TI 2009-010/4
Tinbergen Institute Discussion Paper

Dynamic Factor Analysis in the Presence of Missing Data

B. Jungbacker[1]
S.J. Koopman[1,2]
M. van der Wel[2,3,4]

1 VU University Amsterdam;
[2] Tinbergen Institute;
[3] Erasmus University Rotterdam;
[4] CREATES.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Dynamic factor analysis in the presence of missing data

B. Jungbacker[(a)]*, *S. J. Koopman*[(a,b)] *and* *M. van der Wel*[(c)]

(a) Department of Econometrics, VU University Amsterdam

(b) Tinbergen Institute, Amsterdam

(c) Econometric Institute, Erasmus University Rotterdam and CREATES

February 6, 2009

Abstract

We develop a new model representation for high-dimensional dynamic multi-factor models. It allows the Kalman filter and related smoothing methods to produce optimal estimates in a computationally efficient way in the presence of missing data. We discuss the model in detail together with the implementation of methods for signal extraction and parameter estimation. The computational gains of the new devices are presented based on simulated data-sets with varying numbers of missing entries.

JEL classification: C33; C43.

Some keywords: High-dimensional vector series; Kalman Filter; Maximum likelihood.

1 Introduction

This paper is concerned with factor extraction and parameter estimation procedures for high-dimensional multi-factor models in the presence of missing data. We develop the methodology for the basic dynamic factor model

$$y_t = \Lambda f_t + u_t, \qquad t = 1, \ldots, n, \tag{1}$$

where y_t is the $N \times 1$ vector of time series observations, f_t is an unobserved $q \times 1$ vector of common factors and u_t is the $N \times 1$ vector of idiosyncratic terms. We focus on cases where N is significantly larger than q. The factors are assumed to follow a Gaussian dynamic

*Corresponding author: B.Jungbacker, Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands. Email: bjungbacker@feweb.vu.nl

linear process and the idiosyncratic components in u_t are modelled as autoregressive (AR) processes. The results are also applicable for more general models of this form. We will discuss these issues in detail.

In many applications, the dimension of y_t is large and the model depends on a large number of parameters. The task of signal extraction and parameter estimation is therefore challenging in various respects. Particularly, state space formulations allow us to obtain minimum mean square estimates of the factors together with the corresponding mean square errors by means of the Kalman filter and smoother recursions. The methods can be implemented in a computationally efficient way. However, these model representations are not valid in the presence of missing data. We address this problem by developing a low-dimensional linear state space model with time-varying state dimensions. It is equivalent to the dynamic factor model (1) and is designed to allow for missing entries in the dataset. We also discuss how the state space formulation can be used to obtain maximum likelihood estimates of the parameters by means of the Expectation-Maximization (EM) algorithm of Dempster, Laird, and Rubin (1977) and by direct optimization using a Quasi-Newton scheme.

Traditionally models of the form (1) have been estimated using principal components (PCA), see Forni, Hallin, Lippi, and Reichlin (2000), Bai (2003) and Stock and Watson (2002a). Stock and Watson (2002b) is especially relevant as they show how missing data can be handled in a principal components analysis by means of an EM type algorithm. Sargent and Sims (1977) and Geweke (1977) are the earliest references discussing maximum likelihood methods for dynamic factor models. For a relatively low-dimensional model for wage rates, Engle and Watson (1981) consider maximum likelihood estimation, using Fisher scoring to maximize the likelihood. EM algorithms are developed for maximum likelihood estimation of parameters in state space models by Watson and Engle (1983) and Shumway and Stoffer (1982).

Recently there has been a renewed interest in the use of maximum likelihood estimation for high dimensional models. Doz, Giannone, and Reichlin (2006) show that estimates of the unobserved factors obtained from a likelihood-based analysis are consistent estimators of $f_1, \ldots, f_n$ as $n, N \to \infty$, even if the dynamic factor model is misspecified. Furthermore, they present evidence that in some cases a likelihood-based analysis produces more precise estimates of the factors than a principal component method.

Reis and Watson (2007) consider the dynamic factor model (1) and estimate the parameters by maximum likelihood using the approach of Watson and Engle (1983). This approach is not applicable when missing data is present. Banbura and Modugno (2008) propose a solution that overcomes the problem but is computationally demanding. Furthermore, their method requires numerical modifications to let the methods work in a satisfactory way. These modifications may distort the purpose of finding exact maximum likelihood estimates. In

this paper we address the same issue but we provide a computationally efficient method that leads to exact maximum likelihood parameter estimates. All methods provide minimum mean square estimates and corresponding mean square errors of the factors and the idiosyncratic components.

The remainder of the paper is organised as follows. The dynamic factor model and its state space representations are discussed in detail in section 2. We develop in section 3 a new representation of the model that is valid when missing data is present. This state space representation allows the computationally efficient application of the Kalman filter and smoother recursions. Signal extraction and likelihood evaluation are explored in section 4. Parameter estimation by maximum likelihood methods are discussed in section 5 while computational comparisons based on simulated data is presented in section 6. A short discussion of the presented results is given in section 7.

2 The dynamic factor model

2.1 Model specification

The dynamic factor model given in (1) links the observation y_t to a set of unobserved factors f_t for $t = 1, \ldots, n$. We assume that $f_1, \ldots, f_n$ are linear combinations of an unobserved $p \times 1$ dimensional vector autoregressive process α_t. Specifically, there is a $q \times p$ selection matrix S that defines the dynamic factor as

$$f_t = S\alpha_t, \tag{2}$$

and there is a transition equation for the state vector α_t as given by

$$\alpha_{t+1} = T\alpha_t + \eta_t, \qquad \eta_t \sim N(0, \Sigma_\eta), \tag{3}$$

for $t = 1, \ldots, n$ with the initial state vector α_1 specified by $\alpha_1 \sim N(0, \Sigma_\alpha)$ and where the $p \times p$ transition matrix T and the $p \times p$ variance matrix Σ_η are assumed fixed (non-stochastic). The matrix S is treated as a known selection matrix while matrices T and Σ_η may depend on a fixed and unknown vector of coefficients θ. In case α_t is a time-invariant stationary process, the relation $\Sigma_\alpha = T\Sigma_\alpha T' + \Sigma_\eta$ applies such that a solution for the (initial) state variance Σ_α exists when matrices T and Σ_η are given.

