g_{r+1} - g_r)'H_r(g_{r+1} - g_r)}$$
  - At  $\theta_{r+1}$  execute the Kalman filter and the smoother and store the outputs.

Note that this is a quasi-Newton algorithm and so we would expect convergence to occur in fewer iterations than the standard EM algorithm. However, once the initial inverted information matrix has been computed, the computation of the algorithm is the same as for the standard EM algorithm except that the OLS regressions of the ‘M-step’ are replaced by the update of  $H$  and a line-search. If the line search increases the likelihood at its first attempt then the computational load is similar to the standard EM algorithm. Thus, the accelerated EM algorithm combines the superior convergence properties of a Newton-type algorithm with the computational efficiency of the standard EM.

## 6. A Comparison of Computing Times

We have made two as yet unsubstantiated claims for our work in the preceding section – that our expression for the information matrix is generally quicker to compute than the Engle and Watson expression, and that our accelerated EM algorithm converges in approximately the same number of steps as the scoring algorithm. We now verify these claims using simulated data. Our first exercise was to compare the computation times of the Engle and Watson (1981) element-by-element construction of the information matrix with the analytical approach described above. To do this we generated simulated data for models with different orders and compared the CPU time using both methods for the same data. The results are shown in table 1. The computations were made using a Pentium 3 with 256Mb RAM and were programmed in Matlab. For simplicity we have set  $q_{ij}=q$ ,  $m_j=m$  and  $v_l=v$  for all  $i=1, \dots, p$ ,  $j=1, \dots, k$  and  $l=1, \dots, p$ .

Table 1 – Computation times for the information matrix

Model (T=100)	Engle and Watson	Analytical
p=10, q=0, k=1, m=1, v=0	23.45s	1.65s
p=10, q=4, k=1, m=1, v=0	176.58s	7.04s
p=10, q=0, k=4, m=1, v=0	138.36s	4.51s
p=10, q=0, k=1, m=4, v=0	30.42s	4.56s
p=10, q=0, k=2, m=1, v=4	172.8s	13.3s

p=20, q=0, k=1, m=1, v=0	161.7s	1.97s
p=100, q=0, k=1, m=1, v=0	not computed	27.52s

p=the number of observable variables; q=the number of lagged factors; k=the number of factors; m=the order of the factors; v=the order of the disturbances.

We started by computing the information matrix for a simple model with a single AR(1) factor, no lagged factors, and white noise disturbances, using both methods. We then in turn raised one of the orders of the model and recomputed the information matrix. For all the models examined the analytical information matrix was computed much faster than its numerical counterpart. The relative efficiency of the analytical approach is reduced when the order of the factor is raised, but for all other changes of order the relative efficiency is either the same as, or better than, that for the simple model. Most striking is the relative efficiency when the number of observable variables (p) is raised to twenty. In this case, the numerical approach takes over two minutes and forty seconds to compute the information matrix, while the analytical approach requires less than two seconds. We had insufficient RAM to compute the information matrix numerically for the model with p = 100, but were able to compute it analytically in less than thirty seconds. In each case we compared the standard errors implied by each information matrix. We found that the approximation provided by the numerical approach is quite accurate.

The next exercise was to compare the number of iterations to convergence taken by the scoring algorithm and the accelerated EM algorithm. In these simulations we generated 100 data sets for each model, and estimated the model for each data set using both algorithms with the same starting values and the same convergence criterion. The results are in table 2.

Table 2 – The average number of iterations to convergence used by the scoring and accelerated EM algorithms

Model	Scoring	Accelerated EM
P=5, q=0, k=1, m=1, n=0	7.03	7.17
P=5, q=2, k=1, m=1, n=0	8.47	5.99
P=5, q=0, k=1, m=3, n=0	7.3	4.91
P=5, q=0, k=1, m=1, n=1	13.26	10.07
P=5, q=0, k=2, m=1, n=0	11.3	9.8
P=30, q=0, k=1, m=1, n=0	6.48	5.1

The figures in this table are the average over 100 simulated data sets of the number of iterations taken by the scoring algorithm and the number of iterations taken by the accelerated EM algorithm to satisfy the same convergence criterion. Starting values were generated using the standard EM algorithm. The iterations of the standard EM algorithm are not included in the table.

Note that in these simulations, the accelerated EM algorithm generally finds the maximum in slightly fewer iterations than the scoring algorithm. It should be remembered that, with the exception of the first, the iterations of the accelerated EM algorithm require only a single pass of the Kalman filter and a smoother. Thus, models that might take several minutes to estimate using the traditional scoring method are usually estimated in seconds by the accelerated EM algorithm.

## 7. Conclusions

Dynamic factor analysis is potentially a useful analytical tool in economics, but its use has been limited by a lack of structural theory and computationally efficient estimation algorithms. Our work shows that the zero-restrictions that are identifying in the static case generalise to the dynamic model. Furthermore, under fairly general conditions such as linear independence of the factor spectra and irreducibility of the filter, the dynamic factor model is identified without further restrictions. While being quite general, our assumptions are clearly not necessary. For example, in the standard signal extraction problem an observable scalar is a multiple of an unobservable AR(1) factor plus white noise. This model is known to be identified, but its identification is not proved by our theorems. This raises the possibility that future research could uncover a more general set of sufficient conditions for identification. The issue of necessary conditions is another worthwhile topic for future research. We have also presented a maximum likelihood algorithm that is much more efficient than the standard approach in terms of CPU time and storage space. Future research might look at the topic of statistically inefficient estimation approaches which have even greater computational efficiency. Stock and Watson (1998) and Forni, Hallin, Lippi and Reichlin (2000) have taken just such an approach by using principal components to consistently estimate factors in models with infinite dimension. Similarly, simple methods to consistently estimate factors in small models would be extremely useful.

