# Likelihood-based analysis for dynamic factor models

*Borus Jungbacker*[(a)] *and* *Siem Jan Koopman*[(a,b)]

[(a)] Department of Econometrics, VU University Amsterdam,
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[(b)] Tinbergen Institute, Amsterdam

June 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 modeled 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 (quasi-)maximum likelihood. An illustration is provided for the analysis of a large panel of macroeconomic time series.

*JEL classification*: C33; C43

*Some keywords*: Common factors; High-dimensional vector series; Kalman Filter; Quasi-maximum likelihood.

We would like to thank seminar participants for the insightful discussions during the presentations at Nuffield College, Oxford, NBER-NSF Time Series Conference 2007 in Iowa, ESRC Econometric Study Group, Annual Conference 2007 in Bristol, University of Pennsylvania, Tinbergen Institute Amsterdam and European University Institute, Florence. We are indebted to Mark W. Watson for kindly making the data-set for our empirical study available on his website. All errors in the paper are our own.

# 1 Introduction

In this paper we consider the dynamic factor model given by

$$y_{it} = \mu_i + x_{it}\beta + \sum_{j=0}^{q_\Lambda} \lambda_{ij}' f_{t-j} + u_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where $y_{it}$ denotes the observed value for the $i$th time series at time $t$, $\mu_i$ is a fixed and unknown constant, $x_{it}$ is a $1 \times K$ vector of covariates, $\beta$ is a $K \times 1$ vector of regression coefficients, $f_t$ is an $r \times 1$ vector of common factors, $\lambda_{ij}$ is an $r \times 1$ vector of loadings associated with the common factors at lag $j$ and $u_{it}$ is the idiosyncratic component. The factors are modeled by linear dynamic processes and the idiosyncratic components by autoregressive processes with mutually correlated zero mean innovations. We particularly focus on the case where a high-dimensional panel of $N$ time series depends on a relatively small number of $r$ common dynamic factors.

When the idiosyncratic components $u_{it}$ and the common factors $f_t$ are assumed Gaussian we can evaluate the likelihood function efficiently by means of the Kalman filter. In case the innovations are non-Gaussian, the Gaussian likelihood can be regarded as a quasi-likelihood. The Gaussian likelihood function can be numerically maximized to obtain maximum likelihood or quasi-maximum likelihood (QML) parameter estimates. This is the approach taken by Engle and Watson (1981) for a Gaussian model with one common factor. Watson and Engle (1983) use the expectation-maximization (EM) algorithm of Dempster, Laird, and Rubin (1977) to find the optimum of the likelihood, see also Shumway and Stoffer (1982) and Quah and Sargent (1993). However, in many of the recent applications of the dynamic factor model, the high-dimensional panel of time series and the resulting large number of parameters make such an approach 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 both signal extraction and likelihood evaluation. Finding the optimum of a likelihood function is not straightforward if there is a large number of parameters. We present new devices for an effective implementation of the optimization methods. We also derive an algorithm to efficiently evaluate the marginal Gaussian loglikelihood for dynamic factor models of the form (1). In case of time series models with regression effects, estimators based on the marginal likelihood are found to perform better in small samples than the corresponding maximum likelihood estimators, see, for example, Tunnicliffe-Wilson (1989) and Shephard (1993).

To demonstrate the applicability of our results we estimate a number of dynamic factor models on a panel of 132 macroeconomic time series. The largest model that we estimate includes seven dynamic factors and more than 1000 parameters. The new results presented in this paper enable us to estimate these parameters in minutes.

The key insight for our results is that the observed time series can be split into a low-dimensional vector series and a high-dimensional vector series. For the estimation of the factors and the evaluation of the likelihood function, we need to apply the computationally intensive Kalman filter methods to the low-dimensional series while simple regression-style calculations suffice for the high-dimensional part. As a result, we are able to achieve large computational gains.

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. The increasing availability of high-dimensional vector series in economics and finance has motivated work on alternative methods to estimate the common factors. 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 recent 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.

The likelihood-based approach has a number of advantages over the principal components method. Since the factors are explicitly modeled and the estimation method takes account of the model specification, the factors can represent aspects of economic theory. Hypothesis tests can be formulated and tested. The techniques of this paper allow real-time estimation of the underlying factors, estimation of past factors as well as prediction of factors and future observations. The Kalman filter further produces mean squared errors of the factor estimates without an extra computational effort. Unbalanced data-sets are also easily handled in this framework. Finally, in case the data generating process can be represented as a Gaussian dynamic factor model, the parameter estimators are asymptoticaly efficient. Moreover, Doz, Giannone, and Reichlin (2006) show, under mild conditions, that the factor estimates from the QML procedure are consistent for the true factors when $T \to \infty$ and $N \to \infty$, even if the model is misspecified. They also present Monte Carlo evidence that the QML factor estimates are often more precise than the principal component estimates.

The remainder of the paper is organized as follows. The dynamic factor model and its state space form are presented in Section 2. Parameter estimation is discussed in Section 3. The key results of this paper are presented in Section 4 with proofs and derivations given in the Appendix. Section 5 discusses the new devices for parameter estimation using quasi-maximum likelihood methods. An empirical illustration is provided in Section 6 while Section 7 concludes.

# 2 Generalized dynamic factor model with covariates

We assume that the dynamic characteristics of a time series of observed $N \times 1$ vectors $y_1, \ldots, y_T$ can be described by the dynamic factor model (1). The vector form of model (1) is given by

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

where $y_t = (y_{1t}, \ldots, y_{Nt})'$, $u_t = (u_{1t}, \ldots, u_{Nt})'$, $\bar{\mu} = (\mu_1, \ldots, \mu_N)'$, $\bar{X}_t = (x'_{1t}, \ldots, x'_{Nt})'$ and matrix lag polynomial $\Lambda(L) = \Lambda_0 + \sum_{j=1}^{q_\Lambda} \Lambda_j L^j$ with $\Lambda_j = (\lambda_{1j}, \ldots, \lambda_{Nj})'$ for $j = 0, \ldots, q_\Lambda$, lag-operator $L$ and non-negative integer $q_\Lambda$. The vector $f_t$ of common factors is modeled by the vector autoregressive moving average (VARMA) process

$$\Phi(L)f_t = \Theta(L)\zeta_t, \tag{3}$$

where $\zeta_t$ is a vector of innovations and the matrix lag polynomials are $\Phi(L) = I - \sum_{j=1}^{q_\Phi} \Phi_j L^j$ and $\Theta(L) = I + \sum_{j=1}^{q_\Theta} \Theta_j L^j$ with $r \times r$ autoregressive coefficient matrices $\Phi_j$ for $j = 1, \ldots, q_\Phi$ and $r \times r$ moving average coefficient matrices $\Theta_j$ for $j = 1, \ldots, q_\Theta$. We model the idiosyncratic component vector $u_t$ in (2) as the vector autoregressive (VAR) process

$$\Psi(L)u_t = \varepsilon_t, \tag{4}$$

where $\varepsilon_t$ is a vector of innovations and the matrix lag polynomial is $\Psi(L) = I - \sum_{j=1}^{q_\Psi} \Psi_j L^j$ with $N \times N$ autoregressive coefficient matrix $\Psi_j$ for $j = 1, \ldots, q_\Psi$. Finally, we denote the set of all parameters in the model by $\psi$. The set of parameters in $\psi$ excluding $\bar{\mu}$ and $\beta$, are denoted by $\theta$, that is

$$\psi = (\bar{\mu}' \;,\; \beta' \;,\; \theta')'.$$

We adopt the following set of assumptions for model (2) – (4):

(i) The permissible parameter space $S_\psi$ is a compact sub-set of the Euclidean space. The true parameter $\psi_0$ is an interior point of $S_\psi$.

(ii) For all $\psi$ in $S_\psi$ and $|z| \leq 1$, we have $|\Phi(z)| \neq 0$ and $|\Psi(z)| \neq 0$.

(iii) Denote by $\mathcal{F}_t$ the $\sigma$-algebra generated by $y_1, \ldots, y_t$, with $\mathcal{F}_0$ the trivial $\sigma$-algebra, then

$$\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0, \qquad \mathbb{E}(\zeta_t|\mathcal{F}_{t-1}) = 0, \qquad \mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma_\varepsilon, \qquad \mathbb{E}(\zeta_t\zeta_t'|\mathcal{F}_{t-1}) = \Sigma_\zeta,$$

for $t = 1, \ldots, T$. We assume that $\Sigma_\varepsilon$ is a nonsingular matrix.

(iv) The vector sequences $\{\varepsilon_t\}$ and $\{\zeta_t\}$ are uncorrelated and have finite fourth moments.

(v) The covariate sequence $\{\bar{X}_t\}$ is independent of the innovation sequences $\{\varepsilon_t\}$ and $\{\zeta_t\}$.

(vi) Matrix $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \bar{X}_t \bar{X}_{t+j}'$ exists and is finite for every non-negative integer $j$.

(vii) Let $\Gamma_y(h;\psi) = \mathrm{Cov}_\psi(y_t, y_{t+h})$ for $\psi \in S_\psi$, then for any $\psi^* \in S_\psi$ such that $\psi^* \neq \psi_0$, $\Gamma_y(s;\psi) \neq \Gamma_y(s;\psi_0)$ for at least one value of $s \in \mathbb{Z}_+$.

(viii) Denote $\bar{\mu}_0$ and $\beta_0$ as the true values of $\bar{\mu}$ and $\beta$, respectively. The process $y_t - \bar{\mu}_0 - \bar{X}_t\beta_0$ can be written as a VAR process $\Pi(L;\psi_0)(y_t - \bar{\mu}_0 - \bar{X}_t\beta_0) = \tilde{u}_t$, where $\Pi(z;\psi_0) = I - \sum_{i=1}^{\infty} \Pi_i(\psi_0) z^i$, $\mathbb{E}(\tilde{u}_t|\mathcal{F}_{t-1}) = 0$ and the elements of $\Pi_1, \Pi_2 \ldots$ are absolutely summable.

Assumption (ii) implies that the dynamic factor model is stationary for all admissable parameter vectors. The assumption in (iii) of $\Sigma_\varepsilon$ nonsingular is not restrictive since any dynamic factor model with a singular matrix $\Sigma_\varepsilon$ can be rewritten to satisfy assumption (iii). Assumption (vii) is an identifiability assumption. In practice, for this assumption to hold, we need to put restrictions on $\Lambda_0$, $\Lambda_1, \ldots, \Lambda_{q_\Lambda}$, $\Phi_1, \ldots, \Phi_{q_\Phi}$, $\Theta_1, \ldots, \Theta_{q_\Theta}$ and $\Sigma_\zeta$. Parameter restrictions are common in the literature on factor models, see *e.g.* Geweke and Zhou (1996) for further discussions. Examples of the general model specification (2) – (4) are given in Illustrations 1 and 2 below.