It follows that the dynamic factor model (1) can be expressed in terms of the state vector

$$y_t = Z\alpha_t + u_t, \tag{4}$$

where $Z = \Lambda S$. The factor loading matrix Λ is treated as fixed and it depends on coefficient

vector θ. The idiosyncratic component u_t is modelled as a vector autoregressive process with r lags as given by

$$u_{t+1} = \phi_1 u_t + \cdots + \phi_r u_{t-r+1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma_\varepsilon), \tag{5}$$

where $\phi_1, \ldots, \phi_r$ and Σ_ε are $N \times N$ matrices and the initial vector u_1 is specified as $u_1 \sim N(0, \Sigma_u)$. In general $\phi_1, \ldots, \phi_r$ will be chosen such that u_t is a stationary process and Σ_u will be set to the stationary variance of u_t. In case of $r = 1$, the solution is implied by $\Sigma_u = \phi_1 \Sigma_u \phi_1' + \Sigma_\varepsilon$.

In the remainder of the paper, we consider the dynamic factor model as specified above. However, our results below apply to more general settings. We discuss these generalizations in some detail in section 3.3.

2.2 Two state space representations

The dynamic factor model specification (4), (3) and (5) is close to the well-known state space model formulations of Harvey (1989) and Durbin and Koopman (2001). The Kalman filter and smoothing methods produce estimates of the state vector α_t with minimum mean square linear properties. In our formulation of a linear Gaussian dynamic factor model, we obtain minimum mean square estimates, see the discussions in Duncan and Horn (1972) and Anderson and Moore (1979). However, these optimal properties only apply when the observation equation (4) has disturbances u_t that are not serially correlated. We can re-formulate the dynamic factor model in two ways to ensure that the optimal properties of the Kalman filter and smoothing methods are preserved. The two formulations are given as A and B below for the special case of $r = 1$. The higher-order case of $r > 1$ follows straightforwardly but is notationally more cumbersome.

A. A basic approach is to express the dynamic factor model in terms of $(1 - \phi_1 L) y_t$ where L is the lag-operator. When the polynomial function $1 - \phi_1 L$ is applied to both sides of (4), we obtain

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + Z\alpha_t - \phi_1 Z \alpha_{t-1} + \varepsilon_t \\ &= c_t + \begin{pmatrix} Z \, , & -\phi_1 Z \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \end{pmatrix} + \varepsilon_t, \end{aligned} \tag{6}$$

with $y_1 = Z\alpha_1 + u_1$ and where $c_t = \phi_1 y_{t-1}$ for $t = 2, \ldots, n$. The transition equation for the augmented state vector is given by

$$\begin{pmatrix} \alpha_{t+1} \\ \alpha_t \end{pmatrix} = \begin{bmatrix} T & 0 \\ I & 0 \end{bmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \end{pmatrix}, \tag{7}$$

for $t = 1, \ldots, n$. The introduction of c_t in the observation equation does not cause further complications; it can be incorporated in the Kalman filter since c_t is known at time t.

B. An alternative formulation is obtained by augmenting the state vector with u_t and is given by

$$y_t = (\, Z \,,\, I \,) \begin{pmatrix} \alpha_t \\ u_t \end{pmatrix}, \qquad \begin{pmatrix} \alpha_{t+1} \\ u_{t+1} \end{pmatrix} = \begin{bmatrix} T & 0 \\ 0 & \phi_1 \end{bmatrix} \begin{pmatrix} \alpha_t \\ u_t \end{pmatrix} + \begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix}, \quad (8)$$

for $t = 1, \ldots, n$. The initial condition for the state vector process is straightforwardly determined. The observation disturbance vector has disappeared from this formulation. This loss does not cause complications in the application of the Kalman filter.

Both formulations will lead to the same results when initialisation issues are properly accounted for. Watson and Engle (1983) and, more recently, Reis and Watson (2007) have adopted formulation A while Banbura and Modugno (2008) have adopted formulation B.

2.3 Missing data

In this paper we are concerned with the application of the Kalman filter and smoothing methods to the dynamic factor model in the presence of missing observations. The model formulation B is valid when y_t contains missing entries while formulation A is not valid since c_t cannot be determined when y_{t-1} is partly missing. An exact treatment of filtering and smoothing is therefore not possible when we adopt formulation A. The replacement of c_t by $\hat{c}_t = \phi_1 E(y_t|y_1, \ldots, y_{t-1})$ in the Kalman filter applied to formulation A may lead to a practical solution but it clearly does not lead to an exact solution. This assessment has led Banbura and Modugno (2008) to adopt formulation B for their dynamic factor analysis. The solution is however computationally inefficient since the dimension of the state vector α_t can become very large when N increases. In the empirical study of Banbura and Modugno (2008), the observation dimension is close to $N = 100$ such that their state vector dimension is larger than 100. Although this solution is applicable, a huge dimensional state vector slows down the Kalman filter enormously and may even lead to numerical inaccuracies. It further prohibits its application in a routine manner. Therefore we prefer formulation A where the increase of the state dimension is moderate. The main contribution of this paper is a re-formulation of A that enables Kalman filter and smoothing methods to produce optimal estimates in the presence of missing data. Furthermore we show that the recent developments reported in Jungbacker and Koopman (2008) can be exploited as well in the new state space formulation given below.