### Appendix 1 – Proofs of Theorems

Proof of Theorem 1: Consider the spectral representation of two dynamic factor models  $M$  and  $M^*$ . Without loss of generality, assume that  $k^* \leq k$ . Under the stated conditions we may write

$$\beta_{\omega} = \begin{pmatrix} \beta_1 \\ \beta_{k+1} \\ \beta_2 \\ \beta_3 \end{pmatrix} \text{ and } \beta_{\omega^*} = \begin{pmatrix} \beta_1^* \\ \beta_{k+1}^* \\ \beta_2^* \\ \beta_3^* \end{pmatrix}$$

for each value of  $\omega$ , where  $\beta_1$  and  $\beta_2$  are full-rank  $k \times k$  matrices,  $\beta_{k+1}$  is a  $1 \times k$  vector,  $\beta_1^*$  and  $\beta_2^*$  are  $k \times k$  matrices and  $\beta_{k+1}^*$  is a  $1 \times k^*$  vector. The subscript  $\omega$  is suppressed in order to simplify the notation.

For each value of  $\omega$  we may write

$$\beta_{\omega} S_{\omega}^f \beta_{\omega}^H = \begin{pmatrix} \beta_1 S_f \beta_1^H & \beta_1 S_f \beta_{k+1}^H & \beta_1 S_f \beta_2^H & \beta_1 S_f \beta_3^H \\ \beta_{k+1} S_f \beta_1^H & \beta_{k+1} S_f \beta_{k+1}^H & \beta_{k+1} S_f \beta_2^H & \beta_{k+1} S_f \beta_3^H \\ \beta_2 S_f \beta_1^H & \beta_2 S_f \beta_{k+1}^H & \beta_2 S_f \beta_2^H & \beta_2 S_f \beta_3^H \\ \beta_3 S_f \beta_1^H & \beta_3 S_f \beta_{k+1}^H & \beta_3 S_f \beta_2^H & \beta_3 S_f \beta_3^H \end{pmatrix}$$

where  $\omega$  is again suppressed, and we may write  $\beta_\omega^* S_\omega^f \beta_\omega^{*H}$  in a similar fashion.

Consider the  $(k+1) \times (k+1)$  submatrix of  $\beta_\omega S_\omega^f \beta_\omega^H$  occupying block rows 1 and 2 and block columns 2 and 3.

$$V = \begin{pmatrix} \beta_1 S_f \beta_{k+1}^H & \beta_1 S_f \beta_2^H \\ \beta_{k+1} S_f \beta_{k+1}^H & \beta_{k+1} S_f \beta_2^H \end{pmatrix}$$

and the corresponding  $(k+1) \times (k+1)$  submatrix of  $\beta_\omega^* S_\omega^f \beta_\omega^{*H}$

$$V^* = \begin{pmatrix} \beta_1^* S_f^* \beta_{k+1}^{*H} & \beta_1^* S_f^* \beta_2^{*H} \\ \beta_{k+1}^* S_f^* \beta_{k+1}^{*H} & \beta_{k+1}^* S_f^* \beta_2^{*H} \end{pmatrix}$$

Since  $\beta_\omega S_\omega^f \beta_\omega^H$  is of rank  $k$ ,  $V$  being  $(k+1) \times (k+1)$  must be singular. Hence,

$$|V| = (-1)^k |\beta_{k+1} S_f \beta_{k+1}^H| |\beta_1 S_f \beta_2^H| + f(\beta, S_f) = 0 \quad (2.1)$$

where  $f(\dots)$  is a bounded, real-valued function of the elements of its matrix arguments.

If  $M$  and  $M^*$  are observationally equivalent then  $S_\omega^y = S_\omega^{y*}$ . Since  $S_\omega^\varepsilon$  and  $S_\omega^{\varepsilon*}$  are diagonal it follows that the off-diagonal elements of  $\beta_\omega^* S_\omega^f \beta_\omega^{*H}$  are equal to the corresponding elements of  $\beta_\omega S_\omega^f \beta_\omega^H$ . Thus,  $\beta_1^* S_f^* \beta_{k+1}^{*H} = \beta_1 S_f \beta_{k+1}^H$ ,  $\beta_1^* S_f^* \beta_2^{*H} = \beta_1 S_f \beta_2^H$  and  $\beta_{k+1}^* S_f^* \beta_2^{*H} = \beta_{k+1} S_f \beta_2^H$ . It follows that

$$V^* = \begin{pmatrix} \beta_1 S_f \beta_{k+1}^H & \beta_1 S_f \beta_2^H \\ \beta_{k+1}^* S_f^* \beta_{k+1}^{*H} & \beta_{k+1} S_f \beta_2^H \end{pmatrix}$$

Similarly, since  $V^*$  is a  $(k+1) \times (k+1)$  submatrix of a matrix of rank  $k^* \leq k$  we have

$$|V^*| = (-1)^k |\beta_{k+1}^* S_f^* \beta_{k+1}^{*H}| |\beta_1 S_f \beta_2^H| + f(\beta, S_f) = 0 \quad (2.2)$$

Since  $\beta_1$ ,  $S_f$ , and  $\beta_2$  are of full rank,  $|\beta_1 S_f \beta_1^H| \neq 0$ . Thus, equations (2.1) and (2.2) yield  $\beta_{k+1}^* S_f^* \beta_{k+1}^{*H} = \beta_{k+1} S_f \beta_{k+1}^H$ .

