Likelihood-based Analysis for Dynamic Factor Models

Jungbacker, B.M.J.P.; Koopman, S.J.

2008

TI 2008-0007/4
Tinbergen Institute Discussion Paper

Likelihood-based Analysis for Dynamic Factor Models

Borus Jungbacker
*Siem Jan Koopman**

VU University Amsterdam.

* *Tinbergen Institute*

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.

Likelihood-based analysis for dynamic factor models

Borus Jungbacker and Siem Jan Koopman
Department of Econometrics, VU University Amsterdam,
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

January 2008

Abstract

We present new results for the likelihood-based analysis of the dynamic factor model that possibly includes intercepts and explanatory variables. The latent factors are modelled by stochastic processes. The idiosyncratic disturbances are specified as autoregressive processes with mutually correlated innovations. The new results lead to computationally efficient procedures for the estimation of the factors and parameter estimation by maximum likelihood and Bayesian methods. An illustration is provided for the analysis of a large panel of macroeconomic time series.

JEL classification: C33; C43

Some keywords: EM algorithm; Kalman Filter; Forecasting; Latent Factors; Markov chain Monte Carlo; Principal Components; State Space.

1 Introduction

In this paper we consider likelihood-based inference for the dynamic factor model given by

$$y_{it} = \mu_i + x_{it}\beta + \lambda_i' f_t + u_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where y_{it} denotes the observed value for the i^{th} time series at time t, μ_i is a fixed and unknown constant, x_{it} is a $1 \times K$ vector of covariates, β is a $K \times 1$ vector of regression coefficients, λ_i is an $r \times 1$ vector of factor loadings, f_t is an $r \times 1$ vector of common factors and u_{it} is the idiosyncratic component. We focus on the case where a high-dimensional panel of time series depends on a relatively small number of dynamic factors, $N > r$. The

factors are modelled by dynamic linear Gaussian processes. The idiosyncratic components are modelled by autoregressive processes with mutually correlated zero mean innovations. Since the number of unknown parameters can be very large in such models, likelihood-based methods are usually regarded as computationally infeasible.

In this paper we present new results that lead to computationally efficient methods for a likelihood-based analysis of high-dimensional dynamic factor models. We cover signal extraction, parameter estimation via maximum likelihood and Bayesian estimation using Markov chain Monte Carlo methods. In the empirical section of this paper we treat a panel of 132 time series from which a maximum of seven dynamic factors are extracted. This model requires the estimation of more than 1300 parameters. The new results of this paper enable us to estimate this set of parameters in minutes.

We show how model (1) can be represented by two vector series, a low-dimensional vector and a high-dimensional vector. The low-dimensional vector series is sufficient for extracting the latent dynamic factors using Kalman filter methods while standard regression calculations are employed to estimate mean and regression effects. The loglikelihood function is evaluated efficiently since most computations are only concerned with the low-dimensional series. As a result, large computational gains are achieved in classical as well as Bayesian estimation procedures. We present evidence of the computational savings. Estimating a large number of coefficients in high-dimensional dynamic factor models is not straightforward. We present new devices for an effective implementation of the estimation methods. Finally, we provide a typical empirical illustration for a macroeconomic data set to emphasize the relevance of the new results.

Sargent and Sims (1977) and Geweke (1977) were the first to propose a dynamic factor model. They obtain parameter estimates by maximizing the spectral likelihood function. Engle and Watson (1981) propose the use of Fisher scoring to maximize the likelihood in the time domain and apply this method to a one-factor model of wage rates. In Watson and Engle (1983) the expectation-maximization (EM) algorithm of Dempster, Laird, and Rubin (1977) is adopted, see also Shumway and Stoffer (1982). The EM algorithm has the advantage that it is stable and it is sure to converge to an optimum. Watson and Engle (1983) found however that convergence is often slow. Quah and Sargent (1993) use the EM algorithm to estimate a factor model on employment data.

The high-dimensional data sets available in economics and finance have motivated the work on alternative methods to estimate dynamic factor models. Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1988, 1993) show that if N goes to infinity the factors are estimated consistently using the method of principal components. More re-

cent contributions have focused on extending the inferential theory of this method, see *e.g.* Stock and Watson (2002a) and Bai (2003). Stock and Watson (2002b, 2006) demonstrate the value of this approach for the purpose of constructing diffusion indexes that can be used in forecasting macroeconomic time series. Forni, Hallin, Lippi, and Reichlin (2000) propose a different estimation procedure, based on frequency domain methods, that provides consistent estimates of the factors for a general class of dynamic factor models.

Doz, Giannone, and Reichlin (2006) show that estimates of the factors obtained from a likelihood-based analysis are consistent if both N and T are large, 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.

The remainder of the paper is organised as follows. The dynamic factor model and its state space form are presented in Section 2. The key results are presented in Section 3. The new devices for parameter estimation using maximum likelihood methods and Bayesian methods are discussed in Section 4. An empirical illustration is provided in Section 5 while Section 6 concludes.

2 The dynamic factor model

2.1 Model specification

In vector form the dynamic factor model (1) is given by

$$y_t = \bar{\mu} + \bar{X}_t\beta + \bar{\Lambda}f_t + u_t, \qquad t = 1, \ldots, T, \tag{2}$$

where $\bar{\mu} = (\mu_1, \ldots, \mu_N)'$, $\bar{X}_t = (x_{1t}' \cdots x_{Nt}')'$, $u_t = (u_{1t}, \ldots, u_{Nt})'$, and $\bar{\Lambda} = (\lambda_1 \cdots \lambda_N)'$. The common factors f_t are modelled as a stationary Vector Autoregressive Moving Average (VARMA) process:

$$f_t = \Phi_1 f_{t-1} + \cdots + \Phi_{q_\Phi} f_{t-q_\Phi} + \zeta_t + \Theta_1\zeta_{t-1} + \cdots + \Theta_{q_\Theta}\zeta_{t-q_\Theta}, \qquad \zeta_t \sim NIID(0, \Sigma_\zeta), \tag{3}$$

for $t = 1, \ldots, T$, where Φ_j are $r \times r$ autoregressive coefficient matrices for $j = 1, \ldots, q_\Phi$ and Θ_j are $r \times r$ moving average coefficient matrices for $j = 1, \ldots, q_\Theta$. The idiosyncratic components u_{it} in (1) are modelled as a Vector Autoregressive (VAR) process

$$u_t = \Psi_1 u_{t-1} + \cdots + \Psi_{q_\Psi} u_{t-q_\Psi} + \varepsilon_t, \qquad \varepsilon_t \sim NIID(0, \Sigma_\varepsilon), \qquad t = 1, \ldots, T, \tag{4}$$

where Σ_ε is a $N \times N$ variance matrix and Ψ_i is a $N \times N$ matrix for $i = 1, \ldots, q_\Psi$. The innovations, ε_s and ζ_t, are independent for $s, t = 1, \ldots, T$.

Generally, not all parameters in $\bar{\Lambda}$, $\Phi_1, \ldots, \Phi_{q_\Phi}$, $\Theta_1, \ldots, \Theta_{q_\Theta}$ and Σ_ζ can be identified. In such cases, it will therefore be necessary to restrict these matrices in a suitable manner, depending on the specific form of the model. Parameter restrictions are common in the literature on factor models, see *e.g.* Geweke and Zhou (1996) for further discussions. Examples are given in Illustrations 1 and 2.

2.2 State space representation

Next we show that the dynamic factor model (2), (3) and (4) can be expressed in the form

$$y_t = \mu + d_t + X_t\beta + Z\alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim NIID(0, \Sigma_\varepsilon), \tag{5}$$

$$\alpha_t = H\alpha_{t-1} + R\eta_t, \qquad \eta_t \sim NIID(0, \Sigma_\eta), \tag{6}$$

where $y_t = (y_{1t}, \ldots, y_{Nt})'$, μ and d_t are vectors of constants, X_t is a $N \times K$ matrix of covariates, Z is a $N \times p$ matrix, α_t is the $p \times 1$ state vector with initial specification $\alpha_1 \sim N(a, P)$ for given a and P of suitable dimension, H is the $p \times p$ transition matrix, R is a $p \times q$ matrix, Σ_η is a $q \times q$ variance matrix and the disturbances η_s and ε_t are independent of each other for $s, t = 1, \ldots, T$. The model consisting of the equations (5) and (6) is a special case of the linear Gaussian state space model. The likelihood function for this model can therefore be calculated by means of the Kalman filter while the unobserved factors f_t can be estimated using the associated smoothing algorithm. Detailed accounts of the state space methodology can be found in textbooks such as Anderson and Moore (1979), Harvey (1989) and Durbin and Koopman (2001).

We denote the lag operator by L and define the lag polynomial $\Psi(L)$ as follows

$$\Psi(L) = \sum_{i=1}^{q_\Psi} \Psi_i L^i,$$

so that $\varepsilon_t = \{I - \Psi(L)\}u_t$ in (4). Applying $I - \Psi(L)$ to both sides of equation (2) and rearranging terms, we obtain

$$y_t = \mu + d_t + X_t\beta + \Lambda F_t + \varepsilon_t, \qquad \varepsilon_t \sim NIID(0, \Sigma_\varepsilon),$$

for $t = q_\Psi + 1, \ldots, T$, where $F_t = (f_t', f_{t-1}', \ldots, f_{t-q_\Psi}')'$, $\mu = (I - \sum_{i=1}^{q_\Psi} \Psi_i)\,\bar{\mu}$, $d_t = \Psi(L)y_t$,

$X_t = \{I - \Psi(L)\}\bar{X}_t$ and

$$\Lambda = \left(\bar{\Lambda}\,,\, -\Psi_1\bar{\Lambda}\,,\, \cdots\,,\, -\Psi_{q_\Psi}\bar{\Lambda}\right).$$

The stacked vector F_t can be written as $G\alpha_t$ where G is a matrix of full rank and α_t is of the form (6). Finally, we can write

$$Z = \Lambda G. \tag{7}$$

This shows that the state space model provides a general representation of dynamic factor models.

To handle the initial stretch of observations $y_1, \ldots, y_{q_\Psi}$ explicitly, we can consider the observation equation (5) with different system matrices Z and Σ_ε for $t = 1, \ldots, q_\Psi$. We give an example in Illustration 1 below. In the remainder of the paper, we assume for convenience that all system matrices are time-invariant. However, all results hold for time-varying system matrices subject to some minor modifications.

Illustration 1. *Consider the dynamic factor model $y_t = \bar{\Lambda} f_t + u_t$ with $N \times r$ factor loading matrix $\bar{\Lambda}$ and where the $r \times 1$ vector f_t follows a VAR(1) process, that is equation* (3) *with $q_\Phi = 1$ and $q_\Theta = 0$, and the idiosyncratic components u_{it} are modelled as independent AR(1) processes, that is equation* (4) *with $q_\Psi = 1$ and both Ψ_1 and Σ_ε diagonal, for $i = 1, \ldots, N$. To ensure that all parameters are identified, we restrict $\Sigma_\zeta = I - \Phi_1\Phi_1'$ in* (3) *such that $Var(f_t) = I_r$ for $t = 1, \ldots, T$. Additionally, we set $\bar{\Lambda} = (\bar{\Lambda}_1', \bar{\Lambda}_2')'$ where $\bar{\Lambda}_1$ is an $r \times r$ lower triangular matrix and $\bar{\Lambda}_2$ is an $(N - r) \times r$ full matrix. The diagonal elements of $\bar{\Lambda}_1$ are positive.*

The state vector α_t is specified as $\alpha_t = F_t = (f_t'\,,\, f_{t-1}')'$ so that $G = I$ in (7). *The observation equation* (5) *for $t = 2, \ldots, T$ has $\mu = 0$, $d_t = \Psi_1 y_{t-1}$, $\beta = 0$ and $Z = \Lambda = (\bar{\Lambda}\ ,\, -\Psi_1\bar{\Lambda})$. Since $u_1 \sim N\{0, (I - \Psi_1^2)^{-1}\Sigma_\varepsilon\}$, observation equation* (5) *for $t = 1$ has $d_1 = 0$, $Z = (\bar{\Lambda}\,,\, 0)$ and $\varepsilon_1 = u_1$. The state equation* (6) *has*

$$H = \begin{bmatrix} \Phi_1 & 0 \\ I & 0 \end{bmatrix}, \qquad R = \begin{bmatrix} I \\ 0 \end{bmatrix}, \qquad \Sigma_\eta = \Sigma_\zeta,$$

$a = 0$ and P is set to the unconditional variance of the stationary vector series $(f_t', f_{t-1}')'$.

Illustration 2. *Suppose y_t is modelled by the one-factor model $y_t = \bar{\Lambda} f_t + u_t$ with $N \times 1$ factor loading vector $\bar{\Lambda}$ and the single factor f_t is modelled as an ARMA(1,1) process defined as the autoregressive moving average process* (3) *with $r = q_\Phi = q_\Theta = 1$. We have*

$$f_t = \phi f_{t-1} + \zeta_t + \theta\zeta_{t-1}, \qquad \zeta_t \sim NIID(0, \sigma_\zeta^2),$$

where ϕ, θ *and* $\sigma_\zeta^2 > 0$ *are unknown scalar coefficients. Furthermore, suppose that the idiosyncratic components are independent Gaussian disturbances, that is* $q_\Psi = 0$ *and* $u_t = \varepsilon_t$ *in* (4). *Identifiability of all parameters is guaranteed via the restriction* $\sigma_\zeta^{-2} = 1 + \{(\theta - \phi)^2 / (1 - \phi^2)\}$ *such that* $Var(f_1) = 1$.