3 State space formulation in presence of missing data

In this section we will show how the model of Section 2 can be written as a Gaussian state space model. For ease of notation we pursue the special case of $r = 1$ but with a diagonal coefficient matrix ϕ_1. Section 3.3 discusses the consequences of more general model specifications.

3.1 Notation

Consider some $N \times 1$ vector v_t. The vector $v_t(o_s)$ contains all elements of v_t that correspond to observed entries in the $N \times 1$ data vector y_s for $t, s = 1, \ldots, n$. In a similar way, $v_t(m_s)$ contains all elements of v_t that correspond to missing entries in y_s. In case all entries in y_s are observed, $v_t(m_s)$ is an empty vector. The vector $v_t(o_s, m_{s'})$ contains all elements of v_t that correspond only to observed entries in y_s *and* missing entries in $y_{s'}$ for $t, s, s' = 1, \ldots, n$. Using this notation we can split the vector v_t into four mutually exclusive subvectors $v_t(o_s, o_{s'})$, $v_t(o_s, m_{s'})$, $v_t(m_s, o_{s'})$ and $v_t(m_s, m_{s'})$. In case we have no missing data, vectors $v_t(m_s)$ and $v_t(m_s, m_{s'})$ are empty while $v_t = v_t(o_s) = v_t(o_s, o_{s'})$. We further note that

$$\{v_t\} = \{v_t(o_s), v_t(m_s)\} = \{v_t(o_s, o_{s'}), v_t(o_s, m_{s'}), v_t(m_s, o_{s'}), v_t(m_s, m_{s'})\}\,.$$

To illustrate the notation, consider $N = 5$ and

$$y_t = (1, m, 2, m, 3)', \qquad y_{t-1} = (m, m, m, 4, 5)', \qquad v_t = (6, 7, 8, 9, 10)',$$

where m denotes a missing entry. It follows that

$$\begin{aligned} &v_t(o_t) = (6, 8, 10)', \quad v_t(m_t) = (7, 9)', \\ &v_t(o_t, o_{t-1}) = 10, \quad v_t(o_t, m_{t-1}) = (6, 8)', \quad v_t(m_t, o_{t-1}) = 9, \quad v_t(m_t, m_{t-1}) = 7, \end{aligned}$$

The notation applies to matrices in a similar way. Consider the $N \times k$ matrix V. Matrix $V(o_t; \cdot)$ contains selected rows of V that correspond to the observed entries in y_t while all columns are retained. In case of a $k \times N$ matrix V, the selection $V(\cdot; o_t)$ applies to columns. In case of a $N \times N$ matrix, the selection $V(o_t; m_t)$ applies to both rows and columns.

3.2 The missing data state space formulation

We develop a state space formulation for the observation vector

$$y_t^o = \begin{pmatrix} y_t(o_t, o_{t-1}) \\ y_t(o_t, m_{t-1}) \end{pmatrix},$$

for $t = 1, \ldots, n$. We accomplish the formulation on the basis of the augmented state vector

$$\dot{\alpha}_t = \left[\alpha_t' \,,\, \alpha_{t-1}' \,,\, u_t(o_t, m_{t-1})' \,,\, u_t(m_t, m_{t-1})' \,,\, u_t(m_t, o_{t-1})' \right]'.$$

The state vector is augmented both with α_{t-1} and a selection of u_t. The new formulation below can therefore be interpreted as a mix of formulations A and B in section 2.2.

The observation equation that links the observation vector y_t^o and the state vector $\dot{\alpha}_t$ is obtained straightforwardly as

$$y_t^o = c_t^o + \begin{bmatrix} Z(o_t, o_{t-1}; \cdot) & -\phi_t^o Z(o_t, o_{t-1}; \cdot) & 0 & 0 & 0 \\ Z(o_t, m_{t-1}; \cdot) & 0 & I & 0 & 0 \end{bmatrix} \dot{\alpha}_t + \begin{pmatrix} \varepsilon_t(o_t, o_{t-1}) \\ 0 \end{pmatrix}, \qquad (9)$$

where $c_t^o = [\{\phi_t^o y_{t-1}(o_t, o_{t-1})\}', 0]'$ and $\phi_t^o = \phi_1(o_t, o_{t-1}; o_t, o_{t-1})$. Matrix ϕ_t^o is diagonal consisting of (a subset of), possibly reshuffled, diagonal elements of ϕ_1. The specification for $y_t(o_t, o_{t-1})$ relies on formulation A while for $y_t(o_t, m_{t-1})$ it relies on formulation B. The major difference of our formulation with B is that we only include those entries of u_t in the state vector that correspond to missing entries in y_t and/or y_{t-1}. For those entries of y_t where both y_t and y_{t-1} are observed, we can compute the corresponding entries in c_t^o and rely on formulation A.

The transition equation for the state process $\dot{\alpha}_t$ is obtained as follows. The updates for α_t and α_{t-1} are given as in (7) for formulation A. Next we develop equations for $u_{t+1}(o_{t+1}, m_t)$ and $u_{t+1}(m_{t+1}, m_t)$ which are effectively the selection $u_{t+1}(m_t)$ (re-ordered). The transition from $u_t(m_t)$ to $u_{t+1}(m_t)$ is the autoregressive update (5) with $r = 1$ in our case. For our selection of u_t, we have

$$u_{t+1}(m_t) = \phi_1(m_t; m_t) u_t(m_t) + \varepsilon_t(m_t), \qquad u_t(m_t) = \begin{pmatrix} u_t(m_t, m_{t-1}) \\ u_t(m_t, o_{t-1}) \end{pmatrix},$$

for $t = 1, \ldots, n$. To place $u_{t+1}(m_t)$ into $\dot{\alpha}_{t+1}$, we need to re-order it into

$$\begin{pmatrix} u_{t+1}(o_{t+1}, m_t) \\ u_{t+1}(m_{t+1}, m_t) \end{pmatrix} = J_t u_{t+1}(m_t),$$