Similarly, it may be shown that the other diagonal elements of  $\beta_\omega^* S_\omega^f \beta_\omega^{*H}$  are equal to the corresponding elements of  $\beta_\omega S_\omega^f \beta_\omega^H$  for all  $\omega$ . Thus,  $S_\omega^\varepsilon = S_\omega^{\varepsilon*}$  for all  $\omega$ . #

Proof of theorem 2: Consider two factor models  $M$  and  $M^*$ . Without loss of generality, assume that  $k^* < k$ . From theorem 1 we have  $\beta_\omega S_\omega^f \beta_\omega^H = \beta_\omega^* S_\omega^f \beta_\omega^{*H}$ . Let  $\beta_{1\omega}$  be a  $k \times k$

full-rank sub-matrix of  $\beta_\omega$  and  $\beta_{1\omega}^*$  the  $k \times k$  matrix of corresponding rows from  $\beta_\omega^*$ . Then we have  $\beta_{1\omega} S_\omega^f \beta_{1\omega}^H = \beta_{1\omega}^* S_\omega^{f*} \beta_{1\omega}^{*H}$ . Since  $\beta_{1\omega}$  is of full rank  $|\beta_{1\omega}^* S_\omega^{f*} \beta_{1\omega}^{*H}| = |\beta_{1\omega} S_\omega^f \beta_{1\omega}^H| > 0 \Rightarrow k^* \geq k$ , a contradiction. #

Proof of lemma 1: From theorems 1 and 2,  $S_\omega^e$  and  $k$  are identified. Therefore, if  $M$  and  $M^*$  are observationally equivalent, it must be true that

$$\beta_\omega S_\omega^f \beta_\omega^H = \beta_\omega^* S_\omega^{f*} \beta_\omega^{*H}$$

Under the restrictions of theorem 1 we may partition  $\beta_\omega$  into a full-rank  $k \times k$  matrix  $\beta_{1\omega}$ , and a  $(p-k) \times k$  matrix  $\beta_{2\omega}$ , and partition  $\beta_\omega^*$  similarly. Thus, we have

$$\begin{pmatrix} \beta_{1\omega} \\ \beta_{2\omega} \end{pmatrix} S_\omega^f \begin{pmatrix} \beta_{1\omega}^H & \beta_{2\omega}^H \end{pmatrix} = \begin{pmatrix} \beta_{1\omega}^* \\ \beta_{2\omega}^* \end{pmatrix} S_\omega^{f*} \begin{pmatrix} \beta_{1\omega}^{*H} & \beta_{2\omega}^{*H} \end{pmatrix}$$

a set of four matrix equations, two of which are

$$(2.3) \quad \beta_{1\omega} S_\omega^f \beta_{1\omega}^H = \beta_{1\omega}^* S_\omega^{f*} \beta_{1\omega}^{*H}$$

$$(2.4) \quad \beta_{2\omega} S_\omega^f \beta_{1\omega}^H = \beta_{2\omega}^* S_\omega^{f*} \beta_{1\omega}^{*H}$$

from (2.3) we can write

$$(2.3a) \quad S_\omega^{f*} = M_\omega S_\omega^f M_\omega^H \text{ where}$$

$$(2.3b) \quad M_\omega = \beta_{1\omega}^{*-1} \beta_{1\omega}$$

$$(2.3c) \quad \beta_{1\omega}^* = \beta_{1\omega} M_\omega^{-1}$$

and  $S_\omega^f = M_\omega^{-1} S_\omega^{f*} M_\omega^{H-1}$ , from (2.4) we have

$$\beta_{2\omega} M_\omega^{-1} S_\omega^{f*} M_\omega^{-H} \beta_{1\omega}^H = \beta_{2\omega}^* S_\omega^{f*} M_\omega^{-H} \beta_{1\omega}^H$$

so (2.3d)  $\beta_{2\omega}^* = \beta_{2\omega} M_\omega^{-1}$ . Stacking (2.3c) and (2.3d) gives

$$(2.5) \quad \beta_\omega^* = \beta_\omega M_\omega^{-1}$$

Since  $\beta$  and  $\beta^*$  are irreducible and  $\beta_1$  and  $\beta_1^*$  are constructed from rows of  $\beta$  and  $\beta^*$  respectively, it follows that  $\beta$  and  $\beta_1$  are right coprime and that  $\beta^*$  and  $\beta_1^*$  are right coprime. Thus, from the Simple Bezout Identity (Kailath 1980, section 6.3) there exist polynomial matrices  $X_{1\omega}, X_{2\omega}, X_{3\omega}, X_{4\omega}$  such that

$$(2.6) X_{1\omega}\beta_\omega + X_{2\omega}\beta_{1\omega} = I$$

$$(2.7) X_{3\omega}\beta_\omega^* + X_{4\omega}\beta_{1\omega}^* = I$$

Substituting (2.5) and (2.3b) into (2.6) and (2.7) yields  $M_\omega^{-1} = X_{1\omega}\beta_\omega^* + X_{2\omega}\beta_{1\omega}^*$  and  $M_\omega = X_{3\omega}\beta_\omega + X_{4\omega}\beta_{1\omega}$ . Since  $X_{1\omega}, X_{2\omega}, X_{3\omega}, X_{4\omega}, \beta_\omega, \beta_{1\omega}, \beta_\omega^*$ , and  $\beta_{1\omega}^*$  are polynomial matrices,  $M_\omega$  and  $M_\omega^{-1}$  are polynomial matrices. It follows that  $M_\omega$  is unimodular. #

Proof of theorem 3: We define a permutation matrix as a square matrix in which each column and each row has exactly one element with a value of unity. All other elements are zero.