The state vector α_t *is specified as* $\alpha_t = (f_t\,,\, \theta\zeta)'$ *and since* $F_t = f_t$ *we have* $G = (1\,,\, 0)$ *in* (7). *Then,* $\mu = 0$, $d_t = 0$, $\beta = 0$ *and* $Z = \Lambda G$ *with* $\Lambda = \bar{\Lambda}$ *in equation* (5) *for* $t = 1, \ldots, T$. *The state equation* (6) *for* $t = 1, \ldots, T$ *has*

$$H = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix}, \qquad R = \begin{pmatrix} 1 \\ \theta \end{pmatrix}, \qquad \Sigma_\eta = \sigma_\zeta^2,$$

with initial conditions $a = 0$ *and* P *such that* $P - HPH' = \sigma_\zeta^2 RR'$.

3 Estimation of factors and likelihood evaluation

The new results are firstly presented for the dynamic factor model in state space form (5) and (6) without regression effects (and intercepts). We then consider the model with regression effects in Section 3.2. The results are relevant for likelihood evaluation and estimation of factors and regression effects.

3.1 A dynamic factor model without regression effects

The state space form of the dynamic factor model without regression effects is given by

$$y_t = d_t + Z\alpha_t + \varepsilon_t, \qquad \alpha_t = H\alpha_{t-1} + R\eta_t, \qquad \varepsilon_t \sim NIID(0, \Sigma_\varepsilon), \qquad t = 1, \ldots, T, \quad (8)$$

where y_t is the $N \times 1$ observation vector and α_t is the $p \times 1$ state vector. The loglikelihood function of $y_1, \ldots, y_T$ is obtained using the Kalman filter. The Kalman filter and smoothing (KFS) methods of Appendix A.1 can also be used to evaluate

$$a_{t|s} = \mathbb{E}(\alpha_t | y_1, \ldots y_s), \qquad P_{t|s} = \mathrm{Var}(\alpha_t | y_1, \ldots y_s), \qquad s, t = 1, \ldots, T. \quad (9)$$

Vector $a_{t|s}$ is the minimum mean squared error estimator (MMSE) of α_t conditional on $y_1, \ldots, y_s$ and matrix $P_{t|s}$ is its mean squared error. If the panel dimension is very large the KFS methods are computationally infeasible, even when p is of modest size. Anderson and Moore (1979) and Koopman and Durbin (2003) showed that, for models with diagonal Σ_ε, multivariate KFS methods can be made computationally more efficient by processing the

elements of y_t individually rather than the whole vector at once. This modification leads to substantial computational gains, but they are not sufficient for the dimensions common in recent applications of dynamic factor models.

3.1.1 Transforming the observation equation

Consider the dynamic factor model (8). Suppose $y_t^* = Ay_t$, for $t = 1, \ldots, T$, for some invertible matrix A. The MMSEs of α_t in (9) are not affected if $y_1, \ldots, y_s$ is replaced with $y_1^*, \ldots y_s^*$. Furthermore, the loglikelihood functions of $y_1, \ldots, y_T$ and $y_1^*, \ldots, y_T^*$ differ only by the Jacobian term $\log |A|^T$. We will show that for certain choices of A the loglikelihood function and the MMSEs of $\alpha_1, \ldots, \alpha_T$ can be computed more efficiently based on $y_1^*, \ldots, y_T^*$ rather than $y_1, \ldots, y_T$.

Suppose

$$A = \begin{bmatrix} A^L \\ A^H \end{bmatrix}, \qquad y_t^* = \begin{pmatrix} y_t^L \\ y_t^H \end{pmatrix}, \qquad \text{with} \quad y_t^L = A^L(y_t - d_t), \quad y_t^H = A^H(y_t - d_t), \quad (10)$$

with $N \times N$ matrix A, $m \times N$ matrix A^L and $(N-m) \times N$ matrix A^H where $m < N$. The observation vectors y_t^L and y_t^H have dimensions $m \times 1$ and $(N-m) \times 1$, respectively. We aim to choose A such that y_t^L and y_t^H are uncorrelated and only y_t^L depends on α_t. More specifically, the model for y_t^* will be of the form

$$\begin{array}{rcl} y_t^L &=& A^L Z \alpha_t + e_t^L, \\ y_t^H &=& e_t^H, \end{array} \qquad \begin{pmatrix} e_t^L \\ e_t^H \end{pmatrix} \sim NIID\left(0, \begin{bmatrix} \Sigma_L & 0 \\ 0 & \Sigma_H \end{bmatrix}\right), \quad (11)$$

where $\Sigma_L = A^L \Sigma_\varepsilon A^{L\prime}$ and $\Sigma_H = A^H \Sigma_\varepsilon A^{H\prime}$. A suitable matrix A needs to fulfill the following conditions

(i) A is full rank,

(ii) $A^H \Sigma_\varepsilon A^{L\prime} = 0$,

(iii) $\text{Row}\{A^H\} = \text{Col}\{Z\}^\perp$,

where $\text{Col}\{X\}$ and $\text{Row}\{X\}$ denote the row and column spaces of a matrix X, respectively, and the superscript $\perp$ denotes the orthogonal complement. Condition (i) prevents any loss of information due to the transformation Ay_t. Condition (ii) ensures that both equations in (11) are independent and Condition (iii) implies that the second equation does not depend on α_t. Condition (iii) is stronger than strictly necessary. The transformed model will still

be of the form (11) if Condition (iii) is replaced with $A^H Z = 0$. In its current form however, condition (iii) ensures that the reduction in dimension is as large as possible, in the sense that the dimension of y_t^H cannot be enlarged without comprimising the special form of (11). This is in general not the case if we only require $A^H Z = 0$. Finally, we add the following condition

(iv) $|\Sigma_H| = 1$.

Condition (iv) is not restrictive and it simplifies various calculations. For example, we can express the determinant of A in terms of A^L and Σ_ε since

$$|A|^2 = |\Sigma_\varepsilon|^{-1}|A\Sigma_\varepsilon A'| = |\Sigma_\varepsilon|^{-1}|A^L\Sigma_\varepsilon A^{L\prime}||A^H\Sigma_\varepsilon A^{H\prime}| = |\Sigma_\varepsilon|^{-1}|\Sigma_L|. \tag{12}$$

The Conditions (i)–(iii) imply a closed form for A^L, which is given in the following lemma.

Lemma 1. *Consider model* (8)*. Suppose a matrix A is of the form* (10)*, then A^L satisfies (i)–(iii) if and only if*

$$A^L = C\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}, \tag{13}$$

for some nonsingular $m \times m$ matrix C and for some $N \times m$ matrix $\Lambda^\dagger$, such that the columns of $\Lambda^\dagger$ form a basis for the column space of Λ.

Remark 1. It is evident that any matrix A with A^L given by (13) fulfills Conditions (i)–(iii). We prove the necessity part of the lemma in Appendix A.2.

Remark 2. From (7) we have $Z = \Lambda G$ where G has full rank. If Λ is of full column rank, we can choose $\Lambda^\dagger = \Lambda$. Column rank deficiency of Λ is rare in practice. However, if Λ does not have full column rank it is straightforward to construct a suitable $\Lambda^\dagger$.

Remark 3. A closed form expression for A^H will generally not be available. From condition (iii) it follows that A^H has to be chosen such that its rows form a basis for the null space of $\Lambda^{\dagger\prime}$. Condition (iv) can then be satisfied by rescaling the rows. Finding a basis for the null space of a matrix requires computationally intensive numerical methods. Fortunately, we can show that matrix A^H is not required for any of our computations.

Remark 4. The results below are based on transformation (10) and model (11). Although our results are more general and are developed for different purposes, a similar transformation as (10) for a different class of factor models is considered by Fiorentini, Sentana, and Shephard (2004, section 2.4.1).

Illustration 3. *Consider the one-factor model* $y_t = \bar{\Lambda} f_t + \varepsilon_t$ *of Illustration 2. Apply transformation* (10) *to* y_t *where the row vector* A^L *is given by* (13) *with scalar* $C = (\Lambda' \Sigma_\varepsilon^{-1} \Lambda)^{-1}$ *and* $\Lambda^\dagger = \Lambda = \bar{\Lambda}$. *For this choice of* C, *the scalar* y_t^L *is effectively the generalised least squares (GLS) estimator of* f_t *in the "regression model"* $y_t = \Lambda f_t + \varepsilon_t$, *for a given* t. *We have*

$$y_t^L = \left(\Lambda' \Sigma_\varepsilon^{-1} \Lambda\right)^{-1} \Lambda' \Sigma_\varepsilon^{-1} y_t, \qquad t = 1, \ldots, T.$$

Model (11) *for the univariate time series* y_t^L *is then given by*

$$y_t^L = G\alpha_t + e_t^L, \qquad e_t^L \sim NIID(0, C), \qquad t = 1, \ldots, T.$$

3.1.2 Estimation of factors

By considering a matrix A that satisfies the Conditions (i)–(iv) of Section 3.1.1, we can efficiently compute MMSE estimators of the factors and their mean squared errors. Since A has full rank it follows that $\mathbb{E}(\alpha_t | y_1, \ldots, y_s) = \mathbb{E}(\alpha_t | y_1^*, \ldots, y_s^*)$ and $\mathrm{Var}(\alpha_t | y_1, \ldots, y_s) = \mathrm{Var}(\alpha_t | y_1^*, \ldots, y_s^*)$. Furthermore, from (11) it follows that y_t^L and y_t^H are independent of each other and y_t^H does not depend on α_t. Hence,

$$a_{t|s} = \mathbb{E}(\alpha_t | y_1^*, \ldots, y_s^*) = \mathbb{E}(\alpha_t | y_1^L, \ldots, y_s^L), \quad P_{t|s} = \mathrm{Var}(\alpha_t | y_1^*, \ldots, y_s^*) = \mathrm{Var}(\alpha_t | y_1^L, \ldots, y_s^L),$$

for $s, t = 1, \ldots, T$. The MMSEs of the states can therefore be obtained by applying KFS methods to the low-dimensional model

$$y_t^L = A^L Z \alpha_t + e_t^L, \qquad e_t^L \sim NIID(0, \Sigma_L), \qquad t = 1, \ldots, T. \tag{14}$$

The high-dimensional matrix A^H and vector y_t^H are not required for the estimation of α_t. In case of Illustration 3, the estimation of α_t is simply carried out by univariate KFS methods.

3.1.3 Loglikelihood evaluation

Let $\ell(y)$ denote the loglikelihood function for the linear Gaussian state space model (5) – (6), where $y = (y_1', \ldots, y_T')'$. The loglikelihood $\ell(y)$ can be evaluated via the prediction error decomposition, see Schweppe (1965). The prediction error vector v_t and its variance matrix D_t, obtained from the Kalman filter of Appendix A.1 applied to model (8), are required to calculate

$$\ell(y) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log |D_t| - \frac{1}{2} \sum_{t=1}^{T} v_t' D_t^{-1} v_t, \tag{15}$$

see also Harvey (1989, section 3.4).

A computationally more efficient way to evaluate (15) is to choose a matrix A that satisfies the Conditions (i)–(iv) of Section 3.1.1, transform y_t as in (10) and to consider model (11). We then have

$$\ell(y) = \ell(y^L) + \ell(y^H) + T \log |A|, \tag{16}$$

where $y^L = (y_1^{L\prime}, \ldots, y_T^{L\prime})'$ and $y^H = (y_1^{H\prime}, \ldots, y_T^{H\prime})'$. The first term $\ell(y^L)$ can be evaluated by the Kalman filter applied to the low-dimensional model (14). The second term is given by

$$\ell(y^H) = -\frac{(N-m)T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} y_t^{H\prime} \Sigma_H^{-1} y_t^H, \tag{17}$$

since $|\Sigma_H| = 1$. Lemma 2 shows that the last term in equation (17) can be calculated without constructing A^H. The proof is given in Appendix A.3.

Lemma 2. *Consider the dynamic factor model* (8)*, transformation* (10) *and model* (11)*. For definition* (13) *of matrix* A^L*, we have*

$$y_t^{H\prime} \Sigma_H^{-1} y_t^H = e_t' \Sigma_\varepsilon^{-1} e_t, \tag{18}$$

where $e_t = y_t - \Lambda^\dagger \left(\Lambda^{\dagger\prime} \Sigma_\varepsilon^{-1} \Lambda^\dagger\right)^{-1} \Lambda^{\dagger\prime} \Sigma_\varepsilon^{-1} y_t$ *is the GLS residual for data-vector* y_t*, covariate matrix* $\Lambda^\dagger$ *and variance matrix* Σ_ε*. The choice of* C *in* (13) *is irrelevant.*

Given the expression for $|A|^2$ in (12), loglikelihood function (16) can be expressed as

$$\ell(y) = c + \ell(y^L) - \frac{T}{2} \log \frac{|\Sigma_\varepsilon|}{|\Sigma_L|} - \frac{1}{2} \sum_{t=1}^{T} e_t' \Sigma_\varepsilon^{-1} e_t, \tag{19}$$

where c is a constant independent of both y and the parameters. It follows that for the evaluation of the loglikelihood, computation of matrix A^H and vectors y_t^H $(t = 1, \ldots, T)$ is not required. Expression (19) is instrumental for a computationally feasible approach to the likelihood-based analysis of the dynamic factor model.