The dynamic factor model (2) with idiosyncratic component (4) can be expressed in static form as follows

$$y_t = \mu + d_t + X_t\beta + \Lambda F_t + \varepsilon_t, \qquad t = q_\Psi + 1, \ldots, T, \tag{5}$$

where $\mu = \Psi(I)\bar{\mu}$, $d_t = \sum_{j=1}^{q_\Psi} \Psi_j y_{t-j}$, $X_t = \Psi(L)\bar{X}_t$, $F_t = (f_t', f_{t-1}', \ldots, f_{t-s}')'$ and $\Lambda = (\Lambda_0, \Lambda_1^*, \ldots, \Lambda_s^*)$ with

$$\Lambda_k^* = \Lambda_k^+ - \sum_{j=1}^{k} \Psi_j^+ \Lambda_{k-j}^+, \qquad \Lambda_i^+ = \begin{cases} \Lambda_i, & i \le q_\Lambda, \\ 0, & \text{otherwise}, \end{cases} \qquad \Psi_i^+ = \begin{cases} \Psi_i, & i \le q_\Psi, \\ 0, & \text{otherwise}, \end{cases}$$

and $s = q_\Lambda + q_\Psi$ for $i, k = 1, \ldots, s$. The number of static factors in $F_t$ is given by $m = r\,(s+1)$.

The VARMA process (3) for $f_t$ can be expressed as $f_t = G\alpha_t$ for a suitable matrix $G$ and

$$\alpha_t = H\alpha_{t-1} + R\zeta_t, \tag{6}$$

where $p \times 1$ state vector $\alpha_t$ has mean $\mathbb{E}(\alpha_t) = 0$ and variance matrix $\mathrm{Var}(\alpha_t) = Q$. The system matrices $H$ and $R$ are sparse matrices and contain the coefficient matrices in the polynomials $\Phi(L)$ and $\Theta(L)$. Matrices $H$ and $R$ can be constructed such that $G$ consists of rows of the unity matrix and has full row rank. For more details on the state space specification of the VARMA model, see *e.g.* Harvey (1989) and Durbin and Koopman (2001).

The dimension of the state $\alpha_t$ is generally higher than the dimension of $F_t$ when the latent VARMA process (3) with non-zero orders $q_\Phi$ and $q_\Theta$ is specified in state space form. Model (5) can be expressed in terms of the state vector $\alpha_t$ via the observation equation

$$y_t = \mu + d_t + X_t\beta + Z\alpha_t + \varepsilon_t, \tag{7}$$

for $t = q_\Psi + 1, \ldots, T$, with

$$Z = \Lambda G. \tag{8}$$

To handle the initial stretch of observations $y_1, \ldots, y_{q_\Psi}$ explicitly, we need to consider the observation equation (7) with different system matrices $Z$ and $\Sigma_\varepsilon$ 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 $y_t = \Lambda_0 f_t + u_t$ ($t = 1, \ldots, T$) with $N \times r$ factor loading matrix $\Lambda_0$ 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 modeled as independent AR(1) processes, that is equation* (4) *with $q_\Psi = 1$ and both $\Psi_1$ and $\Sigma_\varepsilon$ diagonal. To ensure that all parameters are identified, we set $\Lambda_0 = (\bar{\Lambda}_1', \bar{\Lambda}_2')'$ 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 set to one. Additionally, we restrict $\Sigma_\zeta$ to a diagonal matrix.*

*The state vector* $\alpha_t$ *is specified as* $\alpha_t = F_t = (f_t', f_{t-1}')'$ *so that* $G = I_{2p}$ *in* (8). *The matrices* $H$ *and* $R$ *in* (6) *are given by*

$$H = \begin{bmatrix} \Phi_1 & 0 \\ I_r & 0 \end{bmatrix}, \qquad R = \begin{bmatrix} I_r \\ 0 \end{bmatrix}.$$

*Further, we have* $\mathbb{E}(\alpha_1) = 0$ *and* $Var(\alpha_1) = Q$ *is set to the unconditional variance of the stationary vector series* $(f_t', f_{t-1}')'$. *The observation equation* (7) *for* $t = 2, \ldots, T$ *has* $\mu = 0$, $d_t = \Psi_1 y_{t-1}$, $\beta = 0$ *and* $Z = \Lambda = (\Lambda_0\,,\, -\Psi_1\Lambda_0)$. *Since* $\mathbb{E}(u_1) = 0$ *and* $\mathbb{E}(u_1 u_1') = (I_N - \Psi_1^2)^{-1}\Sigma_\varepsilon$, *observation equation* (7) *for* $t = 1$ *has* $d_1 = 0$, $Z = (\Lambda_0\,,\, 0)$ *and* $\varepsilon_1 = u_1$.

**Illustration 2.** *Suppose* $y_t$ *is modeled by the dynamic factor model* $y_t = \Lambda_0 f_t + u_t$ *(*$t = 1, \ldots, T$*) with* $N \times r$ *factor loading vector* $\Lambda_0$ *and the latent factor* $f_t$ *is modeled by the VARMA(1,1) process defined as* (3) *with* $r = q_\Phi = q_\Theta = 1$. *We have*

$$f_t = \Phi f_{t-1} + \zeta_t + \Theta\zeta_{t-1}, \qquad \mathbb{E}(\zeta_t|\mathcal{F}_{t-1}) = 0, \qquad \mathbb{E}(\zeta_t^2|\mathcal{F}_{t-1}) = \Sigma_\zeta,$$

*where* $\Phi$, $\Theta$ *and* $\Sigma_\zeta$ *are unknown matrices. Furthermore, we suppose that the idiosyncratic components are independent disturbances, that is* $q_\Psi = 0$ *and* $u_t = \varepsilon_t$ *in* (4). *Identifiability of parameters is guaranteed by restricting* $\Lambda_0$ *and* $\Sigma_\zeta$ *as in Illustration 1.*

*In case* $r = 1$ *and with* $\Phi = \phi$, $\Theta = \theta$ *and* $\Sigma_\zeta = \sigma_\zeta^2$, *the state vector* $\alpha_t$ *is specified as* $\alpha_t = (f_t\,,\, \theta\zeta_t)'$ *and since* $F_t = f_t$ *we have* $\Lambda = \Lambda_0$ *in* (5) *and* $G = (1\,,\, 0)$ *in* (8). *Then,* $\mu = 0$, $d_t = 0$, $\beta = 0$ *and* $Z = \Lambda G$ *in* (7) *for* $t = 1, \ldots, T$. *The matrices* $H$ *and* $R$ *in* (6) *are then given by*

$$H = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix}, \qquad R = \begin{pmatrix} 1 \\ \theta \end{pmatrix},$$

*with initial conditions* $\mathbb{E}(\alpha_1) = 0$ *and* $Var(\alpha_1) = Q$ *such that* $Q - HQH' = \sigma_\zeta^2 RR'$.

# 3 Parameter estimation

The state space model (6) and (7) can be written in the form

$$y = \tilde{X}(\theta)\gamma + \xi, \tag{9}$$

where $y = [\,(y_1 - d_1)', \ldots, (y_T - d_T)'\,]'$, $\gamma = (\mu'\,,\, \beta')'$, $\tilde{X}(\theta)$ is a $NT \times (K+N)$ matrix valued function of $\theta$, $\mathbb{E}(\xi) = 0$ and $\mathbb{E}(\xi\xi') = \Sigma(\theta)$ for a matrix valued function $\Sigma(\theta)$. The Gaussian

loglikelihood function is defined as follows

$$\ell(y;\psi) = c - \frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}\{y - \tilde{X}(\theta)\gamma\}'\Sigma(\theta)^{-1}\{y - \tilde{X}(\theta)\gamma\}, \tag{10}$$

where $c$ is a constant independent of $\gamma$ and $\theta$. In case the disturbances $\varepsilon_t$ and $\zeta_t$ in the dynamic factor model (2) – (4) are Gaussian, equation (10) is the exact loglikelihood function. In other cases, the loglikelihood function is generally intractable. If the Gaussian assumption does not apply, the likelihood is designated as a quasi-likelihood. Quasi-maximum likelihood (QML) estimators of the parameters are obtained by maximizing (10) with respect to $\psi$. These QML estimators are strongly consistent as $T \to \infty$ under the assumptions of Section 2. Additionally, the QML estimators are asymptotically Gaussian, see Hannan, Dunsmuir, and Deistler (1980) for details and proofs.

For models where the Gaussian assumption for the initial factor $f_1$ and the disturbances $\varepsilon_t$ and $\zeta_t$ holds, Tunnicliffe-Wilson (1989) and Shephard (1993) show that maximum likelihood estimators of $\theta$ can be severely biased in small samples. Sample size can be small in macroeconomic applications where the dynamic factor model is considered. For example, Quah and Sargent (1993) analyze a data-set where $T$ is as small as 42. In small samples, estimators obtained by maximizing the marginal Gaussian likelihood are known to have better properties.

Suppose $J_1$ is a $(NT - N - K) \times NT$ matrix, such that the distribution of $J_1 y$ does not depend on $\gamma$. Given a $NT \times NT$ non-singular matrix of the form $J = (J_1' \ , \ J_2')'$ with $|J| = 1$, we define the marginal Gaussian loglikelihood function by

$$\ell_d(y;\theta) = \ell(J_1 y;\theta) = \ell(y;\psi) - \ell(J_2 y;\psi). \tag{11}$$

It can be shown that the marginal likelihood is invariant to the choice of $J$, see Ansley and Kohn (1985). This definition of the marginal Gaussian likelihood is equivalent to the conditional likelihood of Cox and Reid (1987) and the modified profile likelihood of Barndorff-Nielsen (1983), see also the discussion in Bellhouse (1990). In case of a small $T$ we prefer to estimate the parameters by maximizing (11).