From theorem 2 and part of lemma 1 we need only consider k-factor models of the form

$$S_\omega^y = \beta_\omega^* S_\omega^f \beta_\omega^{*H} + S_\omega^\varepsilon$$

where  $\beta_\omega^* = \beta_\omega M_\omega^{-1}$  and  $S_\omega^{f*} = M_\omega S_\omega^f M_\omega^H$

and  $M_\omega$  is a  $k \times k$  non-singular operator.

Define  $\beta_{1\omega}$  and  $\beta_{1\omega}^*$  as  $k \times k$  lower triangular matrices constructed from rows of  $\beta_\omega$  and  $\beta_\omega^*$  respectively. We may then write  $M_\omega = \beta_{1\omega}^{-1} \beta_{1\omega}^*$ . Since  $\beta_{1\omega}$  is lower triangular so is  $\beta_{1\omega}^{-1}$ . It follows that  $M_\omega$  is also lower triangular. Now consider  $S_\omega^{f*} = M_\omega S_\omega^f M_\omega^H$ . Since  $M_\omega$  is lower triangular and  $S_\omega^f$  and  $S_\omega^{f*}$  diagonal we have  $M_\omega^{-1} S_\omega^{f*} = S_\omega^f M_\omega^H$  where the left-hand side is lower triangular and the right-hand side upper triangular. It follows that  $M_\omega$  is diagonal. Accordingly, we may write  $M_\omega = M_\omega^d \frac{1}{2} M_\omega^u$  where  $M_\omega^d$  is a  $k \times k$  non-singular diagonal operator and  $M_\omega^u$  is a  $k \times k$  unitary operator. We then have  $M_\omega^u S_\omega^f M_\omega^H = M_\omega^d S_\omega^{f*}$ . It follows that the right-hand side is a diagonal matrix of eigenvalues of  $S_\omega^f$  and  $M_\omega^u$  the corresponding matrix of eigenfunctions. Since  $S_\omega^f$  is diagonal it follows that the absolute value of  $M_\omega^u$  is a permutation matrix and  $M_\omega^d$  has absolute value  $I_k$ . It follows that the absolute value of  $M_\omega$  is a permutation matrix. #

Proof of theorem 4: From lemma 1 we may write  $\sum_{j=0}^{\max q_{ij}^*} \beta_j^* e^{-ij\omega} = \sum_{l=0}^s \sum_{j=0}^{\max q_{ij}} \beta_j M_l e^{-i(j+l)\omega}$  for

finite, non-negative  $s$ . Since  $M$  is of full-rank it is non-zero so the transformation of  $\beta_\omega$  cannot reduce its order. Thus,  $\max_{i,j} q_{ij}^* \geq \max_{i,j} q_{ij}$ . The unimodularity of  $M_\omega$  under the

assumed conditions allows us to similarly argue that  $\max_{i,j} q_{ij} \geq \max_{i,j} q_{ij}^*$ . Therefore

$\max_{i,j} q_{ij}^* = \max_{i,j} q_{ij}$  and we may match the terms in the above equation and solve using

assumption 2 to find that  $M_\omega = M$ , a real constant. The factor-spectra of the set of

observationally equivalent models are therefore  $S_{\omega}^{f*} = MS_{\omega}^f M'$ . For this equality to hold, we require  $\text{diag}(S_{\omega}^f)(M_i \circ M_j) = 0$  for  $i \neq j$ , where  $M_i$  is the  $i^{\text{th}}$  row of  $M$  and  $\circ$  is the Hadamard product. Under assumption 3 we therefore have  $(M_i \circ M_j) = 0$  for  $i \neq j$ . Thus,  $MM'$  is diagonal. This allows us to write  $M = M^{\frac{d}{2}} M^o$  where  $M^d$  is a  $k \times k$  non-singular diagonal matrix and  $M^o$  is a  $k \times k$  orthogonal matrix. We then have  $M^o S_{\omega}^f M^{oH} = M^d S_{\omega}^{f*}$ . The diagonality of the right-hand side implies that  $M^d S_{\omega}^{f*}$  is the matrix of eigenvalues of  $S_{\omega}^f$ . Since  $S_{\omega}^f$  is diagonal,  $M^d = I_k$  and  $S_{\omega}^{f*}$  is the spectrum of a permutation of the factor vector  $f_t$ . #

## Appendix 2 – The information matrix

The derivation of the information matrix is described on p.11. The elements are as follows.