where J_t is implicitly defined as a selection matrix of ones and zeroes. The bottom part of $\dot{\alpha}_{t+1}$ is $u_{t+1}(m_{t+1}, o_t)$ and corresponds to observed entries in y_t. Therefore, we have

$$\begin{aligned} u_{t+1}(m_{t+1}, o_t) &= \phi_t^* u_t(m_{t+1}, o_t) + \varepsilon_t(m_{t+1}, o_t) \\ &= \phi_t^* \left[y_t(m_{t+1}, o_t) - Z_t^* \alpha_t \right] + \varepsilon_t(m_{t+1}, o_t), \end{aligned}$$

where $\phi_t^* = \phi_1(m_{t+1}, o_t; m_{t+1}, o_t)$ and $Z_t^* = Z(m_{t+1}, o_t; \cdot)$. The transition equation for $\dot{\alpha}_t$ is

therefore

$$\dot{\alpha}_{t+1} = d_t + \begin{bmatrix} T & 0 & 0 & 0 \\ I & 0 & 0 & 0 \\ 0 & 0 & 0 & J_t\phi_1(m_t; m_t) \\ -\phi_t^* Z_t^* & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \\ u_t(o_t, m_{t-1}) \\ u_t(m_t) \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \\ J_t\varepsilon_t(m_t) \\ \varepsilon_t(m_{t+1}, o_t) \end{pmatrix}, \quad (10)$$

where $d_t = [\,0\,,\,0\,,\,0\,,\,\{\,\phi_t^* y_t(m_{t+1}, o_t)\,\}'\,]'$, for $t = 1, \ldots, n$.

3.3 Discussion of the new formulation

The equations (9) and (10) define the state space model for the observed values while the missing observations are accounted for by including the relevant u_t's in the state vector. In case we have no missing data, the vectors $u_t(o_t, m_{t-1})$ and $u_t(m_t)$ are empty and we return to formulation A. Entries of u_t only appear in the state vector when they correspond to missing entries in y_t or in y_{t-1}. In this way we keep the dimension of the state to a minimum while at all times we are able to produce optimal estimates using Kalman filter and smoothing.

In most cases the dimension of $\dot{\alpha}_t$ will be smaller than the dimension of $(\,\alpha_t', u_t'\,)'$, the state vector in model formulation B. In case $\dot{\alpha}_t$ has a larger dimension than $(\,\alpha_t', u_t'\,)'$, due to a large number of missings in y_{t-1} or y_t, it is possible to reduce the dimension of $\dot{\alpha}_t$ by dropping α_{t-1} (partially and temporarily) from the state vector $\dot{\alpha}_t$. Since the resulting computational gains will be relatively small, we will not pursue this further.

The new formulation does imply time-varying system matrices in the observation and transition equations. In fact, the dimension of the state vector also varies over time. Fortunately, the Kalman filter can treat varying dimensions for the state vector. The implementation of such a Kalman filter requires attention but it comes with the benefit of a dynamic factor analysis that is computationally feasible when missing data is present. We give some more details in the next section.

The results presented in this paper apply to more general settings. For example, the dynamic specification of f_t may also depend on non-stationary time series processes such as a random walk. The Kalman filter and smoothing need to deal with the initialisation problem but existing solutions can be applied straightforwardly. Lagged factors and explanatory variables can be included in the observation equation of the dynamic factor model, see the discussion in Jungbacker and Koopman (2008). Given our solution, the case of $r > 1$ is trivial but requires more notation in the expositions of sections 3 and 4. A particular concern is the case of a non-diagonal autoregressive coefficient matrix ϕ_1 since it mixes the lag-dependence of idiosyncratic components associated with missing entries to those associated with observed entries. We therefore need to modify the system matrices in (10) accordingly. This exercise is straightforward but the notation is somewhat cumbersome.

4 Signal extraction and likelihood evaluation

In this section we discuss computationally efficient approaches to signal extraction and likelihood evaluation. These methods are also relevant for parameter estimation that is discussed in section 5.

4.1 Estimation of states and idiosyncratic components

Given the state space formulation of the dynamic factor model, we can adopt the Kalman filter and associating smoothing methods (KFS) to obtain

$$\dot{a}_{t|s} = \mathbb{E}(\dot{\alpha}_t|Y_s), \qquad \dot{P}_{t|s} = \mathrm{Var}(\dot{\alpha}_t|Y_s),$$

for $t, s = 1, \ldots, n$ where $Y_s = (y_1^o, \ldots, y_s^o)$, see Durbin and Koopman (2001) for an exposition of these methods. Prediction refers to $s = t-1$, concurrent filtering to $s = t$ and smoothing to $s = n$. The Kalman filter can also be used to evaluate the log-likelihood function using the prediction error decomposition result of Schweppe (1965), see section 4.4.

In terms of the dynamic factor model (4), (3) and (5), KFS produces estimates (as well as the mean square errors) of α_t and of those entries of u_t that are associated with missing entries in y_t and y_{t-1}, that is $u_t^m = [\, u_t(o_t, m_{t-1})' \,,\, u_t(m_t)' \,]'$. We can also obtain estimates and corresponding mean square errors of $u_t^o = u_t(o_t, o_{t-1})$ using the identity $u_t = y_t - Z\alpha_t$ in (4). Let $a_{t|s} = \mathbb{E}(\alpha_t|Y_s)$ and $P_{t|s} = \mathrm{Var}(\alpha_t|Y_s)$ for $t, s = 1, \ldots, n$. Obviously, $a_{t|s}$ and $P_{t|s}$ are the upper (block) parts of $\dot{a}_{t|s}$ and $\dot{P}_{t|s}$, respectively. It follows that

$$\begin{aligned} \mathbb{E}(u_t^o|Y_s) &= y_t^o - Z_t^o a_{t|s}, & \mathrm{Var}(u_t^o|Y_s) &= Z_t^o P_{t|s} Z_t^{o\prime}, \\ \mathrm{Cov}(u_t^o, \alpha_t|Y_s) &= -Z_t^o P_{t|s}, & \mathrm{Cov}(u_t^o, u_t^m|Y_s) &= -Z_t^o \mathrm{Cov}(\alpha_t, u_t^m|Y_s), \end{aligned}$$

where $Z_t^o = Z(o_t, o_{t-1}; \cdot)$ and $\mathrm{Cov}(u_t^m, \alpha_t|Y_s)$ is part of $\dot{P}_{t|s}$ for $t, s = 1, \ldots, n$.