Since the marginal likelihood does not depend on $\gamma$, we need to estimate these parameters separately, for example by generalized least squares (GLS). The GLS estimator and its variance are given by

$$\hat{\gamma}(\theta) = \{\tilde{X}'\Sigma^{-1}\tilde{X}\}^{-1}\tilde{X}'\Sigma^{-1}y, \qquad \mathrm{Var}_\theta(\hat{\gamma}(\theta)) = \{\tilde{X}'\Sigma^{-1}\tilde{X}\}^{-1}, \tag{12}$$

where we suppressed the dependence of $\tilde{X}$ and $\Sigma$ on $\theta$. Suppose $\hat{\theta}$ denotes the estimator of $\theta$ obtained by maximizing $\ell_d(y;\theta)$. We can then obtain an estimator of $\gamma$ by substituting $\hat{\theta}$ in (12). In case the QML estimators of $\gamma$ and $\theta$, $\bar{\gamma}$ and $\bar{\theta}$, respectively, are obtained by maximizing (10), we have $\hat{\gamma}(\bar{\theta}) = \bar{\gamma}$.

In the next section we present computationally efficient algorithms for evaluating the Gaussian loglikelihood function (10), the marginal Gaussian loglikelihood function (11) and the GLS estimator (12) for any value of $\theta$.

# 4 Estimation of factors and likelihood evaluation

Since the dynamic factor model can be represented as the state space model (6) and (7), we can use the Kalman filter to obtain the Gaussian loglikelihood function (10). Furthermore, the Kalman filter and smoothing (KFS) methods of Appendix A.1 can be adopted to evaluate

$$a_{t|s} = P(\alpha_t|y_1,\ldots,y_s;\psi), \qquad Q_{t|s} = \mathbb{E}\left[(\alpha_t - a_{t|s})(\alpha_t - a_{t|s})'|y_1,\ldots,y_s;\psi\right], \qquad (13)$$

for $s,t = 1,\ldots,T$, where $P(\alpha_t|y_1,\ldots,y_s;\psi)$ denotes the minimum mean squared error linear estimator (MMSLE) of $\alpha_t$ based on $y_1,\ldots,y_s$ for given $\psi$ and $Q_{t|s}$ is its mean squared error matrix. Here we have suppressed the dependence of $a_{t|s}$ and $Q_{t|s}$ on the parameter vector $\psi$. 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).

If the dimension $N$ of $y_t$ is very large, the KFS methods are computationally infeasible, even when the dimension $p$ of the state vector $\alpha_t$ is modest. Anderson and Moore (1979) and Koopman and Durbin (2003) show that, for models with diagonal $\Sigma_\varepsilon$, 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.

In this section we present new results that allow for the computationally efficient evaluation of the likelihood functions and GLS estimator of Section 3 as well as the state estimates and mean squared errors in (13).

## 4.1 Transforming the observation equation

Consider the state space model (6) and (7) for a given parameter vector $\psi$. Define $y_t^+ = Ay_t$, for $t = 1,\ldots,T$, for some non-singular matrix $A$. The MMSLEs of $\alpha_1,\ldots,\alpha_T$ in (13) 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 $A$, factor estimates and likelihood functions can be computed more efficiently based on $y_1^+, \ldots, y_T^+$ rather than $y_1, \ldots, y_T$.

Suppose we partition $N \times N$ matrix $A$ and $N \times 1$ vector $y_t^* = A(y_t - \mu - d_t - X_t\beta)$ as

$$A = \begin{bmatrix} A^L \\ A^H \end{bmatrix}, \qquad y_t^* = \begin{pmatrix} y_t^L \\ y_t^H \end{pmatrix}, \tag{14}$$

where

$$y_t^L = A^L(y_t - \mu - d_t - X_t\beta), \qquad y_t^H = A^H(y_t - \mu - d_t - X_t\beta),$$

with $m \times N$ matrix $A^L$ and $(N-m) \times N$ matrix $A^H$. 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 $\alpha_t$. More specifically, the model for $y_t^*$ will be of the form

$$y_t^L = A^L Z \alpha_t + e_t^L, \qquad y_t^H = e_t^H, \tag{15}$$

where

$$\mathbb{E}(e_t^L|\mathcal{F}_{t-1}) = 0, \qquad \mathbb{E}(e_t^H|\mathcal{F}_{t-1}) = 0,$$

$$\mathbb{E}(e_t^L e_t^{L\prime}|\mathcal{F}_{t-1}) = \Sigma_L, \qquad \mathbb{E}(e_t^H e_t^{H\prime}|\mathcal{F}_{t-1}) = \Sigma_H, \qquad \mathbb{E}(e_t^H e_t^{L\prime}|\mathcal{F}_{t-1}) = 0,$$

for $t = 1, \ldots, T$, with $\Sigma_L = A^L \Sigma_\varepsilon A^{L\prime}$ 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 $e_t^L$ and $e_t^H$ in (15) are uncorrelated and condition (iii) implies that the second equation does not depend on $\alpha_t$. Condition (iii) is stronger than strictly necessary. The transformed model will still be of the form (15) 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 (15). Finally, we add the condition

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

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

$$|A|^2 = |\Sigma_\varepsilon|^{-1}|A\Sigma_\varepsilon A'| = |\Sigma_\varepsilon|^{-1}|A^L\Sigma_\varepsilon A^{L\prime}||A^H\Sigma_\varepsilon A^{H\prime}| = |\Sigma_\varepsilon|^{-1}|\Sigma_L|. \tag{16}$$

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

**Lemma 1.** *Consider model* (6) – (7). *Suppose matrix $A$ is of the form* (14) *and $A^H$ satisfies (iii), then $A$ satisfies (i)–(iii) if and only if*

$$A^L = \Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}, \tag{17}$$

*where the columns of the $N \times r_\Lambda$ matrix $\Lambda^\dagger$ form a basis for the column space of $\Lambda$.*

**Remarks**

(a) The columns of $\Lambda^\dagger$ also form a basis of the column space of $Z$, which follows from the fact that $Z = \Lambda G$, for a full row rank matrix $G$. It is therefore easily verified that any matrix $A$ with $A^L$ given by (17) and $A^H$ satisfying (iii), fulfills conditions (i)–(iii). We prove the necessity part of Lemma 1 in Appendix A.2.

(b) Since column rank deficiency of $\Lambda$ is rare in practice, we can generally choose

$$\Lambda^\dagger = \Lambda C, \tag{18}$$

for any $r_\Lambda \times r_\Lambda$ nonsingular matrix $C$. In case $\Lambda$ does not have full column rank, it is often straightforward to construct a suitable $\Lambda^\dagger$. An example of such a situation is Illustration 1 if $\Psi_1 = \varphi I_N$ with scalar $-1 < \varphi < 1$.

(c) A closed form expression for $A^H$ is generally not available. For $A^H$ to satisfy (iii), we need to choose $A^H$ 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 will show that matrix $A^H$ is not required for any of our computations.

(d) The results below are based on transformation (14) and model (15). Although our results are more general and are developed for different purposes, a similar transformation as (14) for a different class of factor models is considered by Fiorentini, Sentana, and Shephard (2004, section 2.4.1).

**Illustration 3.** *Consider the dynamic factor model $y_t = \Lambda f_t + \varepsilon_t$ of Illustration 2. Apply transformation* (14) *to $y_t$ where matrix $A^L$ is given by* (17) *and* (18) *with $C = (\Lambda' \Sigma_\varepsilon^{-1} \Lambda)^{-1}$. For this choice of $C$, vector $y_t^L$ is effectively the GLS estimator of $f_t$ in the "regression model" $y_t = \Lambda f_t + \varepsilon_t$, for each $t$. We have*

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

*In case $r = r_\Lambda = 1$, model* (15) *for the univariate time series $y_t^L$ is then given by*

$$y_t^L = G\alpha_t + e_t^L, \qquad \mathbb{E}(e_t^L e_t^{L\prime} | \mathcal{F}_{t-1}) = C, \qquad t = 1, \ldots, T,$$

*where vector $G = (1, 0)$.*

## 4.2 Estimation of factors

By considering a matrix $A$ that satisfies the conditions (i)–(iv) in Section 4.1, we are able to efficiently compute MMSLEs of the factors. Since matrix $A$ has full rank, we have $P(\alpha_t | y_1, \ldots, y_s; \psi) = P(\alpha_t | y_1^*, \ldots, y_s^*; \psi)$. Furthermore, from (15) it follows that $y_t^L$ and $y_t^H$ are uncorrelated and that $y_t^H$ does not depend on $\alpha_t$. Hence,

$$a_{t|s} = P(\alpha_t | y_1^*, \ldots, y_s^*; \psi) = P(\alpha_t | y_1^L, \ldots, y_s^L; \psi),$$

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

$$y_t^L = A^L Z \alpha_t + e_t^L, \qquad \mathbb{E}(e_t^L | \mathcal{F}_{t-1}) = 0, \quad \mathbb{E}(e_t^L e_t^{L\prime} | \mathcal{F}_{t-1}) = \Sigma_L, \qquad t = 1, \ldots, T. \tag{19}$$

The high-dimensional matrix $A^H$ and vector $y_t^H$ are not required for the estimation of $\alpha_t$. In case of Illustration 3 with $r = 1$, the estimation of $\alpha_t$ is simply carried out by univariate KFS methods. The low-dimensional KFS also produces the correct mean squared error matrices $Q_{t|s}$ in (13) for $s, t = 1, \ldots, T$. The KFS methods provide solutions for prediction ($s = t-1$), filtering ($s = t$), smoothing ($s = T$) and forecasting ($t > s$) of observation and state vectors.

The procedures of this section can still be used if observed vectors $y_t$ do not all have the same dimension due to, for example, missing values. In this case, a different matrix $A$ must be constructed for $t = 1, \ldots, T$. This solution is also adopted in cases where the system matrices of the state space form (6) – (7) vary over time.

## 4.3 Evaluation of the Gaussian loglikelihood

Let $v_t$ denote the prediction error, $v_t = y_t - P(y_t|y_1, \ldots, y_{t-1}; \psi)$, and $D_t$ the mean squared error matrix of $v_t$, $D_t = \mathbb{E}(v_t v_t'|y_1, \ldots, y_{t-1}; \psi)$. The Gaussian loglikelihood $\ell(y; \psi)$ defined in (10) can be evaluated via the prediction error decomposition,

$$\ell(y;\psi) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|D_t| - \frac{1}{2}\sum_{t=1}^{T} v_t' D_t^{-1} v_t, \tag{20}$$

see Schweppe (1965) and Harvey (1989, section 3.4). The quantities $v_t$ and $D_t$ are obtained from the Kalman filter of Appendix A.1 applied to the state space model (6) and (7).