Illustration 4. *In Illustration 3, the transformation* (10) *is based on matrix* A^L *defined in* (13) *with* $C = (\Lambda^{\dagger\prime} \Sigma_\varepsilon^{-1} \Lambda^\dagger)^{-1}$*. It is more convenient to choose* C *such that* $C'C = (\Lambda^{\dagger\prime} \Sigma_\varepsilon^{-1} \Lambda^\dagger)^{-1}$ *with* C *upper-triangular. For this choice, the variance matrix* Σ_L *in* (14)

is the identity matrix and the loading matrix in (14) *is* $A^L\Lambda^\dagger = C^{-1\prime}$. *We obtain the model*

$$y_t^L = C^{-1\prime}G\alpha_t + e_t^L, \qquad e_t^L \sim NIID(0, I), \qquad t = 1, \ldots, T.$$

Since $\Sigma_L = I$, *the fast KFS methods discussed in Koopman and Durbin (2003) can be applied straightforwardly. Furthermore, the loglikelihood function* (19) *reduces to*

$$\ell(y) = c + \ell(y^L) - \frac{T}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\sum_{t=1}^{T} e_t'\Sigma_\varepsilon^{-1}e_t.$$

The computations for $|\Sigma_\varepsilon|$ *and* Σ_ε^{-1} *can exploit special structures in* Σ_ε *such as the matrix being diagonal or having Toeplitz, spatial or block structures.*

3.2 Factor model with intercepts and regression effects

Consider the state space model (5) and (6) with intercepts and regression effects μ and β. Suppose μ and β are treated as random variables $(\mu', \beta')' \sim N(0, \kappa I_{N+K})$ for some $\kappa > 0$. In this case, μ and β can be added to the state vector and estimated by the Kalman filter. If $\kappa \to \infty$, the Kalman filter estimators for μ and β converge to their minimum mean squared error linear estimators (MMSLE), see *e.g.* Shephard (1993). Such random variables with arbitrarily large variance matrices are said to be diffuse.

Following Ansley and Kohn (1985) and de Jong (1991), we define the diffuse likelihood for the model (5) and (6) with regression effects μ and β by

$$p_d(y) \stackrel{def.}{=} \lim_{\kappa\to\infty} \kappa^{-(N+K)/2} \iiint p(y|\mu, \beta, \alpha)p(\alpha)p_\kappa(\mu, \beta)\, d\mu\, d\beta\, d\alpha. \tag{20}$$

The diffuse likelihood $p_d(y)$ can be evaluated using a modified version of the Kalman filter, see Appendix A.1. Unknown parameters can be estimated by maximizing the diffuse likelihood. Tunnicliffe-Wilson (1989) and Shephard (1993) show that for small sample sizes such estimators are preferred over maximum likelihood estimators. This is especially important in macroeconomic applications, where the time series dimension can be small. Quah and Sargent (1993), for example, consider a data set where T is only 42.

3.2.1 Transforming the observation equation

To facilitate evaluation of the diffuse likelihood and estimation of the regression coefficients in the framework of Section 3.1 we proceed as follows. Choose matrix A such that Conditions (i)–(iii) of Section 3.1.1 are satisfied. Pre-multiplying the observations with A, we obtain

the model

$$\begin{aligned} y_t^L &= \mu^L + X_t^L\beta + A^L Z\alpha_t + e_t^L, \\ y_t^H &= A^H\mu + X_t^H\beta + e_t^H, \end{aligned} \qquad \begin{pmatrix} e_t^L \\ e_t^H \end{pmatrix} \sim NIID\left(0, \begin{bmatrix} \Sigma_L & 0 \\ 0 & \Sigma_H \end{bmatrix}\right), \tag{21}$$

where $\mu^L = A^L\mu$, $X_t^L = A^L X_t$ and $X_t^H = A^H X_t$ and the other variables are defined below (11). We show next that both estimation of μ, β and α_t and likelihood evaluation can be carried out in two steps: first processing the original time series y_t and second applying KFS methods to the time series y_t^L. The computational gains of Section 3.1 can therefore be preserved.

3.2.2 Estimation of intercepts, regression coefficients and factors

Denote the MMSLEs of μ, β and $\alpha_1, \ldots, \alpha_T$, based on y, by $\widehat{\mu}$, $\widehat{\beta}$ and $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_T$. We develop a two-step procedure for the computation of these MMSLEs by treating the two equations of (21) separately. Since we are interested in the MMSLE of μ, not of μ^L and $A^H\mu$ in (21), and since β appears in both equations of (21), we propose the following method of estimation. Define

$$P_\Lambda = \Lambda^\dagger(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}C^{-1}, \qquad M_\Lambda = I - \Lambda^\dagger(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}.$$

We have $P_\Lambda A^L = I - M_\Lambda$ and $\mu = \mu^H + P_\Lambda\mu^L$ where $\mu^H \overset{def.}{=} M_\Lambda\mu$. The MMSLE of μ is then given by

$$\widehat{\mu} = \widehat{\mu}^H + P_\Lambda\widehat{\mu}^L, \tag{22}$$

with mean squared error matrix

$$\begin{aligned} \mathrm{Var}(\widehat{\mu}) &= \mathrm{Var}(\widehat{\mu}^H + P_\Lambda\widehat{\mu}^L) \\ &= \mathrm{Var}(\widehat{\mu}^H) + P_\Lambda\mathrm{Cov}(\widehat{\mu}^L, \widehat{\mu}^H) + \mathrm{Cov}(\widehat{\mu}^H, \widehat{\mu}^L)P_\Lambda' + P_\Lambda\mathrm{Var}(\widehat{\mu}^L)P_\Lambda', \end{aligned} \tag{23}$$

where $\widehat{\mu}^L$ and $\widehat{\mu}^H$ are estimates of μ^L and μ^H, respectively.

Our two-step procedure for computing the MMSLEs $\widehat{\mu}^L$, $\widehat{\beta}$ and $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_T$, together with their mean squared error matrices, consists of:

Step 1. Compute

$$b = B^{-1}\sum_{t=1}^{T}\widetilde{X}_t'\Sigma_\varepsilon^{-1}\widetilde{y}_t, \qquad B = \sum_{t=1}^{T}\widetilde{X}_t'\Sigma_\varepsilon^{-1}\widetilde{X}_t, \tag{24}$$

where $\widetilde{y}_t$ and $\widetilde{X}_t$ are given by

$$\widetilde{y}_t = M_\Lambda(y_t - d_t - \bar{y}), \qquad \widetilde{X}_t = M_\Lambda(X_t - \bar{X}), \tag{25}$$

for $t = 1, \ldots, T$, and

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T}(y_t - d_t), \qquad \bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t. \tag{26}$$

Step 2. Apply KFS methods to the state space model with observation equation

$$y_t^L = \mu^L + X_t^L\beta + A^L Z\alpha_t + e_t^L, \qquad e_t^L \sim NIID(0, \Sigma_L), \qquad t = 1, \ldots, T, \tag{27}$$

where μ^L has the diffuse prior distribution $\mu^L \sim N(0, \kappa I_N)$, with $\kappa \to \infty$, β has the initial distribution $\beta \sim N(b, B^{-1})$ while the initial state vector has $\alpha_1 \sim N(a, P)$. The KFS produces $\widehat{\beta}$, $\widehat{\mu^L}$ and $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_T$ and the mean squared errors of these estimators.

Step 1 relies on standard regression computations while Step 2 applies the KFS to the low-dimensional model (27) with the state vector α_t extended by μ^L and β.

We prove the validity of the two-step procedure in Appendix A.4. It is also shown that

$$\widehat{\mu}^H = M_\Lambda(\bar{y} - \bar{X}\widehat{\beta}), \qquad \mathrm{Var}(\widehat{\mu}^H) = M_\Lambda\{\bar{X}\mathrm{Var}(\widehat{\beta})\bar{X}' + T^{-1}\Sigma_\varepsilon\}M_\Lambda'. \tag{28}$$

and

$$\mathrm{Cov}(\widehat{\mu}^H, \widehat{\mu}^L) = -M_\Lambda\bar{X}\mathrm{Cov}(\widehat{\beta}, \widehat{\mu}^L). \tag{29}$$

These results enable us to obtain $\widehat{\mu}$ from (22) and its mean squared error matrix from (23).

3.2.3 Likelihood evaluation

The diffuse loglikelihood function for the dynamic factor model with fixed effects and regression effects is computed as part of the two step estimation procedure of Section 3.2.2. Suppose $\ell_d(y^L)$ is the diffuse likelihood of a state space model consisting of (27) and (6). We have

$$\ell_d(y) = \ell_d(y^L) + \ell_d(y^H) - \frac{1}{2}(T-1)\log\frac{|\Sigma_\varepsilon|}{|\Sigma_L|}, \tag{30}$$

where $\ell_d(y^H)$ is the diffuse likelihood for the model of y^H in equation (21) and is given by

$$\ell_d(y^H) = -\frac{(N-m)}{2}(T\log 2\pi + \log T) - \frac{T}{2}\log|B| - \frac{1}{2}\sum_{t=1}^{T}\widetilde{y}_t'\Sigma_\varepsilon^{-1}\widetilde{y}_t, \tag{31}$$

where B and $\widetilde{y}_t$ are defined in (24) and (25), respectively. We prove expressions (30) and (31) in Appendix A.5. Note that the evaluation of $\ell_d(y)$ does not require y_t^H.

3.3 Computational gains

The main purpose of the results of the previous sections is to obtain computationally efficient inference procedures for the class of dynamic factor models discussed in Section 2. In this section we report the gains in computing times that are achieved by our new methods.

The computational gains depend primarily on the panel dimension N and state vector dimension p. To obtain some insight in the size of this gain we calculate the likelihood, respectively diffuse likelihood, for different values of N and p using both the standard Kalman filter as well as the methods of Sections 3.2.3 and 3.2.3. In Table 1 we present the ratios of the CPU times of these methods. The results are encouraging. For a typical model without a constant and with $N = 250$ and $p = 5$, the Kalman filter computations are carried out 15 times faster as a result of our new device. We further see that the computational savings are substantial for moderate values of N and relatively small values of p, say, 5 or 10. If p is relatively large, say, 25, the gains are less dramatic but still substantial by any means.

We also achieve substantial computational gains if we apply the method of Section 3.2.3 to dynamic factor models with a constant vector μ, that is state space model (5) and (6) with $\beta = 0$. The reported ratios are high because applying standard KFS methods to this model requires the augmentation of the $p \times 1$ state vector α_t by the $N \times 1$ constant vector μ. The augmentation of the state vector for the model (27) and (6) is limited to the $r \times 1$ constant vector μ^L.

Table 1: Computational Gains

The two panels below present the gains in computing time when evaluating the loglikelihood respectively the diffuse loglikelihood functions of two types of dynamic factor models. Model A is of the form $y_{it} = \lambda_i' f_t + \varepsilon_{it}$ and model B of the form $y_{it} = \mu_i + \lambda_i' f_t + \varepsilon_{it}$, where f_t is a VAR(1), $\varepsilon_{it} \sim NIID(0, \sigma^2)$, for some positive scalar σ and μ_i is a scalar. The ratio d_1/d_2 is reported: d_1 is the CPU time for the standard (diffuse) Kalman filter and d_2 is CPU time for the algorithms of Sections 3.1.3 and 3.2.3. The ratios are reported for different panel dimensions N and different state vector dimensions p.

	Model A					Model B				
$N \backslash p$	1	5	10	25	50	1	5	10	25	50
10	2.0	1.3	–	–	–	10.4	2.3	–	–	–
50	5.7	4.7	3.1	1.5	–	50.6	40.0	18.0	3.4	–
100	6.7	7.5	5.6	2.5	1.5	55.0	62.0	47.2	13.5	3.2
250	8.7	14.8	12.4	5.5	3.0	79.0	82.2	82.9	63.6	22.6
500	12.5	15.9	21.2	10.2	5.4	107.5	108.9	109.5	108.7	69.7

4 Likelihood-based methods for estimation

In this section we discuss procedures for the estimation of the unknown vectors and matrices λ_i $(i = 1, \ldots, N)$, Φ_i $(i = 1, \ldots, q_\Phi)$, Θ_i $(i = 1, \ldots, q_\Theta)$, Ψ_i $(i = 1, \ldots, q_\Psi)$ and Σ_ε in the dynamic factor model of Section 2.1. These vectors and matrices depend on a vector of unknown parameters that we denote by ψ. The dimension of ψ can be as high as $1,000$ or $2,000$. The results of Section 3 imply that likelihood evaluation can be fast for high-dimensional dynamic factor models. We show in Section 4.1.1 that the new results also lead to a very fast procedure for computing the score vector of ψ. Section 4.1.2 shows that the same arguments can apply to the EM method of estimation. Finally, we show in Section 4.2 that the new results can also be useful for Bayesian methods.

4.1 Maximum likelihood estimation

Numerical optimization procedures, such as the quasi-Newton BFGS algorithm, see *e.g.* Nocedal and Wright (1999), can be adopted to maximize the loglikelihood function with respect to ψ. These methods require evaluation of the score vector. Since the dimension of ψ is high, computing the score vector numerically is infeasible, even if the results of Section 3 are used. Fortunately, we can show that the exact score vector can be obtained by a single KFS applied to the low-dimensional model (14) or (27).