4.2 KFS with a collapsed observation vector

The computational effort for the KFS depends on the dimensions of both the state and observation vectors. Consider the dynamic factor model (1) with $q \times 1$ vector $f_t = S\alpha_t$ and state space representation (4) and (3) but with u_t replaced by $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$, that is

$$y_t = Z\alpha_t + \varepsilon_t, \qquad \alpha_{t+1} = T\alpha_t + \eta_t, \tag{11}$$

for $t = 1, \ldots, n$ with $N \times 1$ observation vector y_t and $p \times 1$ state vector α_t. In most practical applications of the dynamic factor model, the dimension of y_t is significantly larger than the dimension of α_t. Jungbacker and Koopman (2008) demonstrate that in such circumstances,

when $N > q$, computational efficiency of KFS can significantly be improved by a simple computational device. Recall that $Z = \Lambda S$ and define the $N \times N$ and $q \times N$ matrices

$$A = \begin{bmatrix} A^L \\ A^H \end{bmatrix}, \qquad A^L = C^{-1}\Lambda'\Sigma_\varepsilon^{-1},$$

respectively, where C can be any invertible matrix and A^H is chosen such that matrix A is full rank and $A^L\Sigma_\varepsilon A^{H\prime} = 0$. It follows that $A^H Z = 0$. We assume that Λ has full column rank. In most cases of practical interest this assumption will be valid. If matrix Λ does not have full rank, it can be replaced with any full rank matrix that spans the column space of Λ, see the discussion in Jungbacker and Koopman (2008). Matrix A^H exists by construction but it does not need to be evaluated for our purposes. By choosing C such that $CC' = Z'\Sigma_\varepsilon^{-1}Z$, we have

$$Ay_t = \begin{pmatrix} A^L y_t \\ A^H y_t \end{pmatrix} = \begin{pmatrix} C'S \\ 0 \end{pmatrix} \alpha_t + \begin{pmatrix} A^L\varepsilon_t \\ A^H\varepsilon_t \end{pmatrix}, \quad \begin{pmatrix} A^L\varepsilon_t \\ A^H\varepsilon_t \end{pmatrix} \sim N\left(0, \begin{bmatrix} I & 0 \\ 0 & A^H\Sigma_\varepsilon A^{H\prime} \end{bmatrix}\right),$$

for $t = 1, \ldots, n$. The equation for α_{t+1} is unaffected by the transformation. It follows that the part $A^H y_t$ does not depend on α_t, it is not correlated with $A^L y_t$ and therefore does not need to be considered for the estimation of α_t. Therefore, the KFS only need to be applied to the collapsed observation (low-dimensional) vector $A^L y_t$ for signal extraction. Since $\mathrm{Var}(A^L\varepsilon_t) = I$, we can adopt the KFS devices discussed in Koopman and Durbin (2003) to further accelerate the computations.

The collapse can lead to high computational savings. To illustrate the reductions that we can achieve in practice, consider model (11) with $N = 100$ and $p = 10$. In this case, the observation vector relevant for the application of the KFS collapses from dimension $N = 100$ to dimension $p = 10$. Jungbacker and Koopman (2008) also demonstrate that likelihood evaluation can rely on the Kalman filter applied to $A^L y_t$, see section 4.4.

4.3 A collapsed KFS in presence of missing data

The computational device of Jungbacker and Koopman (2008) can be modified in the context of the state space formulation developed in section 3.2, in case missing data is present. Consider the observation equation (9). Since this formulation relies on time-varying system matrices, we require the collapsed transformations to vary over time as well.

We carry out a partial collapse of y_t^o and only consider the transformation of $y_t(o_t, o_{t-1})$ with dimension N_t^{oo}. For this purpose, we define

$$A_t^L = C_t^{-1}Z_t^{+\prime}V_t^{-1}, \quad Z_t^+ = \left[\, \Lambda(o_t, o_{t-1}; \cdot)\,, -\phi_t^o\Lambda(o_t, o_{t-1}; \cdot)\,\right], \quad V_t = \Sigma_\varepsilon(o_t, o_{t-1}; o_t, o_{t-1}),$$

where $\phi_t^o = \phi_1(o_t, o_{t-1}; o_t, o_{t-1})$ and C_t is chosen such that

$$C_t C_t' = Z_t^{+\prime} V_t^{-1} Z_t^+,$$

for $t = 1, \ldots, n$. Again, we should make sure that Z_t^+ has full column rank. If this is not the case it is generally easy to find a new matrix with full column rank that spans the same column space. The transformation A_t^L is applied to $y_t(o_t, o_{t-1})$ only and does not require to consider the elements of $\dot{\alpha}_t$ associated with u_t since they do not affect $y_t(o_t, o_{t-1})$. We can extend the transformation towards $y_t(o_t, m_{t-1})$ but this will not lead to further reductions $y_t(o_t, m_{t-1})$.