A computationally more efficient way to evaluate (20) is to choose a matrix $A$ that satisfies the conditions (i)–(iv) in Section 4.1, to transform $y_t$ as in (14) and to consider model (15). We then have

$$\ell(y;\psi) = \ell(y^L;\psi) + \ell(y^H;\psi) + T\log|A|, \tag{21}$$

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 $\ell(y^L;\psi)$ can be evaluated by the Kalman filter applied to the low-dimensional model (19). The second term is

$$\ell(y^H;\psi) = -\frac{(N-m)T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T} y_t^{H\prime}\Sigma_H^{-1} y_t^H, \tag{22}$$

since $|\Sigma_H| = 1$. Lemma 2 shows that the last term in equation (22) can be calculated without constructing $A^H$. The proof is given in Appendix A.3.

**Lemma 2.** *For the state space model* (6) – (7)*, transformation* (14) *and resulting model* (15)*, with* $A^L$ *given by* (17)*, we have the identity*

$$y_t^{H\prime}\Sigma_H^{-1} y_t^H = e_t'\Sigma_\varepsilon^{-1} e_t, \tag{23}$$

*where* $e_t = \left[I_N - \Lambda^\dagger\left(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger\right)^{-1}\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\right](y_t - d_t - \mu - X_t\beta)$.

Given the expression for $|A|^2$ in (16), loglikelihood function (21) can be expressed as

$$\ell(y;\psi) = c + \ell(y^L;\psi) - \frac{T}{2}\log\frac{|\Sigma_\varepsilon|}{|\Sigma_L|} - \frac{1}{2}\sum_{t=1}^{T} e_t'\Sigma_\varepsilon^{-1} e_t, \tag{24}$$

where $c$ is a constant independent of both $y$ and $\psi$. It follows that for the evaluation of the loglikelihood, computation of matrix $A^H$ and vectors $y_t^H$, for $t = 1, \ldots, T$, is not

required. Expression (24) is instrumental for a computationally feasible approach to the quasi-likelihood based analysis of the dynamic factor model.

**Remarks**

(a) The vectors $e_t$ in Lemma 2 have an intuitive interpretation as the residuals of a GLS regression of $y_t - X_t\beta - \mu - d_t$ on the columns of $\Lambda^\dagger$ with variance matrix $\Sigma_\varepsilon$. Since the columns of $\Lambda^\dagger$ also form a basis of the column space of $Z$, this is equivalent to regressing $y_t - X_t\beta - \mu - d_t$ on the columns of $Z$.

(b) The concluding remarks of Section 4.2 concerning missing values and time-varying specifications of the dynamic factor model apply to the evaluation of the loglikelihood via (21) as well.

**Illustration 4.** *In the Illustration 3, the transformation* (14) *is based on the matrix* $A^L$ *defined in* (17) *and* (18) *with* $C = (\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}$. *However, it can be 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* $\Sigma_L$ *in* (19) *is the identity matrix and the loading matrix in* (19) *is* $A^L\Lambda^\dagger = C^{-1\prime}$. *We obtain the model*

$$y_t^L = C^{-1\prime}G\alpha_t + e_t^L, \qquad \mathbb{E}(e_t^L|\mathcal{F}_{t-1}) = 0, \quad \mathbb{E}(e_t^L e_t^{L\prime}|\mathcal{F}_{t-1}) = I, \qquad t = 1, \ldots, T.$$

*Since* $\Sigma_L = I$, *the fast KFS methods discussed in Koopman and Durbin (2003) can be applied straightforwardly. Furthermore, the loglikelihood function* (24) *reduces to*

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

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

## 4.4 Partial concentration of regression coefficients

Maximizing the Gaussian loglikelihood function $\ell(y;\psi)$ is computationally intensive, since the dimension of $\psi$ is generally very high. It is therefore attractive to concentrate the regression coefficients out of the likelihood and maximize the resulting profile likelihood function. In this section we show how the constant vector $\mu$ can be partially concentrated out of the likelihood with minimum effort.

Choose matrix $A$ such that conditions (i)–(iii) in Section 4.1 are satisfied and define

$$\tilde{y}_t^L = A^L(y_t - d_t - X_t\beta), \qquad \tilde{y}_t^H = A^H(y_t - d_t - X_t\beta), \qquad t = 1, \ldots, T, \tag{25}$$

such that $y_t^L = \tilde{y}_t^L - \mu^L$ and $y_t^H = \tilde{y}_t^H - \mu^H$ where $\mu^L = A^L\mu$ and $\mu^H = A^H\mu$, for $t = 1, \ldots, T$. In the likelihood function (21), $\ell(y^L; \psi)$ does not depend on $\mu^H$ while $\mu^H$ only appears in the second term of $\ell(y^H; \psi)$ which can be expressed by

$$-\frac{1}{2}\sum_{t=1}^{T}(\tilde{y}_t^H - \mu^H)'\Sigma_H^{-1}(\tilde{y}_t^H - \mu^H) = -\frac{1}{2}\sum_{t=1}^{T}(\tilde{e}_t - M_\Lambda\mu)'\Sigma_\varepsilon^{-1}(\tilde{e}_t - M_\Lambda\mu), \tag{26}$$

where $M_\Lambda = I_N - \Lambda^\dagger(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}$ and $\tilde{e}_t = M_\Lambda(y_t - d_t - X_t\beta)$ such that $e_t = \tilde{e}_t - M_\Lambda\mu$. The equality in (26) is justified by Lemma 2. It follows from equation (43) in Appendix A.3 that $M_\Lambda\mu$ is a linear function of $\mu^H$. Concentrating out $\mu^H$ is therefore equivalent to concentrating out $M_\Lambda\mu$ from the likelihood function. The GLS estimator of $M_\Lambda\mu$, denoted by $\hat{\mu}_{\perp\Lambda}(\beta, \theta)$, is given by

$$\hat{\mu}_{\perp\Lambda}(\beta, \theta) = \frac{1}{T}\sum_{t=1}^{T}\tilde{e}_t. \tag{27}$$

The (partial) profile loglikelihood function is given by (24) where the last term is replaced by $-0.5\sum_{t=1}^{T}\tilde{e}_t^{m\prime}\Sigma_\varepsilon^{-1}\tilde{e}_t^m$ where $\tilde{e}_t^m = \tilde{e}_t - \hat{\mu}_{\perp\Lambda}(\beta, \theta)$ for $t = 1, \ldots, T$.

The QML estimator of $\mu$ can be obtained via the identity

$$\mu = P_\Lambda\mu^L + M_\Lambda\mu, \qquad \text{where} \quad P_\Lambda = \Lambda^\dagger(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}. \tag{28}$$

The QML estimator of $\mu$ is then given by

$$\tilde{\mu} = \hat{\mu}_{\perp\Lambda}(\tilde{\beta}, \tilde{\theta}) + P_\Lambda\tilde{\mu}^L, \tag{29}$$

where $\tilde{\theta}$, $\tilde{\beta}$ and $\tilde{\mu}^L$ are the QML estimators of $\theta$, $\beta$ and $\mu^L$, respectively, which we obtain by maximizing the profile Gaussian loglikelihood function with respect to $\theta$, $\beta$ and $\mu^L$.

## 4.5 Evaluation of the marginal Gaussian loglikelihood

Different algorithms have been proposed to evaluate the marginal Gaussian loglikelihood function (11), see Ansley and Kohn (1985) and de Jong (1991). The algorithms, consisting of the Kalman filter and smoother augmented by an additional set of $N + K$ dimensional recursions, implicitly carry out the transformation $J_1 y$ in (11). The augmented KFS (AKFS)

methods also produce MMSLEs of regression coefficients together with their mean squared errors as given by (12).

The number of time series in a dynamic factor model can be high and direct application of such algorithms is often infeasible. We can use the earlier results to compute the marginal likelihood based on a much smaller dimension. As a result, inference based on the marginal likelihood becomes feasible for a high-dimensional dynamic factor model (6) – (7). For this purpose, we choose matrix $A$ such that conditions (i)–(iv) in Section 4.1 are satisfied. Pre-multiplying the observations by $A$, we obtain the model

$$\bar{y}_t^L = \mu^L + X_t^L \beta + A^L Z \alpha_t + e_t^L, \qquad \bar{y}_t^H = \mu^H + X_t^H \beta + e_t^H, \tag{30}$$

where $\bar{y}_t^L = A^L(y_t - d_t)$, $\bar{y}_t^H = A^H(y_t - d_t)$, $X_t^L = A^L X_t$ and $X_t^H = A^H X_t$ with $\mu^L$ and $\mu^H$ defined below (25) and the disturbances $e_t^L$ and $e_t^H$ defined below (15). In the remainder of this section we show that the evaluation of $\ell_d(y;\theta)$ can be carried out in two steps: first processing the original time series $y_t$ and second applying AKFS methods to the time series $\bar{y}_t^L$.

Since $\bar{y}^L = (\bar{y}_1^{L\prime}, \ldots, \bar{y}_T^{L\prime})'$ and $\bar{y}^H = (\bar{y}_1^{H\prime}, \ldots, \bar{y}_T^{H\prime})'$ both depend on coefficient vector $\beta$, the marginal Gaussian likelihood function cannot be easily expressed in two independent parts. However, we show in Appendix A.4 that for any given parameter vector $\theta$

$$\ell_d(y;\theta) = L_d(\bar{y}^H;\theta) + L_d(\bar{y}^L;\theta) - \frac{T-1}{2}\log\frac{|\Sigma_\varepsilon|}{|\Sigma_L|}, \tag{31}$$

where $L_d(\bar{y}^H;\theta)$ and $L_d(\bar{y}^L;\theta)$ are obtained by the following two-step algorithm.

**Step 1.** Define

$$b = B^{-1}\sum_{t=1}^{T} \bar{X}_t^{m\prime}\Sigma_\varepsilon^{-1}\bar{e}_t^m, \qquad B = \sum_{t=1}^{T} \bar{X}_t^{m\prime}\Sigma_\varepsilon^{-1}\bar{X}_t^m, \tag{32}$$

where

$$\bar{e}_t^m = M_\Lambda(y_t - d_t - \bar{y}), \qquad \bar{X}_t^m = M_\Lambda(X_t - \bar{X}), \tag{33}$$

for $t = 1, \ldots, T$, with $\bar{y} = T^{-1}\sum_{t=1}^{T}(y_t - d_t)$ and $\bar{X} = T^{-1}\sum_{t=1}^{T} X_t$. Then, compute

$$L_d(\bar{y}^H) = -\frac{(N-m)}{2}(T\log 2\pi + \log T) - \frac{1}{2}\log|B| - \frac{1}{2}\sum_{t=1}^{T}\bar{e}_t^{*\prime}\Sigma_\varepsilon^{-1}\bar{e}_t^*, \tag{34}$$

where $\bar{e}_t^* = \bar{e}_t^m - \bar{X}_t^m b$.