$$E \left[ \frac{\partial^2 L_t}{\partial \phi(B) \partial \phi(B)'} \right] = -\frac{1}{2} \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)'} \right]' \left( B' \Omega_t^{-1} B \otimes B' \Omega_t^{-1} B \right) \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)'} \right] - \left( P_{tt}^{t-1} B' \Omega_t^{-1} B P_{tt}^{t-1} \otimes \Omega_t^{-1} \right) \\ - \left( P_{tt}^{t-1} B' \Omega_t^{-1} \otimes \Omega_t^{-1} B P_{tt}^{t-1} \right) K_{ps} - \frac{1}{2} D_t - \frac{1}{2} D_t' - \left[ \frac{\partial \xi_t}{\partial \phi(B)'} \right]' \Omega_t^{-1} \left[ \frac{\partial \xi_t}{\partial \phi(B)'} \right]$$

$$\text{where } D_t = \left( P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes \Omega_t^{-1} B \right) \left( I_{s^2} + K_{ss} \right) \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)'} \right]$$

$$E \left[ \frac{\partial^2 L_t}{\partial (\text{diag} R) \partial (\text{diag} R)'} \right] = -\frac{1}{2} \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)'} \right]' \left( B' \Omega_t^{-1} B \otimes B' \Omega_t^{-1} B \right) \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)'} \right] - \frac{1}{2} \left( \Omega_t^{-1} \circ \Omega_t^{-1} \right) \\ - \frac{1}{2} F_t - \frac{1}{2} F_t' - \left[ \frac{\partial \xi_t}{\partial (\text{diag} R)'} \right]' \Omega_t^{-1} \left[ \frac{\partial \xi_t}{\partial (\text{diag} R)'} \right]$$

$$\text{where } F_t = \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)'} \right]' \left( B' \Omega_t^{-1} \otimes W_s \right) \circ \left( W_s \otimes B' \Omega_t^{-1} \right)$$

$$E \left[ \frac{\partial^2 L_t}{\partial \phi(A) \partial \phi(A)'} \right] = -\frac{1}{2} \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)'} \right]' \left( B' \Omega_t^{-1} B \otimes B' \Omega_t^{-1} B \right) \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)'} \right] - \left[ \frac{\partial \xi_t}{\partial \phi(A)'} \right]' \Omega_t^{-1} \left[ \frac{\partial \xi_t}{\partial \phi(A)'} \right]$$

$$\begin{aligned} E \left[ \frac{\partial^2 L_t}{\partial \phi(B) \partial \phi(A)} \right] &= -\frac{1}{2} \left\{ \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)} \right]' (B' \Omega_t^{-1} B \otimes B' \Omega_t^{-1} B) + (P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes \Omega_t^{-1} B) (I_{s^2} + K_{ss}) \right\} \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)} \right] \\ &\quad - \left[ \frac{\partial \xi_t}{\partial \phi(B)} \right]' \Omega_t^{-1} \left[ \frac{\partial \xi_t}{\partial \phi(A)} \right] \end{aligned}$$

$$\begin{aligned} E \left[ \frac{\partial^2 L_t}{\partial (\text{diag} R) \partial \phi(A)} \right] &= -\frac{1}{2} \left\{ \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)} \right]' (B' \Omega_t^{-1} B \otimes B' \Omega_t^{-1} B) + (\Omega_t^{-1} B \otimes W_s') \circ (W_s' \otimes \Omega_t^{-1} B) \right\} \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)} \right] \\ &\quad - \left[ \frac{\partial \xi_t}{\partial (\text{diag} R)} \right]' \Omega_t^{-1} \left[ \frac{\partial \xi_t}{\partial \phi(A)} \right] \end{aligned}$$

$$\begin{aligned} E \left[ \frac{\partial^2 L_t}{\partial \phi(B) \partial (\text{diag} R)} \right] &= -\frac{1}{2} \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)} \right]' (B' \Omega_t^{-1} B \otimes B' \Omega_t^{-1} B) \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)} \right] \\ &\quad - \frac{1}{2} (P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes \Omega_t^{-1} B) (I_{s^2} + K_{ss}) \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)} \right] - \frac{1}{2} \left[ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)} \right]' \left[ (B' \Omega_t^{-1} \otimes W_s) \circ (W_s \otimes B' \Omega_t^{-1}) \right] \\ &\quad - \frac{1}{2} K_{sp} \left[ (\Omega_t^{-1} \otimes W_s) \circ (W_p \otimes P_{tt}^{t-1} B' \Omega_t^{-1}) \right] - \frac{1}{2} \left[ (P_{tt}^{t-1} B' \Omega_t^{-1} \otimes W_p) \circ (W_s \otimes \Omega_t^{-1}) \right] \\ &\quad - \left[ \frac{\partial \xi_t}{\partial \phi(B)} \right]' \Omega_t^{-1} \left[ \frac{\partial \xi_t}{\partial (\text{diag} R)} \right] \end{aligned}$$

where  $\circ$  is the Hadamard product,  $W_j$  is a  $j \times 1$  vector of ones,  $s$  is the state dimension,  $K_{ij}$  is the commutation matrix for an  $i \times j$  matrix (Magnus and Neudecker, 1991), and  $\phi(\cdot)$  is a function which stacks the free parameters of its matrix argument into a column vector.

If the model includes autoregressive disturbances, then the extra elements of the information matrix are as follows.

$$E \left[ \frac{\partial^2 L_t}{\partial \phi(\theta_i) \partial \phi(\theta_j)'} \right] = \left[ \frac{\partial \text{vec} B}{\partial \phi(\theta_i)'} \right]' E \left[ \frac{\partial^2 L_t}{\partial \phi(B) \partial \phi(B)'} \right] \left[ \frac{\partial \text{vec} B}{\partial \phi(\theta_j)'} \right] \text{ for } \theta_i, \theta_j = \beta, P;$$

$$E \left[ \frac{\partial^2 L_t}{\partial \phi(\theta_i) \partial \phi(\theta_j)'} \right] = \left[ \frac{\partial \text{vec} B}{\partial \phi(\theta_i)'} \right]' E \left[ \frac{\partial^2 L_t}{\partial \phi(B) \partial \phi(\theta_j)'} \right] \text{ for } \theta_i = \beta, P, \theta_j = A, R.$$