Define matrix

$$A_t = \begin{bmatrix} A_t^L \\ A_t^H \end{bmatrix},$$

where A_t^H is chosen such that $A_t^L V_t A_t^{H\prime} = 0$ and A_t is a full rank matrix. The state space model for the transformed observation vector $A_t y_t(o_t, o_{t-1})$ is given by

$$\begin{pmatrix} A_t^L y_t(o_t, o_{t-1}) \\ A_t^H y_t(o_t, o_{t-1}) \end{pmatrix} = \begin{pmatrix} A_t^L c_t^o \\ A_t^H c_t^o \end{pmatrix} + \begin{bmatrix} C_t' S & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \dot{\alpha}_t + A_t \varepsilon_t(o_t, o_{t-1}), \tag{12}$$

where $q \times p$ matrix S is defined in (2) and $\mathrm{Var}[A_t \varepsilon_t(o_t, o_{t-1})]$ is a block-diagonal variance matrix with the upper-block given by $\mathrm{Var}[A_t^L \varepsilon_t(o_t, o_{t-1})] = I$. It follows that we can remove $A_t^H y_t(o_t, o_{t-1})$ for the application of the KFS and for the same reasons as discussed in section 4.2. In particular, for the application of KFS we can replace (9) by the two observation equations

$$\begin{pmatrix} A_t^L y_t(o_t, o_{t-1}) \\ y_t(o_t, m_{t-1}) \end{pmatrix} = \begin{pmatrix} A_t^L c_t^o \\ 0 \end{pmatrix} + \begin{bmatrix} C_t' S & 0 & 0 & 0 \\ Z_t^{om} & I & 0 & 0 \end{bmatrix} \dot{\alpha}_t + \begin{pmatrix} A_t^L \varepsilon_t(o_t, o_{t-1}) \\ 0 \end{pmatrix}, \tag{13}$$

where $Z_t^{om} = \{ Z(o_t, m_{t-1}; \cdot), 0 \}$. In most cases, the observation vector dimension of the collapsed model will be much lower than the dimension of y_t^o. However, when y_t does contain many missing observations, it may become the case that the dimension of $y_t(o_t, o_{t-1})$ is lower than $2p$. In this case no computational gain can be achieved by transforming the model. The state space model (13) and (10) satisfies the conditions required for the KFS devices discussed in Koopman and Durbin (2003).

In case both observation vectors y_t and y_{t-1} contain no missing entries, we can apply the time-invariant transformation as developed in section 4.2 and based on the state space formulation A of section 2.2. We only require the modifications for collapsing the observation vector presented in this section when missing entries in the observation vectors y_t or y_{t-1} are present.

4.4 Log-likelihood evaluation

For a set of realizations $y_t, \ldots, y_n$ as generated by the state space model, we define the log-likelihood function by

$$\ell(y) = \log p(y_1^o, \ldots, y_n^o; \theta), \qquad y = \{y_t^o\}_{t=1}^n, \tag{14}$$

where $p(\cdot)$ is the Gaussian density function, y is the set of observed data, and θ is the vector of parameters introduced in section 2. The prediction error decomposition result of Schweppe (1965) implies that $\log p(y_1^o, \ldots, y_n^o; \theta) = \log p(y_1^o; \theta) + \sum_{t=2}^n \log p(y_t^o | Y_{t-1}; \theta)$ where $p(y_t^o | Y_{t-1}; \theta)$ can be evaluated by the Kalman filter.

Jungbacker and Koopman (2008) argue that the likelihood function $\ell(y)$ can be obtained by the Kalman filter applied to the collapsed data vector. In our case, we can limit the application of the Kalman filter to the observation equation (13). The log-likelihood function is then evaluated by

$$\ell(y) = \text{constant} + \ell\left(y^L\,,\, y^{om}\right) + \ell\left(y^H\right),$$

where

$$y^L = \{A_t^L y_t(o_t, o_{t-1})\}_{t=1}^n, \qquad y^{om} = \{y_t(o_t, m_{t-1})\}_{t=1}^n, \qquad y^H = \{A_t^H y_t(o_t, o_{t-1})\}_{t=1}^n,$$

and the constant does not depend on θ nor on the obervations. The log-likelihood function $\ell\left(y^L\,,\, y^{om}\right)$ is obtained from the Kalman filter applied to the state space model (13) and (10). The log-likelihood function $\ell\left(y^H\right)$ can be evaluated by

$$\ell\left(y^H\right) = \text{constant} - \frac{1}{2}\sum_{t=1}^n \log|V_t| - \frac{1}{2}\sum_{i=1}^n e_t' V_t^{-1} e_t,$$

where e_t is given by

$$e_t = \left(I - V_t A_t^{L\prime} A_t^L\right)\left[y_t(o_t, o_{t-1}) - \phi_t^o y_{t-1}(o_t, o_{t-1})\right],$$

for $t = 1, \ldots, n$, see Jungbacker and Koopman (2008, Lemma 2).

5 Maximum likelihood estimation

The estimation of θ discussed in section 2 by maximum likelihood methods requires solving a potentialy high dimensional optimization problem. It involves the maximization of $\ell(y; \theta)$ with respect to θ and where $\ell(y; \theta)$ can be evaluated by the methods described in section 4.4.

5.1 Direct maximization

In general, large-scale optimization problems are handled by quasi-Newton algorithms as described in Nocedal and Wright (1999). These algorithms require the evaluation of $\ell(y;\theta)$ and the score function at each iteration of the algorithm. Since the number of parameters in dynamic factor models are typically large, the numerical computation of the score function is not feasible. Fortunately, analytic expressions for the score function are available for dynamic factor models in state space form and which can be computed efficiently, see the discussion in Koopman and Shephard (1992) and Jungbacker and Koopman (2008).