**Step 2.** Set $L_d(\bar{y}^L;\theta)$ equal to the marginal Gaussian loglikelihood for the model

$$\bar{y}_t^L = \mu^L + X_t^L\delta + A^L Z\alpha_t + e_t^L, \tag{35}$$

where $\mu^L$ is treated as an unknown regression coefficient vector and $\delta$ is a random effect with mean $b$ and variance $B^{-1}$. The evaluation of $L_d(\bar{y}^L;\theta)$, together with the estimation of $\mu^L$ and $\delta$, is carried out by the augmented Kalman filter.

## 4.6 Estimation of regression coefficients

For the dynamic factor model (6) – (7) define $\hat{\mu}(\theta)$ and $\hat{\beta}(\theta)$ as the GLS estimators of $\mu$ and $\beta$ as functions of $\theta$. Note that $\hat{\gamma}(\theta) = [\hat{\mu}(\theta)', \hat{\beta}(\theta)']'$ where the GLS estimator $\hat{\gamma}(\theta)$ is given by (12). The GLS estimators are based on the data-set $y_1, \ldots, y_T$. The two-step algorithm of the previous section is also instrumental for computing $\hat{\mu}(\theta)$ and $\hat{\beta}(\theta)$ in a computationally efficient way. The second step of the algorithm produces the MMSLEs of $\delta$ and $\mu^L$, that is $P(\delta|\bar{y}_1^L, \ldots, \bar{y}_T^L;\theta)$ and $P(\mu^L|\bar{y}_1^L, \ldots, \bar{y}_T^L;\theta)$, respectively, as well as their mean squared errors. In Appendix A.5 we prove that for given $\theta$

$$\hat{\beta}(\theta) = P(\delta|\bar{y}_1^L, \ldots, \bar{y}_T^L;\theta), \qquad \hat{\mu}^L(\theta) = P(\mu^L|\bar{y}_1^L, \ldots, \bar{y}_T^L;\theta), \tag{36}$$

where $\hat{\mu}^L(\theta)$ is defined as the GLS estimator of $\mu^L$ based on $y_1, \ldots, y_T$ as a function of $\theta$. The GLS estimator $\hat{\mu}(\theta)$ is follows from (28) and is given by

$$\hat{\mu}(\theta) = P_\Lambda \hat{\mu}^L(\theta) + \hat{\mu}_{\perp\Lambda}(\theta), \tag{37}$$

where matrix $P_\Lambda$ is defined in (28) and $\hat{\mu}_{\perp\Lambda}(\theta)$ denotes $\hat{\mu}_{\perp\Lambda}(\beta,\theta)$, as given by (27), evaluated in $\beta = \hat{\beta}$. The variance matrix of $\hat{\mu}(\theta)$ is given by

$$\begin{aligned}
\mathrm{Var}_\theta(\hat{\mu}(\theta)) &= \mathrm{Var}_\theta\{\hat{\mu}_{\perp\Lambda}(\theta) + P_\Lambda\hat{\mu}^L(\theta)\} \\
&= \mathrm{Var}_\theta\{\hat{\mu}_{\perp\Lambda}(\theta)\} + P_\Lambda \mathrm{Cov}_\theta\{\hat{\mu}_{\perp\Lambda}(\theta), \hat{\mu}^L(\theta)\}' + \mathrm{Cov}_\theta\{\hat{\mu}_{\perp\Lambda}(\theta), \hat{\mu}^L(\theta)\}P_\Lambda' \\
&\quad + P_\Lambda \mathrm{Var}_\theta\{\hat{\mu}^L(\theta)\}P_\Lambda',
\end{aligned}$$

where the dependence of variances and covariances on $\theta$ is made explicit in the notation $\mathrm{Var}_\theta(\cdot)$ and $\mathrm{Cov}_\theta(\cdot,\cdot)$. To evaluate $\mathrm{Var}_\theta\{\hat{\mu}(\theta)\}$, we require the expressions

$$\mathrm{Var}_\theta\{\hat{\mu}_{\perp\Lambda}(\theta)\} = M_\Lambda\left[\bar{X}\mathrm{Var}_\theta\{\hat{\beta}(\theta)\}\bar{X}' + T^{-1}\Sigma_\varepsilon\right]M_\Lambda',$$

and

$$\mathrm{Cov}_\theta\{\hat{\mu}_{\perp\Lambda}(\theta), \hat{\mu}^L(\theta)\} = -M_\Lambda \bar{X} \mathrm{Cov}_\theta\{\hat{\beta}(\theta), \hat{\mu}^L(\theta)\},$$

where $M_\Lambda$ and $\bar{X}$ are defined below (26) and (33), respectively. The variance matrices $\mathrm{Var}_\theta\{\hat{\beta}(\theta)\}$ and $\mathrm{Var}_\theta\{\hat{\mu}^L(\theta)\}$ are equal to the mean squared error matrices of $\hat{\beta}(\theta)$ and $\hat{\mu}^L(\theta)$ in (36), respectively. These two variance matrices, together with the covariance matrix $\mathrm{Cov}_\theta\{\hat{\beta}(\theta), \hat{\mu}^L(\theta)\}$, are evaluated by the AKFS methods from Step 2 of the algorithm in Section 4.5. Derivations and more details are given in Appendix A.5. Estimators of $\mu$ and $\beta$ can be obtained by substituting the QML estimator of $\theta$, found by maximizing the Gaussian marginal likelihood, in (36) and (37).

The AKFS methods applied to the low-dimensional model in step 2 of the algorithm in Section 4.5 also produce the MMSLEs of the state vectors $\alpha_1, \ldots, \alpha_T$ and the mean squared errors as defined in (13). The resulting estimates are identical to the estimates obtained from the usual KFS with $\mu$ and $\beta$ replaced by their GLS estimators $\hat{\beta}$ and $\hat{\mu}$, respectively. However, the mean squared errors of the state vector will not be identical. The AKFS methods adjust the mean squared errors of the state vector for the uncertainty in the estimates of the regression coefficients in $\mu$ and $\beta$.

## 4.7 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 based on the transformed observations $y_1^+, \ldots, y_T^+$ with $y_t^+ = Ay_t$ for $t = 1, \ldots, T$. The gains are relative to the standard application of the Kalman filter based on $y_1, \ldots, y_T$.

The computational gains depend primarily on the panel dimension $N$ and state vector dimension $p$. To obtain some insight in the size of these gains, we calculate the Gaussian loglikelihood and marginal Gaussian loglikelihood functions for different values of $N$ and $p$. The calculations are performed using the Kalman filter and the methods described in Sections 4.3 and 4.5 for the basic factor model given by

$$y_{it} = \mu_i + \lambda_i' f_t + \varepsilon_{it},$$

where $f_t$ is modeled by the VARMA process (3) with $q_\Phi = 1$ and $q_\Theta = 0$ while the innovations $\varepsilon_{it}$ are uncorrelated. For the different model representations in Section 2, we have $\alpha_t = F_t = f_t$ and $p = m = r$. In the first panel of Table 1 we present the ratios of CPU times needed for the evaluation of the two loglikelihood functions. The results are encouraging. If $N = 250$

and $p = 5$, the Kalman filter computations for the loglikelihood are carried out 15 times faster as a result of our new device. Furthermore, 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 achieve even more computational gains if we evaluate the marginal loglikelihood using the method of Section 4.5. The reported ratios in the second panel of Table 1 are so high because the Kalman filter based on $y_1, \ldots, y_T$ requires an $N$ dimensional augmentation for the constant vector $\mu$. The Kalman filter used in Section 4.5 and based on the observation equation (35) requires a limited $p$ dimensional augmentation for the constant vector $\mu^L$.

Table 1: Computational Gains

The two panels below present the gains in computing time when evaluating the Gaussian likelihood respectively the marginal likelihood functions of a basic dynamic factor model. The model considered is $y_{it} = \mu_i + \lambda_i' f_t + \varepsilon_{it}$, where $f_t$ is modeled as a VAR(1) process, $\varepsilon_{it} \sim IID(0, \sigma^2)$, for some positive scalar $\sigma$ and $\mu_i$ is a scalar. The ratio $d_1/d_2$ is reported: $d_1$ is the CPU time for the standard Kalman filter respectively augmented Kalman filter and $d_2$ is CPU time for the algorithms of Sections 4.3 and 4.5. The ratios are reported for different panel dimensions $N$ and different state vector dimensions $p$.

| | Gaussian likelihood | | | | | Marginal Gaussian likelihood | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N \backslash p$ | 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 |

# 5 Maximizing the quasi-likelihood function

In this section we discuss methods to maximize $\ell(y; \psi)$ and $\ell_d(y; \theta)$ with respect to $\psi$ and $\theta$, respectively. The number of parameters can be as high as $1,000$ or $2,000$. The results of Sections 4.3 and 4.5 imply that the Gaussian loglikelihood and the marginal Gaussian loglikelihood functions can be evaluated efficiently for high-dimensional dynamic factor models. Numerical optimization procedures, such as the quasi-Newton BFGS algorithm described in Nocedal and Wright (1999), can be adopted to maximize the loglikelihood function with respect to $\psi$ or $\theta$. These methods require evaluation of the score vector. Since the number of parameters is high, evaluating the score vector numerically is infeasible, even if the results of Section 4.3 are used. Fortunately, we can show that the exact score vector can be obtained by a single KFS or AKFS applied to the low-dimensional model (19) or (35). Alternatively,

the EM algorithm can be used to obtain the QML estimates. In Section 5.2 we show that each EM step relies on a single KFS or AKFS.