We may write  $\phi(B) = \sum_{j=0}^v (I_{s+v} \otimes P_j) L_j \phi(\beta)$  where  $L_j = \begin{pmatrix} 0_{(j-1) \times p} \\ I_p \\ 0_{(v-j) \times p} \end{pmatrix}$  and  $P_0 = I_p$  and derive

$$\frac{\partial \phi(B)}{\partial \phi(\beta)'} = \sum_{j=0}^v (I_{s+v} \otimes P_j) L_j \text{ and } \frac{\partial \phi(B)}{\partial P_j'} = \begin{pmatrix} 0_{p \times p} \\ \text{diag}(\beta_j) \\ \vdots \\ \text{diag}(\beta_j) \\ 0_{p(v-j) \times p} \end{pmatrix}$$

Where  $\text{diag}(\beta_j)$  is a  $p \times p$  diagonal matrix in which the diagonal is the  $j^{\text{th}}$  column of  $\beta$ .

### Appendix 3 – Derivatives of the Kalman filter

Consider the state space model  $y_t = BZ_t + \varepsilon_t$ ,  $Z_t = AZ_t + \eta_t$  where  $y_t$  is the  $p \times 1$  observable vector,  $Z_t$  the  $s \times 1$  state vector,  $E(\varepsilon_t \varepsilon_t') = R$ ,  $E(\eta_t \eta_t') = Q$ . The Kalman filter is then

$$\begin{aligned} Z_t^{t-1} &= AZ_{t-1}^{t-1} \\ Z_t^t &= Z_t^{t-1} + K_t \xi_t \\ K_t &= P_t^{t-1} B' \Omega_t^{-1} \\ P_t^{t-1} &= A P_{t-1, t-1}^{t-1} A' + Q \\ P_t^t &= P_t^{t-1} - K_t B P_t^{t-1} \\ \Omega_t &= B P_t^{t-1} B' + R \\ \xi_t &= y_t - B Z_t^{t-1} \end{aligned}$$

where  $Z_j^i = E_i(Z_j)$  and  $P_{jj}^i = E_i(Z_j - E_i(Z_j))(Z_j - E_i(Z_j))'$ . The starting value for the expected state is  $Z_0^0 = 0$ . In principle the starting value for the state covariance  $P_{00}^0$  should be found by solving the Lyapunov equation. However, in practice we have found starting from the identity matrix to work better. By differentiating we derive the following filters.

Derivatives with respect to A:

$$\begin{aligned}\frac{\partial Z_t^{t-1}}{\partial \phi(A)'} &= \left( Z_{t-1}^{t-1} \otimes I_s \right) + A \frac{\partial Z_{t-1}^{t-1}}{\partial \phi(A)'} \\ \frac{\partial Z_t^t}{\partial \phi(A)'} &= (I_s - K_t B) \frac{\partial Z_t^{t-1}}{\partial \phi(A)'} + \left[ \xi' \Omega_t^{-1} B \otimes (I_s - P_{tt}^{t-1} B' \Omega_t^{-1} B) \right] \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)'} \\ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)'} &= (I_{s^2} + K_{ss}) \left( A P_{t-1}^{t-1} \otimes I_s \right) + (A \otimes A) \frac{\partial \text{vec} P_{t-1}^{t-1}}{\partial \phi(A)'} \\ \frac{\partial \text{vec} P_{tt}^t}{\partial \phi(A)'} &= \left\{ I_{s^2} - (I_s - K_t B) - \left[ P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes (I_s - P_{tt}^{t-1} B' \Omega_t^{-1} B) \right] \right\} \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)'} \\ \frac{\partial \xi_t}{\partial \phi(A)'} &= -B \frac{\partial Z_t^{t-1}}{\partial \phi(A)'} \\ \frac{\partial \Omega_t}{\partial \phi(A)'} &= (B \otimes B) \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(A)'}\end{aligned}$$

Derivatives with respect to R:

$$\begin{aligned}\frac{\partial Z_t^{t-1}}{\partial (\text{diag} R)'} &= A \frac{\partial Z_{t-1}^{t-1}}{\partial (\text{diag} R)'} \\ \frac{\partial Z_t^t}{\partial (\text{diag} R)'} &= (I_s - K_t B) \frac{\partial Z_t^{t-1}}{\partial (\text{diag} R)'} + \left( \xi' \Omega_t^{-1} B \otimes I_s \right) \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)'} - \left[ \left( \xi' \Omega_t^{-1} \otimes W_s \right) \circ P_{tt}^{t-1} B' \Omega_t^{-1} \right] \\ \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)'} &= (A \otimes A) \frac{\partial \text{vec} P_{t-1}^{t-1}}{\partial (\text{diag} R)'} \\ \frac{\partial \text{vec} P_{tt}^t}{\partial (\text{diag} R)'} &= \left\{ I_{s^2} - (I_s - K_t B) - \left[ P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes (I_s - P_{tt}^{t-1} B' \Omega_t^{-1} B) \right] \right\} \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)'} \\ &\quad + \left( P_{tt}^{t-1} B' \Omega_t^{-1} \otimes W_s \right) \circ \left( W_s \otimes P_{tt}^{t-1} B' \Omega_t^{-1} \right) \\ \frac{\partial \xi_t}{\partial (\text{diag} R)'} &= -B \frac{\partial Z_t^{t-1}}{\partial (\text{diag} R)'} \\ \frac{\partial \Omega_t}{\partial (\text{diag} R)'} &= (B \otimes B) \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial (\text{diag} R)'} + \text{diag}[\text{vec}(I_p)] \left( W_p \otimes I_p \right)\end{aligned}$$