Alternatively, the EM algorithm can be used to obtain the maximum likelihood estimator of ψ. In Section 4.1.2 we show that each EM step also relies on a single KFS.

4.1.1 Exact score for models without regression effects

Koopman and Shephard (1992) develop analytical expressions for the score function of the parameters in a state space model. They adopt the results in Louis (1982) and Ruud (1991) and in particular the identity

$$\frac{\partial \ell(y;\psi)}{\partial \psi}\bigg]_{\psi=\psi^*} = \frac{Q(\psi^*|\psi)}{\partial \psi}\bigg]_{\psi=\psi^*}, \tag{32}$$

where $Q(\psi^*|\psi)$ is the expected complete loglikelihood function, given by

$$Q(\psi^*|\psi) = \mathbb{E}\left(\log p(y, \alpha; \psi)\,|y; \psi^*\right),$$

and $p(y, \alpha; \psi)$ is the joint density of y and $\alpha_1, \ldots, \alpha_T$. For the state space model (5) and (6), $Q(\psi^*|\psi)$ is given by

$$Q(\psi^*|\psi) = c - \frac{T}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\text{tr}Q_\varepsilon - \frac{T-1}{2}\log|\Sigma_\eta| - \frac{1}{2}\text{tr}Q_\eta \\ - \frac{1}{2}\log|P| - \frac{1}{2}\text{tr}[P^{-1}\{(a_{1|T} - a)(a_{1|T} - a)' + P_{1|T}\}], \tag{33}$$

where c is a constant independent of ψ and

$$Q_\varepsilon = \Sigma_\varepsilon^{-1}\sum_{t=1}^{T}\{\widehat{\varepsilon}_t\widehat{\varepsilon}_t' + \text{Var}(\varepsilon_t|y)\}, \qquad Q_\eta = \Sigma_\eta^{-1}\sum_{t=2}^{T}\{\widehat{\eta}_t\widehat{\eta}_t' + \text{Var}(\eta_t|y)\}, \tag{34}$$

where $\widehat{\varepsilon}_t = \text{E}(\varepsilon_t|y)$, $\text{Var}(\varepsilon_t|y)$, $\widehat{\eta}_t = \text{E}(\eta_t|y)$ and $\text{Var}(\eta_t|y)$ can be expressed in terms of $a_{j|T}$ and $P_{j|T}$ for $j = 1, \ldots, T$, which can be evaluated using the KFS methods discussed in Appendix A.1. Since the estimation of factors can be based on the low-dimensional model (14) while matrix A^H and time series y_t^H are not needed, the KFS computations are fast. Expressions for the derivatives of (33) with respect to the system matrices, evaluated at $\psi = \psi^*$, are given in Appendix A.6. The score vector of ψ is then obtained via the chain rule.

4.1.2 The EM algorithm for models without regression effects

The EM algorithm, introduced by Dempster, Laird, and Rubin (1977), is an iterative algorithm that repeatedly performs two types of calculations: (E)xpectation and (M)aximization. For a given value of $\psi = \psi^*$, the E and M steps are given by

- E step: determine the expected complete loglikelihood function $Q(\psi^*|\psi)$ in (33).
- M step: maximize $Q(\psi^*|\psi)$ with respect to ψ.

The M step produces a vector ψ^+ with the property $\ell(y; \psi^+) \geq \ell(y; \psi^*)$. If the EM steps are continuously repeated, convergence to a (local) optimum of $\ell(y; \psi)$ is guaranteed, see Wu (1983) for a more detailed discussion. Shumway and Stoffer (1982) and Watson and Engle (1983) have proposed the use of the EM algorithm in the context of state space models. The details of the EM algorithm are specific to the particular specification of the dynamic factor model. In Appendix A.7 the EM algorithm is reviewed for the model of Illustration 1.

4.1.3 Estimation for models with intercepts and regression effects

For the state space model (5) and (6) with regression effects, the score of the diffuse likelihood can be obtained as in Section 4.1.1. The only difference being that $Q(\psi^*|\psi)$ is evaluated using the methods of Section 3.2.2, see Durbin and Koopman (2001, section 7.3). Similarly, using this expected complete diffuse loglikelihood in the E step of the EM algorithm, the algorithm will converge to the maximum of the diffuse likelihood. The MMSLEs of μ and β and the corresponding mean squared errors are obtained from the diffuse Kalman smoother.

4.2 Bayesian Inference

As an alternative to the maximum likelihood estimation of Section 4.1 we can assume prior distributions for the parameters, $\bar{\mu}$, β and ψ, and perform a Bayesian analysis. Examples of Bayesian approaches to dynamic factor models are Aguilar and West (2000) and Fiorentini, Sentana, and Shephard (2004) in the context of modelling volatility in time series. We follow these contributions in adopting Markov chain Monte Carlo (MCMC) methods because closed form expressions of posterior densities are not available. The MCMC method samples a Markov chain that has the posterior density of the parameters as its stationary distribution. After a burn-in period, the samples can therefore be used as correlated draws from the posterior distributions. Reviews of MCMC algorithms are given by, amongst others, Gilks, Richardson, and Spiegelhalter (1996) and Chib (2001).

4.2.1 A basic MCMC implementation

Consider the dynamic factor model (2), (3) and (4). Denote the set of parameters in $\Phi_1, \ldots, \Phi_{q_\Phi}$ by Φ, the set of parameters in $\Theta_1, \ldots, \Theta_{q_\Theta}$ by Θ and those in $\Psi_1, \ldots, \Psi_{q_\Psi}$ by Ψ. The other parameters are $\bar{\mu}$, β, $\bar{\Lambda}$ and Σ_ε. The set of factors $f_1, \ldots, f_T$ is denoted by f. A typical MCMC algorithm for the dynamic factor model is given by:

(i) Initialize f, Φ, Θ, Ψ, $\bar{\mu}$, β, $\bar{\Lambda}$ and Σ_ε.

(ii) Sample f from $p(f|\bar{\Lambda}, \Phi, \Theta, \Psi, \Sigma_\varepsilon, \bar{\mu}, \beta, y)$.

(iii) Sample Φ, Θ from $p(\Phi, \Theta|f)$.

(iv) Sample $\bar{\mu}$, β, $\bar{\Lambda}$, Ψ and Σ_ε from $p(\bar{\mu}, \beta, \Sigma_\varepsilon, \bar{\Lambda}, \Psi|f, y)$.

(v) Goto (ii).

Steps (iii) and (iv) require the sampling from posterior densities associated with a VARMA model and a VAR model with regression effects, respectively. For more details on the implementation in these specific cases, we refer to the Bayesian literature on linear models, see, for example, Chib (1993).

4.2.2 MCMC step (ii)

The main computational difficulty is step (ii) and is caused by the high dimension of the observation vector y. Fortunately, we can use the methods of Section 3 to perform this step in a computationally efficient way. Since the dynamic factor model is a special case of the linear Gaussian state space model, the sampling step (ii) reduces to sampling the state vector from the conditional density $p(\alpha|y;\psi)$. We can sample the states α_t conditional on the observations, y, using algorithms as in Fruhwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002).

The algorithm of Durbin and Koopman (2002) for sampling states is fast and easy to implement. Consider the dynamic factor model in state space form. Suppose $\alpha_1^+, \ldots, \alpha_T^+$ and $y_1^+, \ldots, y_T^+$ are samples from the unconditional density $p(\alpha, y;\psi)$ where $\alpha = (\alpha_1', \ldots, \alpha_T')'$. A sample $\widetilde{\alpha}$ from $p(\alpha|y;\psi)$ is then constructed as follows

$$\widetilde{\alpha}_t = \alpha_t^+ - \mathbb{E}(\alpha_t|y_1^+, \ldots, y_T^+;\psi) + \mathbb{E}(\alpha_t|y_1, \ldots, y_T;\psi), \qquad t = 1, \ldots, T. \tag{35}$$

By choosing a matrix A such that it fulfills the conditions of Section 3.1.1, we have $p(\alpha|y;\psi) = p(\alpha|y^L;\psi)$. Therefore, we can obtain the samples $\widetilde{\alpha}_t$ by sampling $\alpha_1^+, \ldots, \alpha_T^+$ and $y_1^{L+}, \ldots, y_T^{L+}$ from the state space model with observation equation (11) and setting

$$\widetilde{\alpha}_t = \alpha_t^+ - \mathbb{E}(\alpha_t|y_1^{L+}, \ldots, y_n^{L+};\psi) + \mathbb{E}(\alpha_t|y_1^L, \ldots, y_n^L;\psi), \qquad t = 1, \ldots, T. \tag{36}$$

This is computationally more efficient than the original algorithm since all computations are done using the low dimensional vectors y_t^L, see the comparisons in Section 3.3.

To reduce correlation between subsequent draws, it is in general preferable to sample $\bar{\mu}$ and β in one block with the factors. This can be easily achieved using state space methods. Suppose δ is a a vector of regression coefficients, Durbin and Koopman (2002) show that we can sample from $p(\delta|y)$, by adding the variables to the state and using the diffuse KFS to evaluate equation (35). The unconditional samples δ^+ are set to zero. This suggests an efficient method to sample from $p(\bar{\mu}, \beta, f|y, \psi)$ in two steps. First, perform the two steps of Section 3.2.2. Second, obtain samples $\widetilde{\mu}$, $\widetilde{\beta}$ and $\widetilde{f}$ by

- Sample $\widetilde{\mu}^L$, $\widetilde{\beta}$ and $\widetilde{f}$ conditional on y^L in the model defined in (27).

- Set $\widetilde{\mu} = \widetilde{\mu}^H + P_\Lambda \widetilde{\mu}^L$, where $\widetilde{\mu}^H = M_\Lambda(\bar{y} - \bar{X}\tilde{\beta} + \xi)$ and $\xi \sim N(0, T^{-1}\Sigma_\varepsilon)$.

This algorithm is easily modified to allow informative Gaussian priors for μ and β.

5 An empirical illustration

In this section we present an illustration of the likelihood-based treatment of the dynamic factor model. We consider the data set of Stock and Watson (2005)[1]. From this data set we constructed a balanced panel of $N = 132$ monthly US macroeconomic time series from 1960:1 through 2003:12 (44 years, $T = 528$). The data is transformed and differenced to obtain a stationary set of time series; the details of each series and its transformation are given in Appendix A of Stock and Watson (2005). The 132 series are categorized into the 15 sectors presented in Table 2. Each sector is indexed by a letter. Table 2 also contains the number of time series in each sector. For all series, observations larger than 6 times the standard deviation of the series, σ, (in absolute value) are set to $\pm 6\sigma$. In total, 46 (out of $69,696$) observations are Winsorized in this way (0.066%). Subsequently, each time series is scaled such that its sample variance equals one.

The empirical analysis below differs from Stock and Watson (2005) and the related study in Stock and Watson (2002b) since these studies are based on a principal components analysis (diffusion indexes). Our approach is closer in spirit to the likelihood-based analyses of Bernanke, Boivin, and Eliasz (2005) and Boivin and Giannoni (2006). Maximum likelihood estimates are asymptotically efficient provided that the model is not misspecified. In this section we show that standard diagnostic tests for model misspecification can be carried out. Tests for model specification can be based on standard likelihood ratio statistics. Furthermore, the dynamic properties of the factors can be analyzed by investigating the estimated coefficients that are associated with the factors.

We consider the dynamic factor model discussed in Illustration 1 with intercept $\bar{\mu}$

$$y_t = \bar{\mu} + \bar{\Lambda} f_t + u_t, \qquad f_t = \Phi_1 f_{t-1} + \eta_t, \qquad u_t = \Psi_1 u_{t-1} + \varepsilon_t. \tag{37}$$

We treat the unknown intercepts, $\bar{\mu}$, as diffuse and use the algorithm of Section 3.2. To ensure identifiability of all parameters we adopt the restrictions set forth in Illustration 1. We note that $\mathrm{Var}(f_t) = I$ for $t = 1, \ldots, T$. We have implemented the Bai and Ng (2002) procedure for determining the number of factors based on their principal components analysis. We confirm the findings in Stock and Watson (2005) and find seven factors. Therefore, we include seven

[1]We thank Mark W. Watson for kindly making the data set available on his website.

TABLE 2: LIST OF SECTORS

This table lists the 15 sectors in the data set that we consider in Section 5. For each sector, the code, a short description and the number of series in the sector are given. Details of the 132 time series can be found in Appendix A of Stock and Watson (2005).

Code	Description	Number of Time Series
A	Real Output and Income	17
B	Employment and Hours	30
C	Real Retail	1
D	Manufacturing and Trade Sales	1
E	Consumption	1
F	Housing Starts and Sales	10
G	Real Inventories	3
H	Orders	7
I	Stock Prices	4
J	Exchange Rates	5
K	Interest Rates and Spreads	17
L	Money and Credit Quantity Aggregates	11
M	Price Indexes	21
N	Average Hourly Earnings	3
O	Miscellanea	1

factors in our dynamic factor model ($r = 7$).