## 5.1 Calculating the analytical score

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

$$\frac{\partial \ell(y;\psi)}{\partial \psi}\Big]_{\psi=\psi^*} = \frac{Q(\psi^*|\psi)}{\partial \psi}\Big]_{\psi=\psi^*}, \tag{38}$$

where $Q(\psi^*|\psi)$ is the expected complete Gaussian 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 (6) – (7), with Gaussian innovations $\varepsilon_t$ and $\zeta_t$, $Q(\psi^*|\psi)$ is given by

$$\begin{aligned} Q(\psi^*|\psi) = c &- \frac{T}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\mathrm{tr}Q_\varepsilon - \frac{T-1}{2}\log|\Sigma_\zeta| - \frac{1}{2}\mathrm{tr}Q_\zeta \\ &- \frac{1}{2}\log|P| - \frac{1}{2}\mathrm{tr}[P^{-1}\{(a_{1|T}-a)(a_{1|T}-a)' + Q_{1|T}\}], \end{aligned} \tag{39}$$

where $a = \mathbb{E}(\alpha_1)$, $P = \mathbb{E}\{(\alpha_1 - a)(\alpha_1 - a)'\}$ and $c$ is a constant independent of $\psi$ and

$$Q_\varepsilon = \Sigma_\varepsilon^{-1}\sum_{t=1}^{T}\{\hat{\varepsilon}_t\hat{\varepsilon}_t' + \mathrm{Var}(\varepsilon_t|y)\}, \qquad Q_\zeta = \Sigma_\zeta^{-1}\sum_{t=2}^{T}\{\hat{\zeta}_t\hat{\zeta}_t' + \mathrm{Var}(\zeta_t|y)\}, \tag{40}$$

where $\hat{\varepsilon}_t = \mathbb{E}(\varepsilon_t|y)$, $\mathrm{Var}(\varepsilon_t|y)$, $\hat{\zeta}_t = \mathbb{E}(\zeta_t|y)$ and $\mathrm{Var}(\zeta_t|y)$ can be expressed in terms of $a_{j|T}$ and $Q_{j|T}$ 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 (19) while matrix $A^H$ and time series $y_t^H$ are not needed, the KFS computations are fast. Expressions for the derivatives of (39) with respect to the system matrices, evaluated at $\psi = \psi^*$, are given in Appendix A.6. The score vector with respect to $\psi$ is then obtained using the chain rule. The score vector of the marginal Gaussian loglikelihood function $\ell_d(y;\theta)$ with respect to $\theta$ and evaluated at $\theta = \theta^*$ can be obtained in the same way with the difference that $Q(\theta^*|\theta)$ is obtained by AKFS methods as described in Section 4.5, see Durbin and Koopman (2001, section 7.3).

### 5.2 The EM algorithm

The well-known 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 (39).
- M step: maximize $Q(\psi^*|\psi)$ with respect to $\psi$.

The M step produces a vector $\psi^+$ 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. A feasible EM algorithm for high-dimensional dynamic factor models is obtained by applying the methods of Sections 4.1 and 4.2 in the E step. 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. The EM algorithm can also be adopted to find a (local) optimum of the marginal Gaussian loglikelihood function $\ell_d(y;\theta)$. The necessary modification is to adopt the two-step algorithm of Section 4.5 in the E step.

## 6 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). From this data-set we construct 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 15 sectors as presented in Table 2. Each sector is indexed by a code (A...O). Table 2 also reports the number of time series in each sector. For all series, observations larger than 6 times the standard deviation of the series, $\sigma$, (in absolute value) are set to $\pm 6\sigma$. In total, 46 (out of $69,696$) observations are Winsorized in this way (0.066%). Finally, each time series is scaled such that its sample variance equals one.

The empirical analyses below differ from Stock and Watson (2005) and the related study in Stock and Watson (2002b) since their results are obtained from principal components analyses (diffusion indexes). Our approach is closer in spirit to the likelihood-based analyses of Bernanke, Boivin, and Eliasz (2005) and Boivin and Giannoni (2006). The estimation

Table 2: List of sectors

This table lists the 15 sectors in the data-set that we consider in Section 6. For each sector, the code, a short description and the number of series in the sector are given. More detailed descriptions 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 |

of parameters is based on QML. The dynamic properties of the factors can be analyzed by investigating the estimated coefficients that are associated with the factors. We further show that diagnostic tests for model misspecification can be computed as part of a model-based analysis.

We consider the dynamic factor model which is given by

$$y_t = \bar{\mu} + \Lambda f_t + u_t, \qquad f_t = \Phi_1 f_{t-1} + \zeta_t, \qquad u_t = \Psi_1 u_{t-1} + \varepsilon_t, \tag{41}$$

for $t = 1, \ldots, T$ where $y_t$ is the $N \times 1$ observation vector and $f_t$ is the $r \times 1$ vector of factors with $T = 528$ and $N = 132$. The intercept vector $\bar{\mu}$ is fixed and unknown. The factor loading matrix $\Lambda$ is unknown and its $r$ top rows form a lower-triangular matrix with the diagonal elements restricted to be one. The matrices $\Phi_1$ and $\Psi_1$ are autoregressive coefficient matrices and we assume that $\Psi_1$ is a diagonal matrix. The innovation series $\zeta_t$ and $\varepsilon_t$ are IID processes with mean zero and variance matrices $\Sigma_\zeta$ and $\Sigma_\varepsilon$, respectively. We assume that both $\Sigma_\zeta$ and $\Sigma_\varepsilon$ are diagonal.

We adopt two model specifications: Model I has $r = 7$ and $\Psi_1 = 0$ (such that $u_t = \varepsilon_t$ is IID for all $t$) and Model II has $r = 4$ and a non-zero diagonal matrix $\Psi_1$. Model I is motivated by Stock and Watson (2005) where they adopt the procedure of Bai and Ng

(2002) to conclude that seven static factors are present in this data-set. Model II is motivated by an analysis of Bai and Ng (2007) based on the same data-set and where they advocate 4 dynamic factors which may span over 7 static factors. In Model II we therefore set $r = 4$ and obtain a set of $m = 8$ static factors by introducing autoregressive disturbances of order 1, that is $q_\Psi = 1$. Since the number of static factors in models I and II are comparable (7 and 8, respectively), it is interesting to compare the empirical findings for the two model specifications. The dimensions of the different model specifications and the dimensions of sub-vectors $\bar{\mu}$, $\beta$ and $\theta$ of the parameter vector $\psi$ are reported in Table 3.

Table 3: Dynamic factor model specifications

The table reports dimensions for two dynamic factor model specifications that we consider in the empirical analyses as well the corresponding parameter vectors $\psi$ and its components. The observed series, $y_t$, are modeled as $y_t = \bar{\mu} + \Lambda f_t + u_t$ where $f_t$ is a VAR(1) process, see Section 2 for more details. The state space formulation is discussed in Section 2. In model I, the innovation vector $u_t$ is an IID process with mean zero and diagonal variance matrix $\Sigma_\varepsilon$. In model II, the vector $u_t$ is modeled as a VAR(1) process with a diagonal autoregressive coefficient matrix $\Psi_1$ and a diagonal variance matrix $\Sigma_\varepsilon$. The dimension of $\theta$ is the total of all parameter dimensions excluding $\mu$ and $\beta$ since $\psi = (\bar{\mu}', \beta', \theta')'$.

| | Model formulation §2 | | | | | State Space | | | Parameter vector $\psi$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r$ | $q_\Lambda$ | $q_\Phi$ | $q_\Theta$ | $q_\Psi$ | $s$ | $m$ | $p$ | $\bar{\mu}$ | $\beta$ | $\Lambda$ | $\Phi$ | $\Theta$ | $\Psi$ | $\Sigma_\varepsilon$ | $\theta$ |
| I | 7 | 0 | 1 | 0 | 0 | 0 | 7 | 7 | 132 | 0 | 903 | 49 | 0 | 0 | 132 | 1084 |
| II | 4 | 0 | 1 | 0 | 1 | 1 | 8 | 8 | 132 | 0 | 522 | 16 | 0 | 132 | 132 | 802 |

## 6.1 Parameter estimation

We have estimated the parameters by numerically maximizing the Gaussian loglikelihood function $\ell(y; \psi)$ with respect to $\psi$ using the results of Sections 4 and 5. First we use the EM algorithm to find a point in the neigbourhood of the optimum using the devices of Appendix A.7. We then adopt the BFGS algorithm to maximize $\ell(y; \psi)$ with respect to $\psi$ by starting from the final iteration of the EM algorithm. The numerical maximization routine makes use of the analytical score calculations of Appendix A.6. On a standard computer with 3 GB memory and a 2.2 GHz two-core processor, this took less than 15 minutes for Model I and less than 10 minutes for Model II.

TABLE 4: QUASI-MAXIMUM LIKELIHOOD ESTIMATES OF VAR COEFFICIENTS

We report the QML estimates of the coefficients in the $r \times r$ matrix $\Phi_1$ for Models I ($r = 7$) and II ($r = 4$). The eigenvalues of the estimates of $\Phi_1$ are reported in descending order. For complex eigenvalues we present both the real and imaginary (img) components.

Model I

| | VAR coefficients | | | | | | | Eigenvalues | |
|---|---|---|---|---|---|---|---|---|---|
| Factor | 1 | 2 | 3 | 4 | 5 | 6 | 7 | real | img |
| 1 | 0.17 | -0.15 | 0.18 | -0.031 | -0.14 | 0.062 | -0.031 | 0.95 | 0 |
| 2 | -0.36 | 0.84 | -0.017 | 0.03 | 0.099 | 0.028 | 0.031 | 0.94 | 0.08 |
| 3 | 0.065 | 0.074 | 0.9 | 0.048 | 0.19 | 0.0069 | 0.034 | 0.94 | -0.08 |
| 4 | 0.068 | 0.051 | 0.034 | 0.92 | 0.045 | -0.031 | -0.017 | 0.91 | 0 |
| 5 | -0.075 | 0.025 | 0.014 | -0.073 | 0.25 | -0.1 | 0.043 | 0.28 | 0 |
| 6 | 0.003 | -0.022 | -0.029 | 0.036 | 0.003 | -0.33 | -0.012 | 0.042 | 0 |
| 7 | -0.024 | -0.027 | -0.038 | -0.0002 | -0.049 | -0.028 | 0.97 | -0.33 | 0 |

Model II

| | VAR coefficients | | | | Eigenvalues | |
|---|---|---|---|---|---|---|
| Factor | 1 | 2 | 3 | 4 | real | img |
| 1 | 0.29 | -0.23 | -0.12 | -0.1 | 0.94 | 0.094 |
| 2 | -0.38 | 0.44 | 0.031 | 0.13 | 0.94 | -0.094 |
| 3 | 0.086 | -0.43 | 0.96 | 0.17 | 0.33 | 0 |
| 4 | -0.64 | 0.33 | -0.019 | 0.56 | 0.046 | 0 |

The estimation process relies further on starting values for the parameters which for Model I are given by

$$\Lambda = (I_r \quad 0)', \quad \Sigma_\zeta = I_r, \quad \Sigma_\varepsilon = I_N, \quad \Phi_1 = 0.5 I_r, \quad \bar{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t,$$

with $r = 7$. For model II, the same starting values are adopted with $r = 4$ and with $\Psi_1 = 0.5 I_N$.