Derivatives with respect to B:

$$\begin{aligned}
\frac{\partial Z_t^{t-1}}{\partial \phi(B)'} &= A \frac{\partial Z_{t-1}^{t-1}}{\partial \phi(B)'} \\
\frac{\partial Z_t^t}{\partial \phi(B)'} &= (I_s - K_t B) \frac{\partial Z_t^{t-1}}{\partial \phi(B)'} + \left[ (\xi' \Omega_t^{-1} B \otimes I_s) - (\xi' \Omega_t^{-1} B \otimes P_{tt}^{t-1} B' \Omega_t^{-1} B) \right] \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)'} \\
&\quad + \left[ \xi' \Omega_t^{-1} \otimes (P_{tt}^{t-1} - P_{tt}^{t-1} B' \Omega_t^{-1} B P_{tt}^{t-1}) \right] K_{sp} - (\xi' \Omega_t^{-1} B P_{tt}^{t-1} \otimes P_{tt}^{t-1} B' \Omega_t^{-1} B) - \left( Z_t^{t-1} \otimes K_t \right) \\
\frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)'} &= (A \otimes A) \frac{\partial \text{vec} P_{t-1}^{t-1}}{\partial \phi(B)'} \\
\frac{\partial \text{vec} P_{tt}^t}{\partial \phi(B)'} &= \left[ I_{s^2} - (I_s \otimes K_t B) - (P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes I_s) - (P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes P_{tt}^{t-1} B' \Omega_t^{-1} B) \right] \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)'} \\
&\quad - \left[ P_{tt}^{t-1} B' \Omega_t^{-1} \otimes (P_{tt}^{t-1} - P_{tt}^{t-1} B' \Omega_t^{-1} B P_{tt}^{t-1}) \right] K_{sp} + (P_{tt}^{t-1} B' \Omega_t^{-1} B \otimes P_{tt}^{t-1} B' \Omega_t^{-1} B) - (P_{tt}^{t-1} \otimes K_t) \\
\frac{\partial \xi_t}{\partial \phi(B)'} &= \left( Z_t^{t-1} \otimes I_p \right) - B \frac{\partial Z_t^{t-1}}{\partial \phi(B)'} \\
\frac{\partial \text{vec} \Omega_t}{\partial \phi(B)'} &= (B P_{tt}^{t-1} \otimes I_p) + (I_p \otimes B P_{tt}^{t-1}) K_{ps} + (B \otimes B) \frac{\partial \text{vec} P_{tt}^{t-1}}{\partial \phi(B)'}
\end{aligned}$$

where  $K_{ij}$  is the  $ij$  commutation matrix, but  $K_t$  always refers to the Kalman gain matrix.

Starting values for the filters are  $\frac{\partial Z_0^0}{\partial \phi(\theta)'} = 0$  for  $\theta = B, A, R$ ,  $\frac{\partial \text{vec} P_{00}^0}{\partial \phi(\theta)'} = 0$  for  $\theta = B, A, R$ .

#### Appendix 4

Definitions:

- A scalar polynomial is said to be *monic* if and only if its leading term has a value of unity.
- A polynomial matrix (i.e. a matrix of polynomials) is said to be *unimodular* if and only if its inverse is also a polynomial matrix. It follows that a polynomial matrix is unimodular if and only if its determinant is a non-zero constant.
- If  $P(L)$  is a  $p \times k$  rational transfer function matrix (i.e. a matrix of rational transfer functions), and if  $P(L) = \bar{P}(L) D_R(L)^{-1}$  where  $\bar{P}(L)$  and  $D_R(L)$  are  $p \times k$  and  $k \times k$  rational transfer function matrices respectively, then  $D_R(L)$  is a right divisor of  $P(L)$ . If  $P(L) = D_L(L)^{-1} \bar{P}(L)$  where  $D_L(L)$  is a  $k \times k$  rational transfer function matrix, then  $D_L(L)$  is a left divisor of  $P(L)$ .

- Two matrix polynomials  $P(L)$  and  $Q(L)$  of dimensions  $p \times k$  and  $q \times k$  respectively, are said to be (left or right) *coprime* if and only if they only have unimodular common (left or right) divisors.
- A rational transfer function matrix  $P(L)$  is said to be *proper* if  $\lim_{L \rightarrow \infty} P(L) < \infty$  and *strictly proper* if  $\lim_{L \rightarrow \infty} P(L) = 0$ .
- A polynomial matrix of full column rank is said to be *irreducible* if and only if its rows are right coprime.