Let $Q(\theta|\theta^*)$ denote the complete expected loglikelihood defined by

$$Q(\theta|\theta^*) = \mathbb{E}_{\theta^*}\left[\log p(y^o, \dot{\alpha}_1, \ldots, \dot{\alpha}_n; \theta) \middle| y^o\right],$$

where $p(y^o, \dot{\alpha}_1, \ldots, \dot{\alpha}_n; \theta)$ is the joint density of y^o and $\dot{\alpha}_1, \ldots, \dot{\alpha}_n$ for parameter vector θ. The subscript θ^* in $\mathbb{E}_{\theta^*}$ emphasizes that the expectation is calculated for a given parameter vector θ^*. From the results in Louis (1982) and Ruud (1991), we have

$$\left.\frac{\partial Q(\theta|\theta^*)}{\partial \theta}\right|_{\theta=\theta^*} = \left.\frac{\partial \ell(y;\theta)}{\partial \theta}\right|_{\theta=\theta^*},$$

provided that θ is the true parameter. Note that $\ell(y;\theta)$ is defined in (14). In Appendix 7 we provide the expressions for the derivatives of $Q(\theta|\theta^*)$ with respect to the system matrices of the state space model (9) and (10). From these expressions and the chain rule we can determine the score function for all the parameters in the model of section 2. The computation of the score in this way only requires a single run of the KFS and can be done computationally efficiently due to the results of the previous section.

5.2 Expectation-Maximization algorithm

An alternative method is the expectation-maximization (EM) algorithm for finding the maximum likelihood estimators, see Watson and Engle (1983) and Shumway and Stoffer (1982). The EM algorithm produces a sequence of proposals $\theta^{(1)}, \theta^{(2)}, \ldots$ for the maximum likelihood estimator of θ by repeating the following two steps for $j = 1, 2, \ldots$:

- E-step: obtain the expected complete likelihood $Q(\theta|\theta^{(j)})$ where $\theta^{(j)}$ is the current estimate.
- M-step: maximize $Q(\theta|\theta^{(j)})$ with respect to θ and set $\theta^{(j+1)}$ to the parameter value where this optimum is attained.

The EM algorithm has the attractive property that it always converges to a (local) optimum and the likelihood is ensured to increase with each iteration. The E-step can be performed by

means of the KFS in a relatively straightforward way for the dynamic factor model considered in this paper. The maximization in the M-step can however not be done analytically. Watson and Engle (1983) propose an ad-hoc iterative scheme for the M-step. Alternatively, we can perform the M-step via a quasi-Newton scheme. Since the gradient of $Q(\theta|\theta^{(n)})$ is available analytically, the necessary computations can be done computationally efficient.

6 Computational costs and gains

In this section we explore the computational gains that we can obtain using our new new state space specification of section 3.2 when applying the Kalman filter and associated smoothing algorithm (KFS). We do this by considering the computational costs compared to the state space formulations A and B of section 2.2. Below we will refer to our new model as formulation C. In comparison with A, the computing times for KFS without the presence of missing data will be the same since the two specifications are equivalent in this case. When missing observations are present, formulation A is not valid while formulation C is. The costs of the additional computations are modest when the number of missing entries is modest. When y_t and y_{t-1} have a total of m unique missing entries (the entries that are both missing in y_t and y_{t-1} are counted once), the state vector $\dot{\alpha}_t$ needs to be increased by m (temporarily). This will slow down the KFS computations but it will lead to to exact results while formulation A cannot deal with missing entries. The increase in computing times depend on the number of missings in the data-set. In Table 1(a) we provide some indications of the computational costs for the dynamic factor model (1) with two dynamic factors ($q = 2$) which are modelled as stationary vector autoregressive processes. The comparisons are carried out for three different observation vector dimensions $N = 10, 50, 100$. We have implemented the collapsed KFS as described in section 4.3. The results reveal, for example, that for $N = 50$ and for 1% missing observations (missing entries are randomlychosen in the sample), the computations take 1.5 times longer than those for formulation A (instead of, say, 20 seconds, it takes 30 seconds). When the number of missings has increased by 10%, the computations take 2.6 times longer.

The formulation B also provides exact results when data is missing and it is the approach adopted by Banbura and Modugno (2008). However, we have argued in this paper that formulation C is computationally more efficient. In Table 1(b) we compare the computing times for formulations B and C. The gains of the new formulation compared to B are quite considerable. For the same model as described above with $N = 50$, the KFS for formulation C is almost 88 times faster when we have 1% missings while it is 44 times faster when 10% of the data is missing. These gains are considerable and they are even higher and more dramatic when N increases to higher values. We therefore suggest to use formulation C for dynamic factor analysis in the presence of missing data.

Table 1: Computational costs and gains

The table (a) presents ratios of computing times for the formulation of section 3.2 with missing data divided by those for the formulation without missing data (this is formulation A of section 2.2). For example, the value 2 indicates that the computational demands are as twice as high. The table (b) presents ratios of computing times for the formulation B of section 2.2 divided by those for the formulation of section 3.2 with missing data. For example, the value 2 indicates that the new device is twice as fast. The ratios are presented for different dimensions N of the observation vector y_t and for different percentages of missing data.

(a) Costs relative to A

N	percentage missing		
	1%	10%	25%
10	1.4	1.8	2.3
50	1.5	2.6	8.9
100	1.2	3.9	24.8

(b) Gains relative to B

N	percentage missing		
	1%	10%	25%
10	2.1	1.5	1.1
50	87.9	43.7	11.3
100	625.5	197.5	25.8

7 Conclusions

Dynamic factor analysis has been given renewed attention in the economics and finance literature recently. High-dimensional dynamic models with multiple factors contain many parameters that need to be estimated. In case maximum likelihood estimation is requested and optimization is employed via Fisher-scoring and/or the EM algorithm, computational efficient methods are of key importance. Various problems arise when these methods are used when missing entries in the data-set are present. A standard solution requests to include the idiosyncratic component in the state vector. It will lead to a high-dimensional state vector for a model with a high-dimensional observation vector. The associating computations for Kalman filtering and smoothing will slow down considerably. To circumvent this problem, we propose a new state space formulation of the model that allows for missing values and can exploit existing devices for computational efficiency. Only the idiosyncratic components associated with missing entries for the concurrent and previous time periods are accommodated in the state vector, all other ones are removed from the state vector. As a result, the dimension of the state vector is kept to a minimum. Also we have shown that the existing computational devices to accelerate the computations can be adopted. In our formulation the dimension of the state vector varies over time and therefore the implementation of the Kalman filter and smoothing methods requires attention. However, we can obtain high computational savings even when the number of missing entries is moderate, see Table 1.