The parameter vector ψ consists of elements of $\bar{\Lambda}$ and Φ_1, and the diagonal elements of Ψ_1 and Σ_ε. The number of coefficients to estimate in $\bar{\Lambda}$ is $Nr - r(r-1)/2 = 903$, in Φ_1 is $r^2 = 49$ and in the diagonal matrices Ψ_1 and Σ_ε is $2N = 264$. The dimension of ψ is therefore 1216. Further, we estimate 132 intercepts, one for each time series. The dynamic factor model is formulated in state space form with a state vector of dimension $p = 14$; see Illustration 1. Since the intercepts are treated as diffuse variables we need to add the 14 elements of μ^L to the state vector.

5.1 Parameter estimates

We have estimated the parameters by the method of maximum likelihood using the results of Sections 3.2 and 4.1. First we used the EM algorithm to find a point in the neigbourhood of the optimum. We then used the BFGS algorithm to find the maximum likelihood estimates, starting from the value of the final iteration of the EM algorithm. On a standard computer with 3 GB memory and a 2.2 GHz two-core processor this took less than 10 minutes. It is encouraging that the numerical maximization routine (based on the analytical score calculations of Appendix A.6) converges rapidly to an optimum in a parameter space

of 1216.

Figure 1 presents the estimates of the parameters in the diagonal matrices (i) Ψ_1 and (ii) Σ_ε. The estimated idiosyncratic autoregressive coefficients in Ψ_1 in Panel (i) are generally small in the real sectors (A) real output and income and (B) employment and hours, implying that the dynamic latent factors describe the dynamics appropriately for the series in these sectors. This is not the case for series in sectors (F) housing starts and sales, (G) real inventories and, possibly, (L) money and credit quantity aggregates, since most of these estimated autoregressive coefficients exceed 0.75. We conclude that the dynamics in the (F), (G) and (L) sectors can not be fully captured by the seven dynamic factors. The estimates in Ψ_1 for the series in the other sectors show a more erratic picture although the majority have modest values. Panel (ii) of Figure 1 provides a graphical display of the estimates of the diagonal elements of the variance matrix Σ_ε.

Given the different scalings of the factors, the actual estimates of $\bar{\Lambda}$ are of minor interest. Stock and Watson (2002b) present the R^2 goodness-of-fit statistic obtained from regressing y_{it} $(i = 1, \ldots, N)$ on a constant and each of the principal component estimates (diffusion indexes). These R^2 statistics are then regarded as proxies for the correlations (in absolute values) between the series and each principal component. In our modelling framework, we can evaluate the correlations between the series and each factor directly. The model-based correlations are presented in Figure 2 for the seven dynamic factors. The clustering of correlations within sectors is clearly visible. The first factor is mostly correlated with the real variables, in particular those variables in the sectors (A) real output and income, (B) employment and hours (G) real inventories and (H) orders. We will therefore refer to this factor as the "business cycle" factor. The second factor is most prominently represented by the sectors (I) stock prices and (M) price indexes. We therefore label the second factor "inflation". The correlations with the third factor resemble those with the inflation factor although they are much smaller in the sector (I) stock prices and larger for the employment series in sector (B). Since the correlations between the third factor and both the employment and inflation series have the same sign but mostly an opposite sign with the unemployment series in sector (B), we label this factor the "Phillips curve" factor. The fourth factor mostly represents the industrial production indexes in sector (A). The money supply and monetary base variables in sector (L) are strongly correlated with the fifth factor. The sixth and seventh factors are strongly associated with sector (K) interest rates and spreads while the sixth factor is also correlated with some material and manifacturing series in sector (A). The sector (F) housing starts and sales is hardly represented by any of the factors.

Table 3 presents the maximum likelihood estimates of the VAR coefficients in Φ_1 together

FIGURE 1: MAXIMUM LIKELIHOOD ESTIMATES OF AR PARAMETERS AND VARIANCES

In this figure we present maximum likelihood estimates of the diagonal elements of (i) Ψ_1 and (ii) Σ_ε in model (37). The estimates are grouped in sectors (A)-(O) of Table 2.

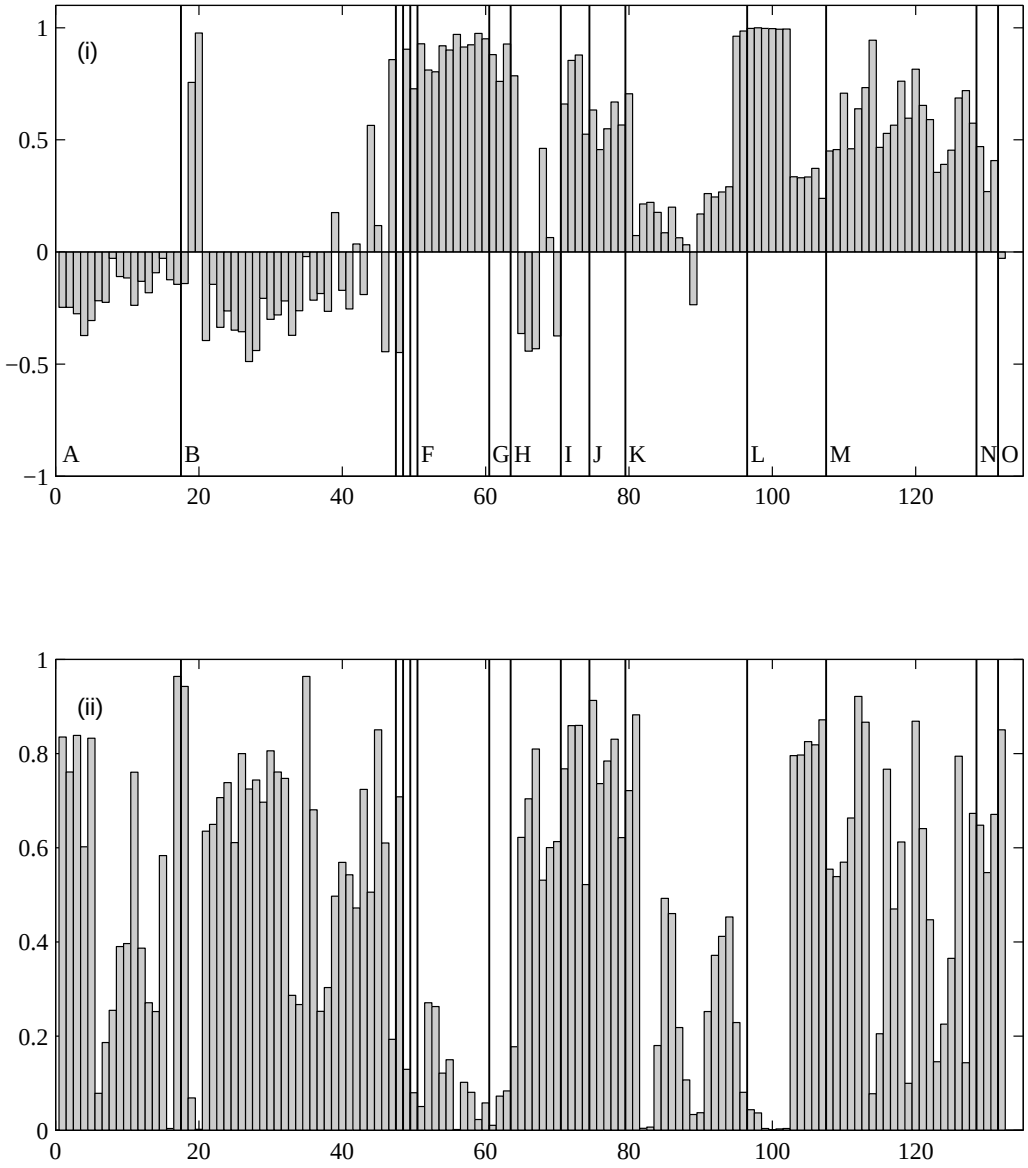


Figure 2: Correlations between observations and factors

In this figure we present the uncondional correlations between the observed series and the seven latent dynamic factors. The correlations presented are those implied by the maximum likelihood estimates of Φ_1, Ψ_1, Σ_ϵ and $\bar{\Lambda}$.

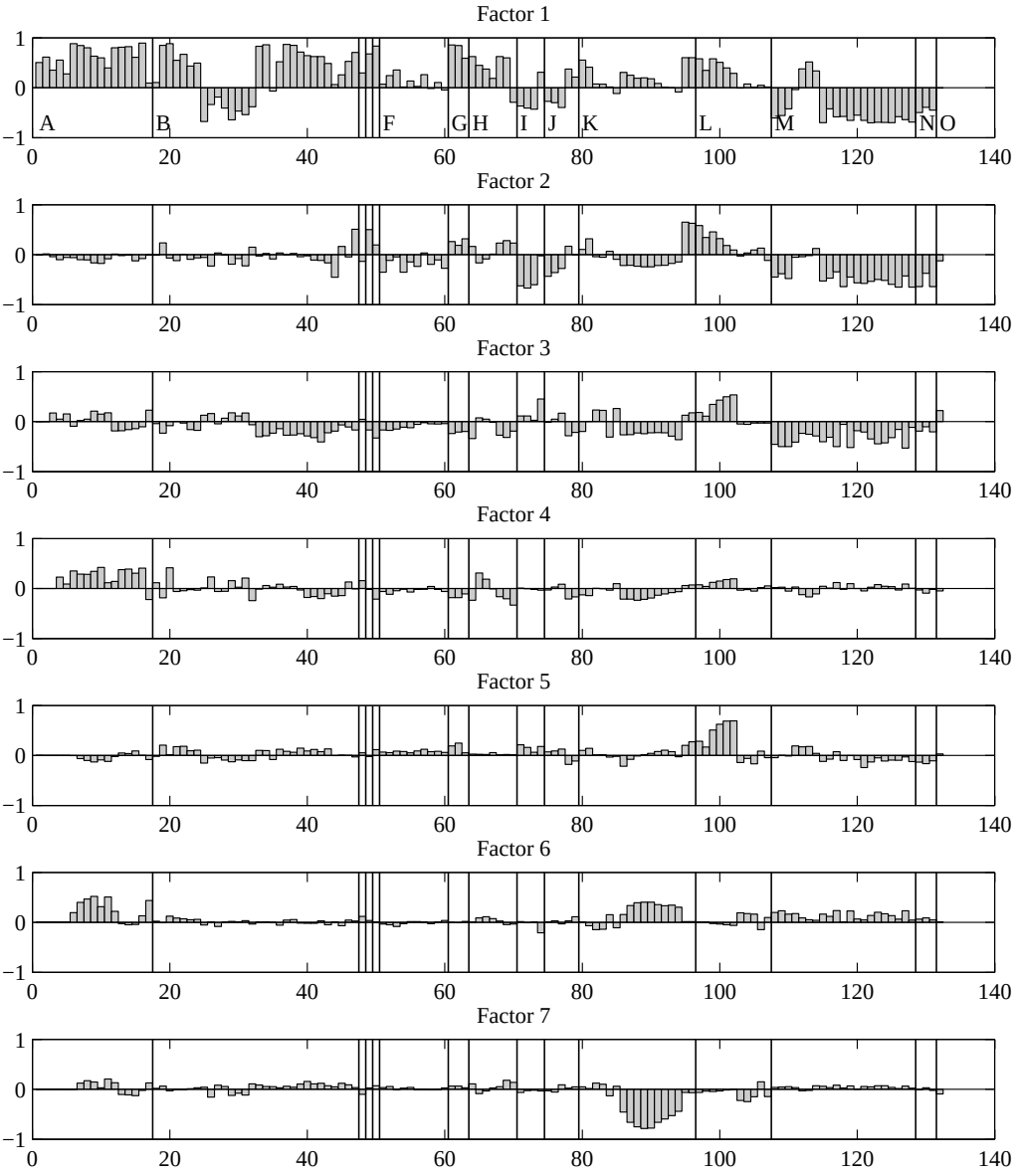


TABLE 3: MAXIMUM LIKELIHOOD ESTIMATE OF VAR COEFFICIENT MATRIX

The table reports the maximum likelihood estimates of the VAR coefficient matrix Φ_1 of model (37) estimated from the data set of Stock and Watson (2005). The eigenvalues of the estimated Φ_1 are reported in descending order. For complex eigenvalues we present both the real and imaginary (img) components.