Table 4 presents the QML estimates of the VAR coefficients in $\Phi_1$ together with the corresponding eigenvalues for Models I and II. The factors in the models are organized in descending order of the eigenvalues of $\Phi_1$. We learn from Table 4 that the factors are estimated as stationary and highly persistent processes given the largest eigenvalue of 0.95. For both models, we find the presence of persistent cyclical behaviour in the factors since one conjugate pair of complex eigenvalues is obtained where the real part is equal to 0.94. The other eigenvalues range from large to small. As in any VAR analysis, it is hard to comment on individual coefficients in $\Phi_1$.

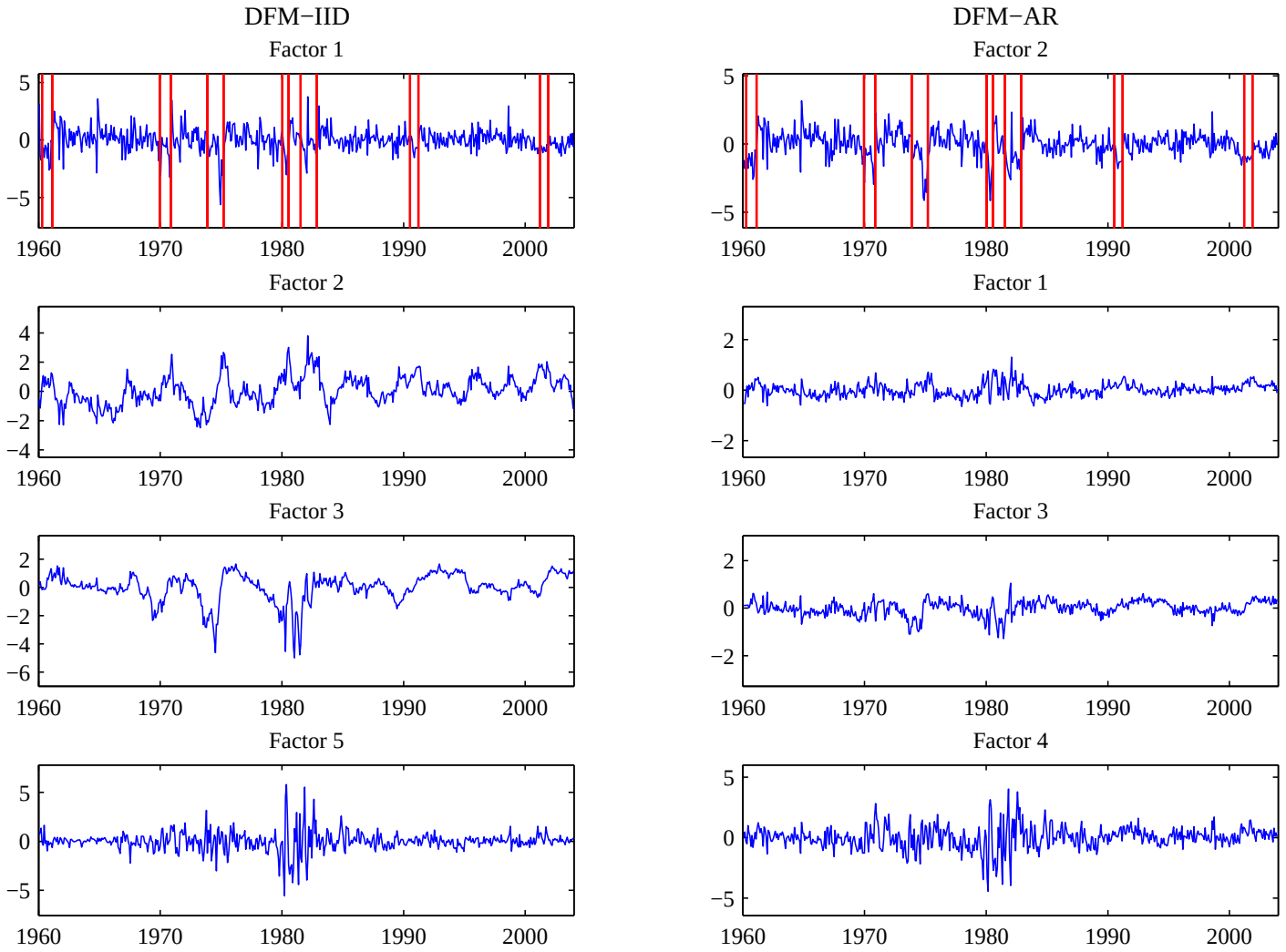
## 6.2 Factor estimates and factor loadings

The estimates of the factors are obtained by application of the low-dimensional KFS methods and are based on all observations in the data-set (smoothed estimates). To facilitate a clear interpretation of the factors we rotated the factors by means of the varimax method. The varimax method tries to construct a rotation such that the resulting factors are as distinct from each other as possible, see *e.g.* Lawley and Maxwell (1971) for details. To facilitate comparisons, we have selected the Factors 1, 2, 3 and 5 of Model I and all factors of Model II to be presented in Figure 1. The first two estimated factors of the two models are similar although the amount of noise in the factors is somewhat different. It is expected that the first factor is associated with the business cycle and is therefore displayed with the NBER business cycle reference dates of peaks and throughs. The NBER dates do not coincide perfectly with the peaks and throughs of the first factor but close enough to justify referring to it as the "business cycle" factor. The same association of the first two estimated factors for Model II with the business cycle applies as we can observe in the right panel of Figure 1. The Factors 3 of both models appear to pick up the first and second oil crisis periods in the mid 1970's and the early 1980's. The turbulence of the interest rates in the early 1980's are present in the estimated Factor 5 of Model I and the estimated Factor 4 of Model II.

The actual estimates of $\Lambda$ are not easy to interpret and therefore Stock and Watson (2002b) propose to focus on the $R^2$ goodness-of-fit statistics which are obtained by regressing the univariate time series $y_{it}$ (for each $i = 1, \ldots, N$) on a constant and a particular principal component estimate (or diffusion index). The series of $N$ regressions can be repeated for each prinicipal component and the resulting $N$ dimensional series of $R^2$ statistics can be displayed as an index plot for each principal component. We present the $N$ series of $R^2$ statistics for the seven factors, estimated by KFS methods, of Model I in the left panel of Figure 2. The clustering of high $R^2$ statistics within one or more sectors is clearly visible. The first factor is highly correlated with the real variables in sector (A) real output & income and weakly correlated with the variables in the sectors (B) employment & hours and (H) orders. The second factor is mostly associated with the sectors (G) real inventories and (H) orders but also correlated with variables in the sectors (B) employment & hours and (F) housing starts & sales. The two individual indices for production and unemployment in sectors (A) and (B) are particularly highly correlated with Factor 2. The third and fifth factors are clearly connected with interest rates and spreads, respectively, from sector (K). The fourth factor does not contribute much to the analysis. Factors 6 and 7 can be interpreted as the price index and the housing market index, respectively. The $R^2$ statistics for the four factors in Model II are presented in the right panel of Figure 2. The third and fourth factors in

## Figure 1: Estimated factors

We present a selection of estimated factors extracted from the observed series by applying the KFS methods as described in Section 4.6. In the left panel the estimated Factors 1, 2, 3 and 5 for Model I are displayed and in the right panel all four estimated factors for Model II are presented.



Model II are strongly connected with interest rates and spreads, respectively, from sector (K). They are similar to Factors 3 and 5 of Model I. However, the first two factors of Model II are not very distinctive and can be regarded as a mix of the first two factors of Model I. In other words, they are associated with the "real" sectors (A) real output & income, (B) employment & hours, (F) housing starts & sales, and (H) orders.

### 6.3 Diagnostic checking

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 the practice of time series analysis, diagnostic test statistics are applied to standardized one-step ahead prediction errors. If the model is correctly specified these errors should be IID. We will not argue that a dynamic factor model is the appropriate specification for a joint analysis of 132 time series. However, the model misspecification diagnostics may indicate how far we are from a reasonable specification.

The Kalman filter allows us to compute the prediction errors for all 132 series in a few seconds. More specifically, we have computed the generalized least squares residuals as advocated by Harvey (1989, section 5.4) to allow for the intercept vector $\bar{\mu}$ in both model specifications I and II. 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 $\chi^2$ 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 3 for $q = 5$. The upper panel presents the $Q(q)$ statistics for the residuals of Model I while the lower panel presents those for Model II. 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 however that Model II is more successful in capturing the collective dynamics in the data-set compared to Model I.

## 7 Conclusions

We have presented a number of 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

FIGURE 2: $R^2$ STATISTICS FOR THE ESTIMATED FACTORS AGAINST EACH VARIABLE

We present two panels of $R^2$ statistics for each estimated factor against all $N = 132$ variables. The $R^2$ presented in the left-panel are those for the seven factors in Model I (with QML estimates for $\bar{\Lambda}$, $\Phi_1$ and $\Sigma_\epsilon$) and in the right-panel for the four factors in Model II (with QML estimates for $\bar{\Lambda}$, $\Phi_1$, $\Psi_1$ and $\Sigma_\epsilon$).