Appendix: the details for score evaluation

Suppose the vector time series $y_1, \ldots, y_n$ is generated by the linear Gaussian state space model as given by

$$\begin{aligned} y_t &= c_t + Z_t\alpha_t + \varepsilon_t & \varepsilon_t &\sim N(0, H_t), \\ \alpha_{t+1} &= d_t + T_t\alpha_t + R_t\eta_t & \eta_t &\sim N(0, Q_t), \end{aligned}$$

with $\alpha_1 \sim N(a_{1|0}, P_{1|0})$ and $t = 1, \ldots, n$. The system vectors and matrices c_t, d_t, Z_t, T_t, H_t and Q_t are fixed and may depend on the parameter vector θ. Selection matrix R_t has full column rank, is fixed and does not depend on θ. Finally, we assume that H_t is a non-singular variance matrix.

Let

$$a_{t|n} = \mathbb{E}(\alpha_t|y_1, \ldots, y_n), \quad P_{t|n} = \mathrm{Var}(\alpha_t|y_1, \ldots, y_n), \quad P_{t+1,t|n} = \mathrm{Cov}(\alpha_{t+1}, \alpha_t|y_1, \ldots, y_n),$$

for $t = 1, \ldots, n$. These quantities can be computed by the Kalman filter and associating state smoothing method, see Durbin and Koopman (2001, Chapter 4). Denote $\ell(y)$ by the log-likelihood of $y_1, \ldots, y_n$. We then have the following expressions for the score vectors with respect to each of the system matrices

$$\begin{aligned} \frac{\partial \ell(y)}{\partial d_t} &= \bar{R}_t' Q_t^{-1} \bar{R}_t (a_{t+1|n} - T_t a_{t|n} - d_t), & \frac{\partial \ell(y)}{\partial T_t} &= \bar{R}_t' Q_t^{-1} \bar{R}_t (M_{T_t} - T_t M_{Z_t}) \\ \frac{\partial \ell(y)}{\partial c_t} &= H_t^{-1}(y_t - Z_t a_{t|n} - c_t), & \frac{\partial \ell(y)}{\partial Z_t} &= H_t^{-1}\left[(y_t - c_t)a_{t|n}' - Z_t M_{Z_t})\right], \\ \frac{\partial \ell(y)}{\partial Q_t} &= Q_t^{-1} M_{Q_t} Q_t^{-1} - \frac{1}{2}\mathrm{diag}\{Q_t^{-1} M_{Q_t} Q_t^{-1}\}, \\ \frac{\partial \ell(y)}{\partial H_t} &= H_t^{-1} M_{H_t} H_t^{-1} - \frac{1}{2}\mathrm{diag}\{H_t^{-1} M_{H_t} H_t^{-1}\}, \end{aligned}$$

where $\bar{R}_t = (R_t' R_t)^{-1} R_t'$ and

$$\begin{aligned} M_{Q_t} &= \mathbb{E}(\eta_t \eta_t'|y_1, \ldots, y_n) - Q_t, & M_{T_t} &= a_{t+1|n} a_{t|n}' + P_{t+1,t|n}, \\ M_{H_t} &= (y_t - Z_t a_{t|n})(y_t - Z_t a_{t|n})' + Z_t P_{t|n} Z_t' - H_t, & M_{Z_t} &= a_{t|n} a_{t|n}' + P_{t|n}. \end{aligned}$$

The matrix M_{Q_t} can be evaluated using the identity $\eta_t = \bar{R}_t(\alpha_{t+1} - T_t\alpha_t - d_t)$ and from $P_{t+1,t|n}$ and $P_{t|n}$.

References

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–72.

Banbura, M. and M. Modugno (2008). Maximum likelihood estimation of large factor model on datasets with arbitrary pattern of missing data. Working paper.

Dempster, A. P., N. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society B 39*, 1–38.

Doz, C., D. Giannone, and L. Reichlin (2006). A quasi maximum likelihood approach for large approximate dynamic factor models. Discussion paper, CEPR.

Duncan, D. B. and S. D. Horn (1972). Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. American Statistical Association 67*, 815–21.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Engle, R. F. and M. W. Watson (1981). A one-factor multivariate time series model of metropolitan wage rates. *J. American Statistical Association 76*, 774–81.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: Identification and estimation. *Rev. Economics and Statistics 82*, 540–54.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. J. Aigner and A. S. Goldberger (Eds.), *Latent variables in socio-economic models*. Amsterdam: North-Holland.

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

Jungbacker, B. and S. J. Koopman (2008). Likelihood-based analysis for dynamic factor models. Working paper.

Koopman, S. J. and J. Durbin (2003). Filtering and smoothing of state vector for diffuse state space models. *J. Time Series Analysis 24*, 85–98.

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*, 823–6.

Louis, T. A. (1982). Finding observed information using the EM algorithm. *J. Royal Statistical Society B 44*, 98–103.

Nocedal, J. and S. J. Wright (1999). *Numerical Optimization*. New York: Springer Verlag.

Reis, R. and M. W. Watson (2007). Relative goods prices and pure inflation. NBER Working paper.

Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *J. Econometrics 49*, 305–41.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. In C. A. S. et al. (Ed.), *New methods in business cycle research*. Minneapolis: Federal Reserve Bank of Minneapolis.

Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory 11*, 61–70.

Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Analysis 3*, 253–64.

Stock, J. H. and M. Watson (2002a). Forecasting using principal components from a large number of predictors. *J. American Statistical Association 97*, 1167–79.

Stock, J. H. and M. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *J. Business and Economic Statist. 20*, 147–162.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression. *J. Econometrics 23*, 385–400.