	VAR coefficients							Eigenvalues	
Factor	1	2	3	4	5	6	7	real	img
1	0.39	0.04	-0.01	-0.19	0.08	-0.23	-0.03	0.99	0
2	0.23	1.07	-0.48	-0.57	-0.12	0.31	-0.83	0.93	-0.09
3	-0.41	-0.26	0.72	0.38	0.02	-0.60	0.32	0.93	0.09
4	-0.58	0.16	0.16	0.42	-0.05	0.03	-0.2	0.64	0
5	-0.14	-0.23	0.24	0.76	1.03	-0.67	0.75	0.39	0
6	-0.30	-0.16	-0.13	-0.11	-0.11	0.21	-0.05	0.11	0
7	0.43	0.10	-0.49	-0.61	-0.17	0.36	0.03	-0.13	0

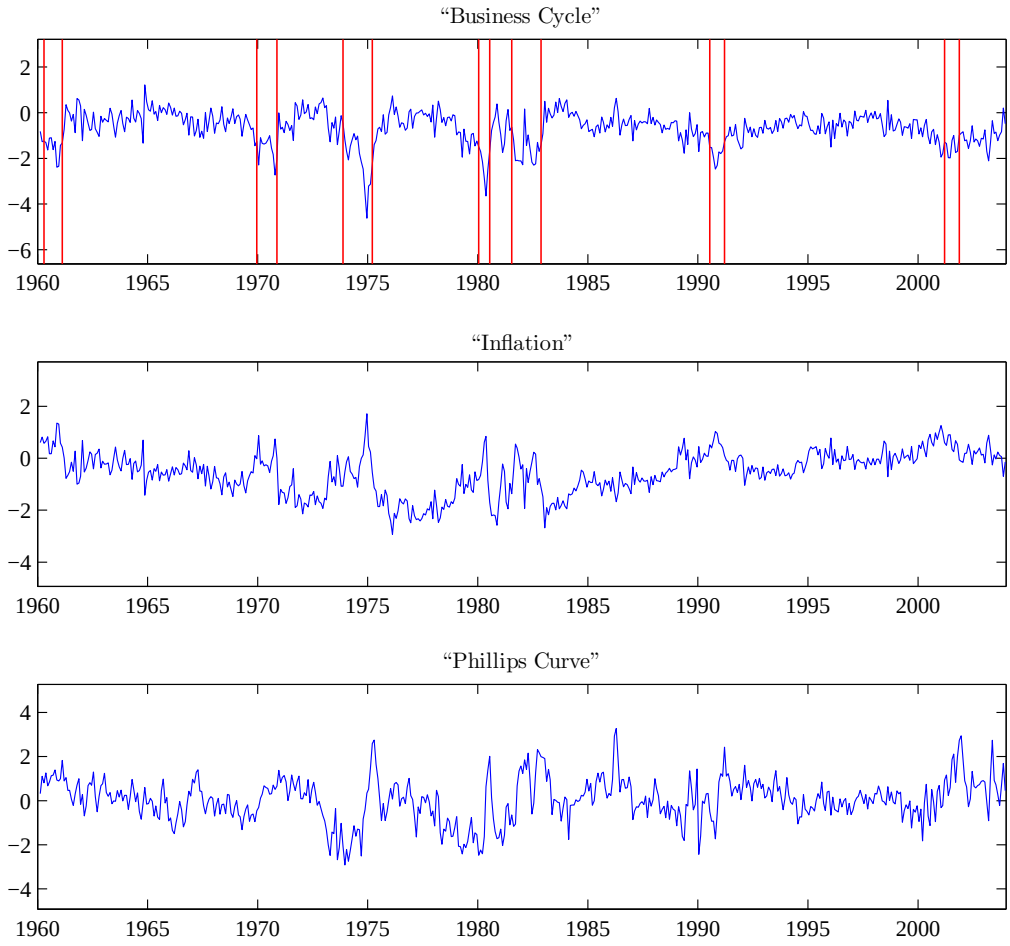
with the eigenvalues of Φ_1. From the eigenvalues we learn that all factors are estimated as stationary processes although some of the factor will be highly persistent because the largest eigenvalue is 0.99. Furthermore, we find the presence of persistent cyclical behaviour in the factors because one conjugate pair of complex eigenvalues is obtained with its real part equal to 0.93. The remaining four eigenvalues are relatively small. Future work may investigate whether changes in the cyclical properties of the factors have occurred before and after the beginning of the 1980s. Since the factors of the VAR process have zero mean and a variance matrix equal to the identity matrix, we can relate the individual coefficients in Φ_1 to each other. However, as in any VAR analysis, it remains hard to comment on individual coefficients in Φ_1.

5.2 Signal extraction and diagnostic checking

The first three estimated factors are displayed in Figure 3. The "business cycle" factor is displayed in the first panel together with the NBER business cycle reference dates of peaks and troughs. The NBER dates do not coincide perfectly with the peaks and troughs of the first factor but close enough to justify referring to it as the "business cycle" factor. The "inflation" factor (opposite sign) shows the steady inflation increases in the 1960s, the volatile inflation shocks in both the mid-1970s and early 1980s and the steady decline of inflation after 1983. We have labelled the third factor the "Phillips-curve" factor given its negative correlations with inflation and employment and its (weak) positive correlations with the unemployment series. The three factors displayed in Figure 3 are relatively persistent and cyclical. These characteristics are due to the three large eigenvalues of Φ_1 (0.99 and

FIGURE 3: ESTIMATED COMMON FACTORS

This figure shows the first three common factors extracted from the observed series by applying the method of Section 3.2.2 for model (37) from the data set of Stock and Watson (2005).



0.93) and the pair of complex eigenvalues reported in Table 3. The remaining four estimated factors are noisy series.

An appealing feature of our model-based analysis is that model misspecification tests and diagnostics concerning normality, heteroskedasticity and serial correlation can be considered as an effective tool for model selection. In time series, diagnostic test statistics are applied to standardised one-step ahead prediction errors. If the model is correctly specified these errors are IID. We will not argue that the dynamic factor model is the appropriate model specification for a joint analysis of 132 time series. However, the model misspecification diagnostics will indicate how far we are from a reasonable specification.

The KFS modifications of Section 3 allow us to compute the prediction errors for all 132

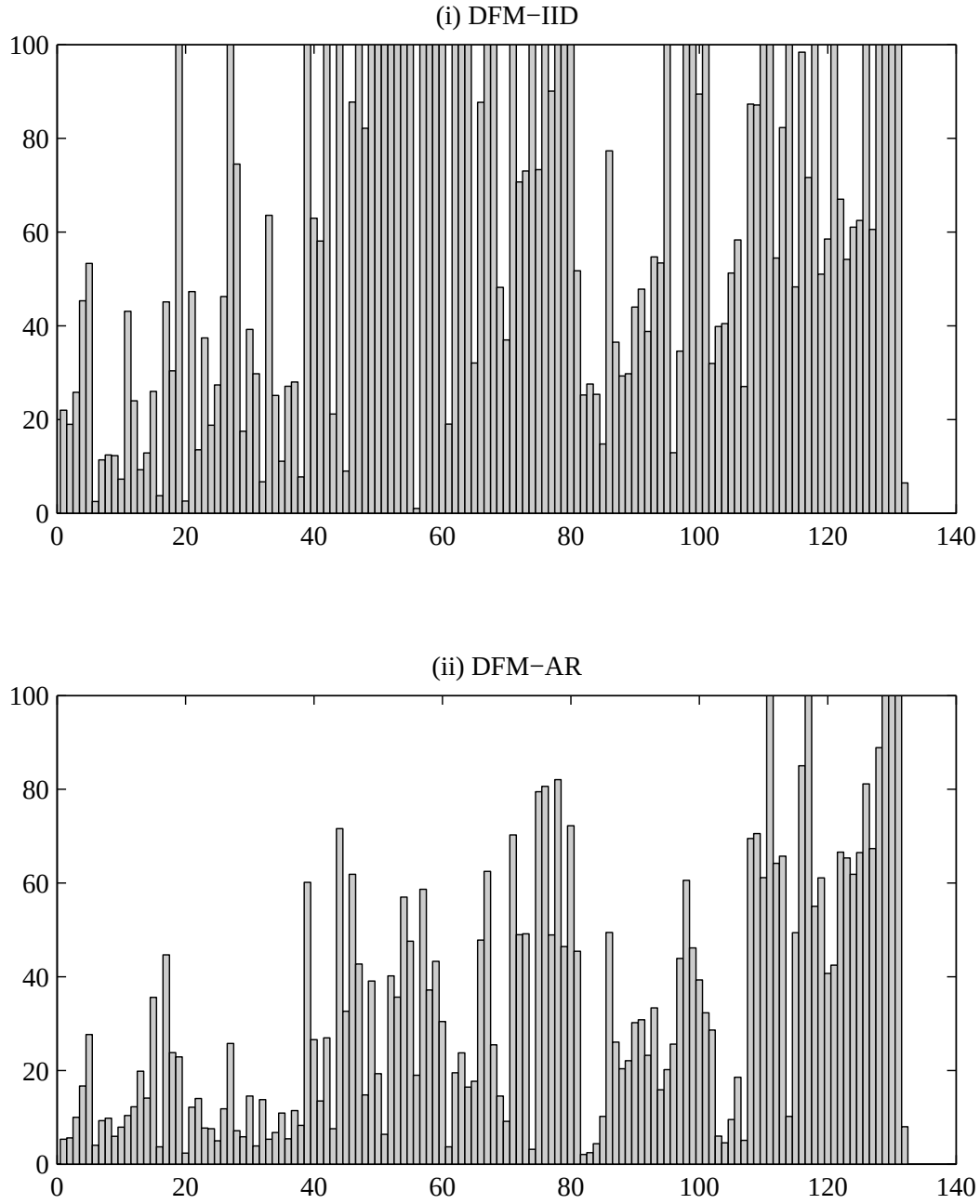
series in a few seconds. More specifically, we have computed the generalised least squares residuals as advocated by Harvey (1989, section 5.4) to allow for the intercept vector $\bar{\mu}$ in model (37). The residuals are standardized. To illustrate the effectiveness of residual diagnostics in the context of dynamic factor analysis, we compute for each residual series the serial correlation portmanteau χ^2 test of Ljung and Box (1978). The Box-Ljung $Q(q)$ statistic is based on the first q sample autocorrelations r_k^*, $k = 1, \ldots, q$, of the residual series and is computed by $Q(q) = \sum_{k=1}^{q} r_k^{*2}$. The Box-Ljung statistics for the 132 time series are graphically presented as index plots in Figure 4 for $q = 5$. The upper and lower index plots are for the residuals from the DFM-IID and DFM-AR models, respectively. The displayed Box-Ljung values are truncated at 100. It is evident that for many series the null hypothesis of no serial correlation in the residuals is rejected. The current dynamic factor models are therefore not fully satisfactory for this panel of macroeconomic time series. We can conclude that the DFM-AR specification is more successful in capturing the collective dynamics in the data set than the DFM-IID model.

6 Conclusions

We have presented new results which are instrumental for an effective likelihood-based analysis of dynamic factor models. We have shown that a high-dimensional dynamic factor model can be reduced to a low-dimensional state space model. This insight leads to substantial computational savings when estimating the factors, evaluating the loglikelihood function or sampling the factors in an MCMC algorithm. This state space formulation also allows us to calculate misspecification diagnostics from the one-step ahead forecasting errors. An important motivation for this paper is macroeconomic forecasting. Stock and Watson (2002b) advocate a two-step approach to the forecasting of macroeconomic time series: (i) extract a sufficient number of principal components from the panel; (ii) include these factors as (lagged) explanatory variables in a forecast model for a sub-set of the panel. With the results of this paper, likelihood-based methods become a viable alternative to the principal component approach. Future work must establish whether these methods produce more accurate forecasts. This paper is based on a fairly general modeling framework and we expect that the new results can be exploited in other applications and for different purposes.

FIGURE 4: BOX-LJUNG STATISTICS

This figure presents Box-Ljung statistics $Q(5)$ for the generalised least squares residuals of two dynamic factor models: (i) DFM-IID and (ii) DFM-AR.



A Appendices

A.1 Kalman filter and backward smoothing recursions

Consider the state space model (5) – (6) with initial state vector $\alpha_1 \sim N(a, P)$. The Kalman filter for a given time series y_t and a given parameter vector ψ is given by

$$\begin{aligned} v_t &= y_t - d_t - \mu - X_t\beta - Za_{t|t-1}, & D_t &= ZP_{t|t-1}Z' + \Sigma_\varepsilon, \\ & & K_t &= HP_{t|t-1}Z'D_t^{-1}, \\ a_{t+1|t} &= Ha_{t|t-1} + K_t v_t, & P_{t+1|t} &= HP_{t|t-1}H' - K_tF_tK_t' + RQR', \end{aligned} \tag{38}$$

for $t = 1, \ldots, T$, with initialisations $a_{1|0} = a$ and $P_{1|0} = P$, where v_t is the one-step ahead prediction error vector and D_t is its mean squared error, the one-step ahead predictor of the state vector α_t based on $y_1, \ldots, y_{t-1}$ is $a_{t|t-1}$, its mean squared error matrix is P_t and the *Kalman gain* matrix is K_t. Vector $a_{t+1|t}$ and matrix $P_{t+1|t}$ are evaluated recursively within the Kalman filter. A proof and more details are provided, amongst others, by Anderson and Moore (1979) and Durbin and Koopman (2001). For a linear state space model, the state predictor $a_{t|t-1}$ is the minimum mean squared error estimator (MMSE) of α_t based on $y_1, \ldots, y_{t-1}$; see Duncan and Horn (1972).

The smoothed estimators of the state vector can be obtained by the backward recursion

$$r_{t-1} = Z'D_t^{-1}v_t + L_t'r_t, \qquad N_{t-1} = Z'D_t^{-1}Z + L_t'N_tL_t, \qquad t = T, T-1, \ldots, 1,$$

with definition $L_t = H - K_tZ$ and initializations $r_T = 0$ and $N_T = 0$. From these recursions, the MMSE of the state vector using $y_1, \ldots, y_T$ is computed by

$$a_{t|T} = a_{t|t-1} + P_{t|t-1}r_{t-1}, \qquad P_{t|T} = P_{t|t-1} - P_{t|t-1}N_{t-1}P_{t|t-1}, \qquad t = T, T-1, \ldots, 1,$$

where $a_{t|T}$ is the MMSE of α_t and $P_{t|T}$ is its minimum mean squared error. Expressions for predictors of the state vector α_t and its mean squared error $P_{t|s}$ based on $y_1, \ldots, y_s$ for $s = t, t+1, \ldots, T-1$ can be found in Durbin and Koopman (2001). An expression for the covariance between α_t and α_{t-1} given $y_1, \ldots, y_T$ and denoted by $P_{t,t-1|T}$ is presented by de Jong and MacKinnon (1988) and given by

$$P_{t,t-1|n} = (P_tN_{t-1} - I)L_{t-1}'P_{t-1}, \qquad t = 2, \ldots, T,$$

which can be evaluated using the earlier recursions.