## References

- Anderson, T. W. and Rubin, H. (1956) "Statistical inference in factor analysis" *Third Berkeley Symposium of Mathematical Statistics and Probability* 5, p.111-150.
- Barnett, S. (1980) *Matrices in Control Theory* Van Nostrand Reinhold Company: New York.
- Chauvet, M. (1998) "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching" *International Economic Review* 39(4), November, p.969-96..
- Connor G. and Korajczyk R.A. (1986) "Performance Measurement with the Arbitrage Pricing Theory: A new Framework for Analysis" *Journal of Financial Economics* 15, p.373-94.
- Dungey, M., Martin, V. L. and Pagan, A. R. (2000) "A Multivariate Latent Factor Decomposition of International Bond Yield Spreads" *Journal of Applied Econometrics* 15, p.697-715.
- 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(376), p.774-81.
- Fisher, R. A. (1925) "Theory of Statistical Estimation" *Proceedings of the Cambridge Philosophical Society* 22, p.700-725.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000) "The Generalized Dynamic Factor Model: Identification and Estimation" *The Review of Economics and Statistics* 82(4), p.540-554.
- Geweke J. (1977) "The Dynamic Factor Analysis of Economic Time-Series Models" in: D.J. Aigner and A.S. Goldberger (eds) *Latent Variables in Socioeconomic Models* North Holland: Amsterdam.
- Geweke, J. F. and K. J. Singleton (1981). "Maximum Likelihood 'Confirmatory' Factor Analysis of Economic Time Series." *International Economic Review* 22(1): 37-54.
- Gregory, A.W. and Head, A. C., (1999) "Common and country-specific fluctuations in productivity, investment, and the current account" *Journal Of Monetary Economics* Vol. 44 (3) pp. 423-451
- Gregory A., Head, A. and Raynauld, J. (1997) "Measuring World Business Cycles" *International-Economic-Review* 38(3), August, p.677-701.

- Hamilton, J. D., (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle" *Econometrica* 57(2), March, p.357-84..
- Harris, D. and Martin, V. L. (1998) "Indirect Estimation of Dynamic Factor Models of the Business Cycle" mimeo, University of Melbourne.
- Howe, W. G. (1995) *Some Contributions to Factor Analysis* Report No. ORNL-1919, Oak Ridge National Laboratory, Oak Ridge Tennessee.
- Jöreskog, K. G. (1967). "Some Contributions to Maximum Likelihood Factor Analysis." *Psychometrika* 32: 443-482.
- Kailath T. (1980) *Linear Systems* Prentice-Hall: New Jersey
- Kim, C. J. and Nelson, C. R. (1998) "Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching" *Review of Economics and Statistics* 80(2), May, p.88-201..
- Koopman, S. J. and Shephard, N. (1992) "Exact Scores for Time Series Models in State Space Form" *Biometrika* 79, p.823-826.
- Lawley, D. N. and A. E. Maxwell (1971). *Factor Analysis as a Statistical Method*. Elsevier: New York.
- Lebow, D. E., (1993) "The Covariability of Productivity Shocks across Industries" *Journal of Macroeconomics* 15(3), Summer, p.483-510.
- Louis, T. A. (1982) "Finding the Observed Information Matrix When Using the EM Algorithm" *Journal of the Royal Statistical Society B* 44, p.226-242.
- Magnus, J. R. and Neudecker, H. (1991) *Matrix Differential Calculus with Applications in Statistics and Econometrics* John Wiley and Sons: Chichester.
- Meilijson, I. (1989) "A Fast Improvement to the EM Algorithm on its Own Terms" *Journal of the Royal Statistical Society B* 51(1), p.127-138.
- Melvin, M. and Schlagenhauf, D. (1986) "Risk in International Lending: A Dynamic Factor Analysis Applied to France and Mexico" *Journal of International Money and Finance* 5, p.S31-S48.
- Norrbin, S. C. and Schlagenhauf, D. E., (1996) "The Role of International Factors in the Business Cycle: A Multi-country Study" *Journal of International Economics* 40(1-2), February, p.85-104.
- Oakes, D. (1999) "Direct Calculation of the Information Matrix via the EM Algorithm" *Journal of the Royal Statistical Society B* 61(2), p.479-482.

- Pearson, K. (1901) "On lines and planes of closest fit to a system of points in space" *Philosophical Magazine* 2, 6<sup>th</sup> series, p.557-572.
- Reiersøl, O. (1950) "On the identifiability of parameters in Thurstone's multiple factor analysis" *Psychometrika* 15, p.121-49.
- Rosenbrock, H. H. (1970) *State-space and multivariable theory* Wiley: New York.
- Ruud, P. A. (1991) "Extensions of Estimation Methods using the EM Algorithm" *Journal of Econometrics*, 49(3), p.305-401.
- Sargent, T. and Sims, C. (1977) "Business cycle modelling without pretending to have too much a priori economic theory" in Sims, C. (ed.) *New Methods in Business Cycle Research* Minneapolis: Federal Reserve Bank of Minneapolis.
- Shumway, R. H. (1988) *Applied Statistical Time Series Analysis* Englewood Cliffs, NJ: Prentice-Hall.
- 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(4), p.253-264.
- Singleton, K. J. (1980) "A latent time series model of the cyclical behaviour of interest rates" *International Economic Review* 21(3), October, p.559-575.
- Spearman, C. (1904) "General Intelligence Objectively Determined and Measured" *American Journal of Psychology*, 15, p.201-293.
- Stock J.H. and Watson M.W. (1998) "Diffusion Indexes" *NBER Working Paper W6702*
- Stock, J. and Watson, M., (1989) "New Indexes of Coincident and Leading Economic Indicators" *Macroeconomics Annual* Vol. 4, M.I.T. Press.
- Stock, J. and Watson, M., (1991) "A Probability Model of the Coincident Economic Indicators" in K. Lahiri and G. Moore (eds) *Leading Economic Indicators: New Approaches and Forecasting Records* CUP.
- Thurstone, L. L. (1947) *Multiple Factor Analysis* Chicago: University of Chicago Press.
- Watson, M. W. and R. F. Engle (1983). "Alternative Algorithms for Estimation of Dynamic MIMIC, Factor, and Time Varying Coefficient Regression Models." *Journal of Econometrics* 23: 385-400.