In Section 3.2 we treat μ and β as diffuse variables and extend the state vector to include μ and β. If the state vector contains diffuse variables, the corresponding diagonal values of P depends on the diffuse scalar $\kappa \to \infty$, see Section 3.2. Harvey and Phillips (1979) propose to handle diffuse variables by setting $\kappa = \kappa^*$ for some large value κ^*. This approach suffers however from numerical instability. Exact KFS methods for state space models with diffuse states are developed in Ansley and Kohn (1985), de Jong (1991) and Koopman (1997).

A.2 Proof of Lemma 1

From Conditions (ii) and (iii), we obtain

$$\text{Col}\{Z\} = \text{Col}\{\Sigma_\varepsilon A^{L\prime}\},$$

since matrix $\Sigma_\varepsilon A^{L\prime}$ has full column rank. It also implies that the columns of $\Sigma_\varepsilon A^{L\prime}$ form a basis for $\text{Col}\{Z\}$. Suppose $\Lambda^\dagger$ is a full column rank matrix such that $\text{Row}\{\Lambda^\dagger\} = \text{Col}\{Z\}$. Then there is a non-singular matrix C such that $\Sigma_\varepsilon A^{L\prime} = \Lambda^{\dagger\prime} C$. This proves the necessity part of Lemma 1.

A.3 Proof of equation (18)

We have

$$\begin{aligned} y_t^{H\prime}\Sigma_H^{-1} y_t^H &= (y_t - d_t)' A^{H\prime}(A^H \Sigma_\varepsilon A^{H\prime})^{-1} A^H (y_t - d_t) \\ &= (y_t - d_t)' J^H \Sigma_\varepsilon^{-1} (y_t - d_t), \end{aligned}$$

where $J^H \stackrel{def.}{=} A^{H\prime}(A^H \Sigma_\varepsilon A^{H\prime})^{-1} A^H \Sigma_\varepsilon$ is the projection matrix for a GLS with covariate matrix $A^{H\prime}$ and variance matrix Σ_ε^{-1}. Similarly, define

$$J^L \stackrel{def.}{=} A^{L\prime}(A^L \Sigma_\varepsilon A^{L\prime})^{-1} A^L \Sigma_\varepsilon,$$

as the GLS projection matrix for covariate matrix $A^{L\prime}$ and variance matrix Σ_ε^{-1}. Since the transformation matrix $A = (A^{L\prime}, A^{H\prime})'$ is full rank and $A^L \Sigma_\varepsilon A^{H\prime} = 0$, we must have

$$J^H = I - J^L.$$

The definition of A^L implies that $J^H = I - \Sigma_\varepsilon^{-1}\Lambda^\dagger(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}\Lambda^{\dagger\prime}$ and

$$J^{H\prime} = \Sigma_\varepsilon A^{H\prime}(A^H \Sigma_\varepsilon A^{H\prime})^{-1} A^H = I - \Lambda^\dagger(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1} \stackrel{def.}{=} M_\Lambda. \tag{39}$$

The proof of (18) is completed by the identity $J^H\Sigma_\varepsilon^{-1} = J^H\Sigma_\varepsilon^{-1}J^{H\prime}$ and the definition $e_t \stackrel{def.}{=} M_\Lambda(y_t - d_t)$ as the GLS residual for data vector $y_t - d_t$, covariate $\Lambda^\dagger$ and variance matrix Σ_ε.

A.4 Proof of two-step procedure in Section 3.2.2

For the proof of the two-step procedure we need the following lemma.

Lemma 3. *Given the linear regression model*

$$y^x = X^x\delta + \xi, \qquad \xi \sim N(0, \Omega), \tag{40}$$

where y^x is an $n \times 1$ vector, X^x is an $n \times k$ matrix of full column rank and Ω is an $n \times n$ positive definite matrix. Suppose that $\delta \sim N(0, \kappa I_k)$ for some scalar $\kappa > 0$ and denote the posterior distribution conditional on y^x by $p_\kappa(\delta|y^x)$. We then have

$$p_d(\delta|y^x) \stackrel{def.}{=} \lim_{\kappa\to\infty} p_\kappa(\delta|y^x) = g(\gamma, \Gamma),$$

where $g(\gamma, \Gamma)$ is the multivariate Gaussian density with mean γ and variance Γ given by

$$\gamma = (X^{x\prime}\Omega^{-1}X^x)^{-1}X^{x\prime}\Omega^{-1}y^x, \qquad \Gamma = (X^{x\prime}\Omega^{-1}X^x)^{-1}.$$

Proof. See *e.g.* Ansley and Kohn (1985) and de Jong (1991). □

Denote $W = (I \quad M_\Lambda')'$. The MMSLE of $M_\Lambda\mu$ and β based on y^H are given by

$$\begin{pmatrix} \widehat{\beta}^H \\ \widehat{M_\Lambda\mu} \end{pmatrix} = W \begin{pmatrix} \sum_t X_t^{H\prime}\Sigma_H^{-1}X_t^H & \sum_t X_t^{H\prime}\Sigma_H^{-1}A^H \\ \sum_t A^{H\prime}\Sigma_H^{-1}X_t^H & \sum_t A^{H\prime}\Sigma_H^{-1}A^H \end{pmatrix}^+ \begin{pmatrix} \sum_t X_t^H\Sigma_H^{-1}y_t^H \\ \sum_t A^{H\prime}\Sigma_H^{-1}y_t^H \end{pmatrix},$$

where M^+ denotes the Moore-Penrose inverse for some matrix M, see Magnus and Neudecker (1988, Theorem 8, p. 273). The corresponding mean squared errors are given by

$$\mathrm{Var}\begin{pmatrix} \widehat{\beta}^H \\ \widehat{M_\Lambda\mu} \end{pmatrix} = W \begin{pmatrix} \sum_t X_t^{H\prime}\Sigma_H^{-1}X_t^H & \sum_t X_t^{H\prime}\Sigma_H^{-1}A^H \\ \sum_t A^{H\prime}\Sigma_H^{-1}X_t^H & \sum_t A^{H\prime}\Sigma_H^{-1}A^H \end{pmatrix}^+ W'.$$

Using the results from Appendix A.3 and a result on the Moore-Penrose inverse of partitioned

matrices, see Rao and Mitra (1971, p. 41), we obtain

$$\widehat{\beta}^H = \left(\sum_t \widetilde{X}_t' \Sigma_\varepsilon^{-1} \widetilde{X}_t\right)^{-1} \sum_t \widetilde{X}_t' \Sigma_\varepsilon^{-1} \widetilde{y}_t, \qquad \mathrm{Var}(\widehat{\beta}^H) = \left(\sum_t \widetilde{X}_t' \Sigma_\varepsilon^{-1} \widetilde{X}_t\right)^{-1}, \tag{41}$$

where $\widetilde{y}_t$ and $\widetilde{X}_t$ are defined in (25).

The following arguments justify the particular application of KFS in the second step. Lemma 3 implies that the MMSLE and the corresponding mean squared errors for β and μ^L are equal to the posterior mean and variance if these coefficients are assumed diffuse. Let $p_d(\beta, \mu^L, \alpha|y)$ be the posterior density for diffuse β and μ^L, then

$$\begin{aligned} p_d(\alpha, \beta, \mu^L|y) &\stackrel{def.}{=} \lim_{\kappa_L, \kappa_H \to \infty} p_{\kappa_L, \kappa_H}(\alpha, \beta, \mu^L|y) = \lim_{\kappa_L, \kappa_H \to \infty} \frac{p(y^L|\alpha, \beta, \mu^L) p(\alpha) p_{\kappa_L}(\mu^L) p_{\kappa_H}(\beta|y^H)}{p_{\kappa_L, \kappa_H}(y^L|y^H)} \\ &= p(y^L|\alpha, \beta, \mu^L) p(\alpha) \lim_{\kappa_L \to \infty} \frac{p_{\kappa_L}(\mu^L)}{p_{d,\kappa_L}(y^L|y^H)} \lim_{\kappa_H \to \infty} p_{\kappa_H}(\beta|y^H) \\ &= \lim_{\kappa_L \to \infty} \frac{p(y^L|\alpha, \beta, \mu^L) p(\alpha) p_{\kappa_L}(\mu^L) p_d(\beta|y^H)}{p_{d,\kappa_L}(y^L|y^H)} \end{aligned} \tag{42}$$

where κ_L and κ_H refer to diffuse variables in the model equations for y_t^L and y_t^H, respectively, and where

$$p_{d,\kappa_L}(y^L|y^H) = \int p(y^L|\beta, \mu^L) p_{\kappa_L}(\mu^L) p_d(\beta|y^H) d\mu^L d\beta.$$

From Lemma 3, $p_d(\beta|y^H)$ is a density with mean $\widehat{\beta}^H$ and variance $\mathrm{Var}(\widehat{\beta}^H)$. Equation (42) represents the smoothing density of the linear state space model (27) and (6). This completes the proof of the two-step procedure.

The MMSLE of $A^H\mu$ is given by

$$\widehat{A^H\mu} = \bar{y}^H - \bar{X}^H\widehat{\beta} = A^H(\bar{y} - \bar{X}\widehat{\beta}).$$

Multiplying both sides of this equation with $\Sigma_\varepsilon A^{H\prime}(A^H\Sigma_\varepsilon A^{H\prime})^{-1}$ and using equation (39), we obtain the MMSLE of $\mu^H = M_\Lambda\mu$ as given by

$$\widehat{M_\Lambda\mu} = M_\Lambda(\bar{y} - \bar{X}\widehat{\beta}).$$

The expressions in (28) and (29) follow straightforwardly.

A.5 Proof of diffuse likelihood in Section 3.2.3

The following two lemmas are required for the main proof.

Lemma 4. *Consider regression model* (40) *in Lemma 3 with* $\delta \sim N(0, \kappa I)$ *and* $\kappa \to \infty$. *The diffuse loglikelihood* $\ell_d(y)$ *defined in* (20) *is given by*

$$\ell_d(y) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Omega| - \frac{1}{2}|X^{x\prime}\Omega^{-1}X^x| - \frac{1}{2}e^{x\prime}\Omega^{-1}e^x,$$

where e^x *is the residual vector from a GLS regression on* y^x *with covariate matrix* X^x *and variance matrix* Ω.

Proof. See *e.g.* Shephard (1993). □

Lemma 5. *Consider regression model* (40) *in Lemma 3 and its counterpart*

$$y^x = \widetilde{X}^x\delta + \xi, \qquad \xi \sim N(0, \Omega), \tag{43}$$

where $\widetilde{X}^x$ *is a full column rank matrix and* $Col\{X^x\} = Col\{\widetilde{X}^x\}$. *Denote the diffuse likelihood of model* (40) *by* $p_d(y^x)$ *and* $\widetilde{p}_d(y^x)$ *as the diffuse likelihood of model* (43) *for* $\delta \sim N(0, \kappa I)$ *and* $\kappa \to \infty$. *Then* $p_d(y^x)$ *and* $\widetilde{p}_d(y^x)$ *are proportional. Without loss of generality, suppose* $\widetilde{X}^x = X^x M$ *for some invertible matrix* M *then,*

$$p_d(y^x) = |M|\widetilde{p}_d(y^x).$$

Proof. From Lemma 4 we get

$$\log p_d(y^x) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Omega| - \frac{1}{2}\log\left|X^{x\prime}\Omega^{-1}X^x\right| - \frac{1}{2}e^{x\prime}\Omega^{-1}e^x,$$
$$\log \widetilde{p}_d(y^x) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Omega| - \frac{1}{2}\log\left|M'X^{x\prime}\Omega^{-1}X^xM\right| - \frac{1}{2}\widetilde{e}^{x\prime}\Omega^{-1}\widetilde{e}^x,$$

where e^x is defined in Lemma 4 and $\widetilde{e}^x$ is the residual vector from a GLS regression on y^x with covariate matrix $\widetilde{X}^x$ and variance matrix Ω. Since $\mathrm{Col}\{X^x\} = \mathrm{Col}\{\widetilde{X}^x\}$, we have $e^x = \widetilde{e}^x$. The result follows from $\left|\widetilde{X}^{x\prime}\Omega^{-1}\widetilde{X}^x\right| = |MX^{x\prime}\Omega^{-1}X^xM| = |M|^2\,|X^{x\prime}\Omega^{-1}X^x|$. □

Define

$$\mu^* = \begin{pmatrix} \mu^L \\ \mu^H \end{pmatrix} = \begin{pmatrix} A^L\mu \\ A^H\mu \end{pmatrix},$$

$y = (y_1', \ldots, y_T')'$ and $y^* = (y_1^{*\prime}, \ldots, y_T^{*\prime})'$, we have

$$\begin{aligned}\int p(y|\mu,\beta)p_\kappa(\mu,\beta)d\mu d\beta &= |A|^T \iint p(y^*|\mu,\beta)p_\kappa(\mu,\beta)d\mu d\beta \\ &= |A|^{T-1} \iint p(y^*|\mu^*,\beta)p_\kappa(\mu^*,\beta)d\mu^* d\beta \qquad (44) \\ &= |A|^{T-1} \iint p(y^H|\mu^H,\beta)p(y^L|\mu^L,\beta)p_\kappa(\mu^*,\beta)d\mu^* d\beta\end{aligned}$$

where equation (44) follows directly from Lemma 5. An application of the monotone convergence theorem and some elementary calculations give

$$\begin{aligned}\lim_{\kappa\to\infty} \kappa^{-(N+K)/2} \iiint p(y^H|\mu^H,\beta)p(y^L|\mu^L,\beta)p_\kappa(\mu^*,\beta)d\mu^L d\mu^H d\beta \\ = p_d(y^H) \lim_{\kappa\to\infty} \kappa^{-m/2} \iiint p(y^L|\mu^L,\beta)p_\kappa(\mu^L)p_d(\mu^H,\beta|y^H)d\mu^L d\mu^H d\beta,\end{aligned} \qquad (45)$$

where $p(y^H)$ is the diffuse likelihood of the second equation,

$$\begin{aligned}p_d(y^H) &= \lim_{\kappa\to\infty} \kappa^{-(N+K-m)/2} \int p(y^H|\mu^H,\beta)p(\mu^H,\beta)d\mu^H d\beta, \\ p_d(\mu^H,\beta|y^H) &= \lim_{\kappa\to\infty} p_\kappa(\mu^H,\beta|y^H).\end{aligned}$$

Similarly to the argument leading to (42) we have

$$\begin{aligned}\lim_{\kappa\to\infty} \kappa^{-m/2} \iiint p(y^L|\mu^L,\beta)p_\kappa(\mu^L)p_d(\mu^H,\beta|y^H)d\mu^L d\mu^H d\beta \\ = \lim_{\kappa\to\infty} \kappa^{-m} \iint p(y^L|\mu^L,\beta)p_\kappa(\mu^L)p_d(\beta|y^H)d\mu^L d\beta \\ = \exp\{\ell_d(y^L)\},\end{aligned}$$

and, using (12), we have

$$\begin{aligned}\ell_d(y) &= (T-1)\log|A| + \log p_d(y^H) + \ell(y^L) \\ &= \frac{T-1}{2}(\log|\Sigma_L| - \log|\Sigma_\varepsilon|) + \log p_d(y^H) + \ell(y^L).\end{aligned}$$

The density $p_d(y^H)$ is the diffuse likelihood of the linear model

$$y_t^H = \mu^H + X_t^H\beta + e_t^H, \qquad e_t^H \sim N(0,\Sigma_H), \qquad t = 1,\ldots,T. \qquad (46)$$

Define $X_t^a = (X_t^H \quad I)$ and $X^a = (X_1^a, \ldots, X_T^a)$ such that (46) can be written as the multiple regression model

$$y^H = X^a \begin{pmatrix} \beta \\ \mu^H \end{pmatrix} + e^H, \qquad e^H \sim N(0, I \otimes \Sigma_H),$$

where $e^H = (e_1^{H\prime}, \ldots, e_T^{H\prime})'$. From Lemma 4 we obtain

$$\log p_d(y^H) = -\frac{T(N-m)}{2}\log 2\pi - \frac{T}{2}\log|\Sigma_H| - \frac{1}{2}\log\left|\sum_t X_t^{a\prime}\Sigma_H^{-1}X_t^a\right| - \frac{1}{2}\sum_t \bar{e}_t'\Sigma_H^{-1}\bar{e}_t,$$

where $\bar{e}_t$ are residuals from GLS on y^H with covariates X^a and variance matrix $I \otimes \Sigma_H$. By using an argument similar to the one used in Appendix A.3, we have

$$\sum_t \bar{e}_t'\Sigma_H^{-1}\bar{e}_t = \sum_t \widetilde{y}_t'\Sigma_\varepsilon^{-1}\widetilde{y}_t.$$

Finally, it can be shown that

$$\left|\sum_t X_t^{a\prime}\Sigma_H^{-1}X_t^a\right| = \frac{\left|\sum_t \widetilde{X}_t'\Sigma_\varepsilon^{-1}\widetilde{X}_t\right|}{|\Sigma_H|}T^{N-m},$$

and from Condition (iv) in Section 3.1.1, that is $|\Sigma_H| = 1$, we have

$$\log p_d(y^H) = -\frac{T(N-m)}{2}\log 2\pi - \frac{N-m}{2}\log T - \frac{T}{2}\log\left|\sum_t \widetilde{X}_t'\Sigma_\varepsilon^{-1}\widetilde{X}_t\right| - \frac{1}{2}\sum_t \widetilde{y}_t'\Sigma_\varepsilon^{-1}\widetilde{y}_t.$$

The result follows.

A.6 The score function of Section 4.1.1

The derivatives of (33) with respect to the system matrices Z, H, Σ_ε and Σ_η of the linear Gaussian state space model (5) – (6), with $\mu = 0$, $\beta = 0$ and $R = I$, are given by

$$\frac{\partial \ell(y)}{\partial Z} = \Sigma_\varepsilon^{-1}\Big(\sum_{t=1}^{T}\{y_t - d_t\}a_{t|T}' - ZS_{1:T}^{(0)}\Big), \qquad \frac{\partial \ell(y)}{\partial \Sigma_\varepsilon} = Q_\varepsilon^*\Sigma_\varepsilon^{-1} - \frac{1}{2}\mathrm{diag}(Q_\varepsilon^*\Sigma_\varepsilon^{-1}),$$

$$\frac{\partial \ell(y)}{\partial H} = \Sigma_\eta^{-1}(S_{2:T}^{(1)} - HS_{2:T}^{(0)}), \qquad \frac{\partial \ell(y)}{\partial \Sigma_\eta} = Q_\eta^*\Sigma_\eta^{-1} - \frac{1}{2}\mathrm{diag}(Q_\eta^*\Sigma_\eta^{-1}).$$

where $Q_\varepsilon^* = Q_\varepsilon - T$, $Q_\eta^* = Q_\eta - T - 1$, with Q_ε and Q_η defined in (34),

$$S_{j:k}^{(0)} = \sum_{t=j}^{k} a_{t|T} a_{t|T}' + P_{t|T}, \qquad S_{j:k}^{(1)} = \sum_{t=j}^{k} a_{t|T} a_{t-1|T}' + P_{t,t-1|T}, \tag{47}$$

for $j, k = 1, \ldots, T$ $(j \leq k)$, where $a_{t|T}$, $P_{t|T}$ and $P_{t-1,t|T} = P_{t,t-1|T}'$ are evaluated by the KFS methods of Appendix A.1. Matrices Q_ε and Q_η depend on the smoothed disturbances $\widehat{\varepsilon}_t = y_t - d_t - Z a_{t|T}$ and $\widehat{\eta}_t = a_{t|T} - F a_{t-1|T}$ together with their variances which depend on $P_{t|T}$, $P_{t,t-1|T}$ and the system matrices. The derivatives given above are evaluated at $\psi = \psi^*$. The system matrices are functions of coefficient vector ψ. For cases where $\mu \neq 0$, $\beta \neq 0$ and/or $R \neq I$, similar expressions can be obtained for the derivatives but the expressions are more lengthy and more intricate, see Koopman and Shephard (1992) for a detailed discussion.

A.7 EM algorithm of Section 4.1.2

The details of the EM algorithm are specific to the model specification. We illustrate the EM for model (2), (3) and (4) with $q_\Theta = 0$, $q_\Psi = 1$ and diagonal matrix Ψ_1. The details are given for the likelihood function conditional on observation y_1. The treatment of initial conditions is intricate in its detail and does not provide further insight. The model considered is also discussed by Watson and Engle (1983, section 5). Given the definitions in the previous subsection, the M step provides new estimates of the system matrices and are given by

$$\begin{aligned}
Z_i^+ &= -\sum_{t=2}^{T} (y_{it} - \Psi_{ii} y_{i,t-1})(a_{t|T} - \Psi_{ii} a_{t-1|T})' [\Psi_{ii}^2 S_{1:T-1}^{(0)} - \Psi_{ii} S_{2:T}^{(1)} - \Psi_{ii} S_{2:T}^{(1)\prime} + S_{2:T}^{(0)}]^{-1}, \\
F^+ &= S_{2:T}^{(1)} S_{2:T}^{(0)\,-1}, \\
\Psi_{ii}^+ &= \sum_{t=2}^{T} Z_i P_{t,t-1|T} Z_i' - \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{i,t-1} \;/\; (Z_i S_{1:T-1}^{(0)} Z_i' + \sum_{t=2}^{T} y_{it} \{\widehat{\varepsilon}_{it} - Z_i a_{t|T}\}),
\end{aligned} \tag{48}$$

where Z_i is the ith row of Z and Ψ_{ii} is the ith diagonal element of Ψ_1 for $i = 1, \ldots, N$. Expressions for Σ_ε^+ and Σ_η^+ are obtained as solutions of $Q_\varepsilon = 0$ and $Q_\eta = 0$, respectively. The system matrices are evaluated at $\psi = \psi^*$. The new coefficients for λ_i are distilled from Z_i^+ for $i = 1, \ldots, N$ while new coefficients for Φ_i $(i = 1, \ldots, q_\Phi)$ are distilled from F^+.

The equations for Z_i^+ and Ψ_{ii}^+ in (48) can not be solved separately. Keeping $S_{jk}^{(0)}$, $S_{jk}^{(1)}$, $a_{t|T}$, $P_{t,t-1|T}$ and $\widehat{\varepsilon}_{it}$ $(i = 1, \ldots, N,\ t = 2, \ldots, T)$ fixed, we obtain a solution by repeatingly solving one equation separately and substituting its solution in the other equation. This same scheme is also used in Watson and Engle (1983). Meng and Rubin (1993) show that this algorithm retains the attractive properties of the EM algorithm. In particular, the

likelihood is monotonically increasing over the iterations and the algorithm converges to an optimum.

References

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *J. Business and Economic Statist. 18*, 338–357.

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

Ansley, C. F. and R. Kohn (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics 13*, 1286–1316.

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

Bai, J. and S. Ng (2002). Determining the number of of factors in approximating factor models. *Econometrica 70*, 191–221.

Bernanke, B., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly J. of Economics 120*, 387–422.

Boivin, J. and M. Giannoni (2006). Has monetary policy become more effective ? *Rev. Economics and Statistics 88*, 445–62.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–53.

Chamberlain, G. and M. Rothschild (1983). Arbitrage factor structure, and mean-variance analysis of large asset markets. *Econometrica 51*, 1281–1304.

Chib, S. (1993). Bayes regression with autoregression errors: A Gibbs sampling approach. *J. Econometrics 58*, 275–294.

Chib, S. (2001). Markov chain Monte Carlo methods: Computation and inference. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–649. Amsterdam: North-Holland.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory : A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

Connor, G. and R. A. Korajczyk (1988). Risk and return in an equilibrium APT : Application of a new test methodology. *Journal of Financial Economics 21*(2), 255–289.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *Journal of Finance 48*(4), 1263–91.

de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics 19*, 1073–83.

de Jong, P. and M. J. MacKinnon (1988). Covariances for smoothed estimates in state space models. *Biometrika 75*, 601–2.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–50.

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 regression 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.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–16.

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.

Fiorentini, G., E. Sentana, and N. Shephard (2004). Likelihood-based estimation of latent generalised ARCH structures. *Econometrica 72*, 1481–1517.

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.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *J. Time Series Analysis 15*, 183–202.

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.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies 9*(2), 557–87.

Gilks, W. K., S. Richardson, and D. J. Spiegelhalter (Eds.) (1996). *Markov chain Monte Carlo in Practice*. London: Chapman & Hall.

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

Harvey, A. C. and G. D. A. Phillips (1979). The estimation of regression models with autoregressive-moving average disturbances. *Biometrika 66*, 49–58.

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for non-stationary time series models. *J. American Statistical Association 92*, 1630–8.

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.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 66*, 67–72.

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

Magnus, J. R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.

Meng, X. L. and D. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 80*, 267–78.

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

Quah, D. and T. J. Sargent (1993). A dynamic index model for large cross sections. In J. H. Stock and M. Watson (Eds.), *Business cycles, indicators and forecasting*, pp. 285–306. Chicago: University of Chicago Press.

Rao, C. R. and S. K. Mitra (1971). *Generalized Inverse of Matrices and Its Applications*. New York: John Wiley & Sons.

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.

Shephard, N. (1993). Maximum likelihood estimation of regression models with stochastic trend components. *J. American Statistical Association 88*, 590–95.

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.

Stock, J. H. and M. Watson (2005). Implications of dynamic factor models for var analysis. Discussion paper.

Stock, J. H. and M. Watson (2006). Forecasting with many predictors. In G. Elliot, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 515–54. Amsterdam: Elsevier Science Publishers.

Tunnicliffe-Wilson, G. (1989). On the use of marginal likelihood in time series model estimation. *J. Royal Statistical Society B 51*, 15–27.

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.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics 11*, 95–103.