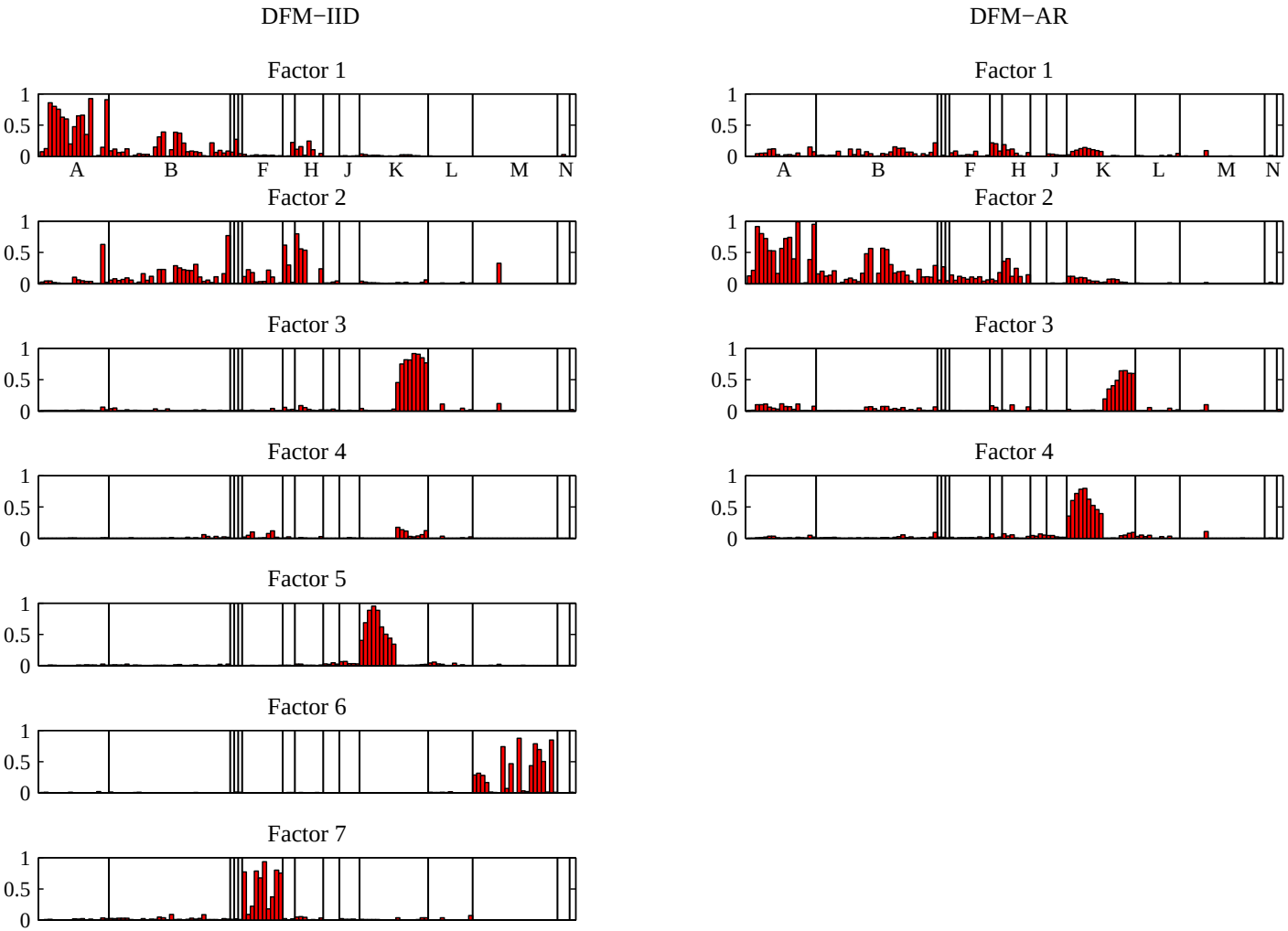
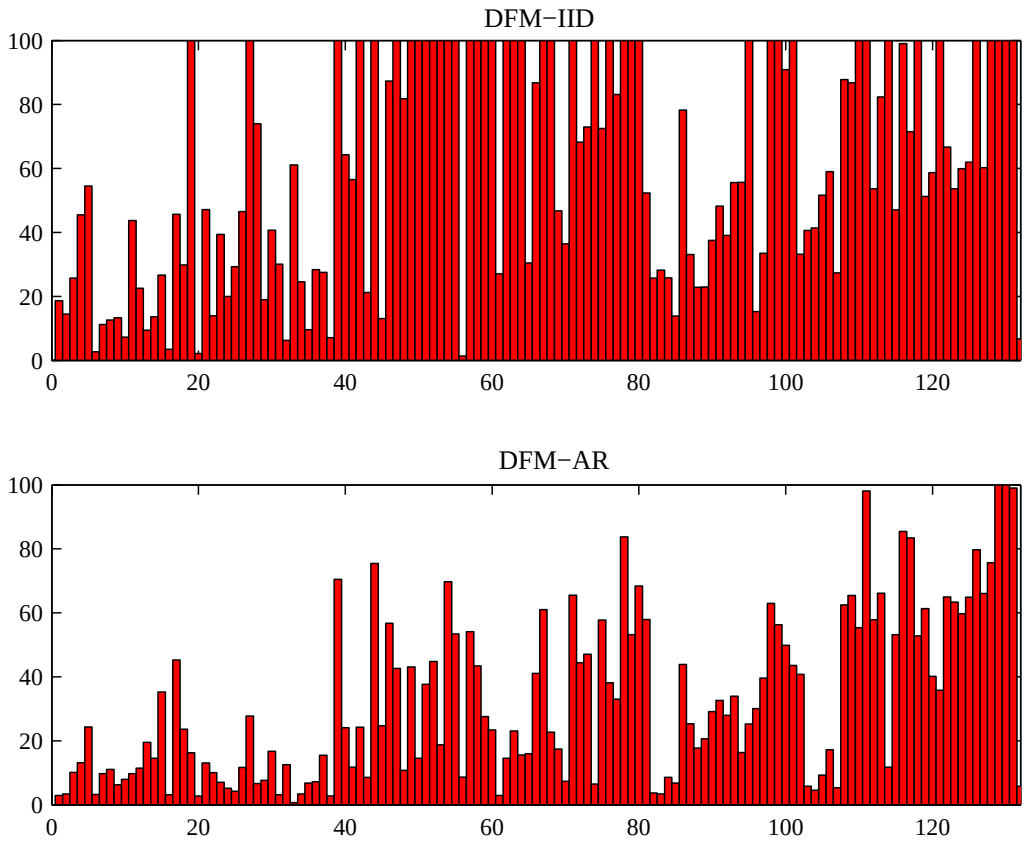


FIGURE 3: BOX-LJUNG STATISTICS

We present the Box-Ljung statistics $Q(5)$ for the generalized least squares residuals of the dynamic factor model specifications I and II.



substantial computational savings when estimating the factors and evaluating the Gaussian likelihood function. State space methods allow us to calculate model misspecification tests and diagnostics from the one-step ahead prediction errors. An important motivation for this research 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. The two steps can be integrated in a model-based dynamic factor analysis. Likelihood-based methods become a viable alternative to the principal component approach using the results of this paper. Future work must establish whether this approach produces 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.

# A Appendices

## A.1 Kalman filter and backward smoothing recursions

Consider the state space model (6)– (7) where initial state vector $\alpha_1$ has mean $\mathbb{E}(\alpha_1) = a$ and variance matrix $\mathrm{Var}(\alpha_1) = Q$. The Kalman filter for a given time series $y_t$ and a given parameter vector $\psi$ is given by

$$\begin{aligned} v_t &= y_t - d_t - \mu - X_t\beta - Za_{t|t-1}, & D_t &= ZQ_{t|t-1}Z' + \Sigma_\varepsilon, \\ & & K_t &= HQ_{t|t-1}Z'D_t^{-1}, \\ a_{t+1|t} &= Ha_{t|t-1} + K_t v_t, & Q_{t+1|t} &= HQ_{t|t-1}H' - K_t D_t K_t' + R\Sigma_\zeta R', \end{aligned} \tag{42}$$

for $t = 1, \ldots, T$, with initialisations $a_{1|0} = a$ and $Q_{1|0} = Q$, 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 $\alpha_t$ based on $y_1, \ldots, y_{t-1}$ is $a_{t|t-1}$, its mean squared error matrix is $Q_{t|t-1}$ and the *Kalman gain* matrix is $K_t$. Vector $a_{t+1|t}$ and matrix $Q_{t+1|t}$ 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 linear estimator (MMSLE) of $\alpha_t$ 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 $L_t = H - K_t Z$ and initializations $r_T = 0$ and $N_T = 0$. From these recursions, the MMSLE of the state vector using $y_1, \ldots, y_T$ is computed by

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

where $a_{t|T}$ is the MMSLE of $\alpha_t$ and $Q_{t|T}$ is its minimum mean squared error. Expressions for predictors of the state vector $\alpha_t$ and its mean squared error $Q_{t|s}$ 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 $\alpha_t$ and $\alpha_{t-1}$ given $y_1, \ldots, y_T$ and denoted by $Q_{t,t-1|T}$ is presented by de Jong and MacKinnon (1988) and given by

$$Q_{t,t-1|T} = (Q_{t|t-1} N_{t-1} - I) L'_{t-1} Q_{t-1|t-2}, \qquad t = 2, \ldots, T,$$

which can be evaluated using the earlier recursions.

Augmented KFS methods to calculate the marginal likelihood are developed in Ansley and Kohn (1985), de Jong (1991) and Koopman (1997). These algorithms require an additional set of recursions with the same dimension as the regression coefficient vector $(\mu', \beta')'$.

## A.2 Proof of Lemma 1

From Conditions (i), (ii) and (iii) in Section 4.1 and using the fact that $\Sigma_\varepsilon A^{L\prime}$ has full column rank, we obtain

$$\text{Col}\{\Sigma_\varepsilon A^{L\prime}\} = \text{Row}\{A^H\}^\perp = \text{Col}\{Z\}.$$

Define $\Lambda^\dagger = \Sigma_\varepsilon A^{L\prime}$ then $A^L = \Lambda^{\dagger\prime} \Sigma_\varepsilon^{-1}$ and $\text{Col}\{\Lambda^\dagger\} = \text{Col}\{Z\}$. Since $Z = \Lambda G$, with $G$ of full row rank, we have $\text{Col}\{\Lambda^\dagger\} = \text{Col}\{\Lambda\}$. This proves the necessity part of Lemma 1.

## A.3 Proof of Lemma 2

We have

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

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

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

as the GLS projection matrix for covariate matrix $A^{L\prime}$ and variance matrix $\Sigma_\varepsilon^{-1}$. 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

$$J^{H\prime} = \Sigma_\varepsilon A^{H\prime}(A^H \Sigma_\varepsilon A^{H\prime})^{-1} A^H = I - \Lambda^\dagger(\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1} \stackrel{def.}{=} M_\Lambda. \tag{43}$$

The proof of (23) is completed by the identity $J^H\Sigma_\varepsilon^{-1} = J^H\Sigma_\varepsilon^{-1}J^{H\prime}$ and the definition $e_t \stackrel{def.}{=} M_\Lambda(y_t - d_t - X_t\beta - \mu)$ as the GLS residual for data vector $y_t - d_t - X_t\beta - \mu$, covariate $\Lambda^\dagger$ and variance matrix $\Sigma_\varepsilon$.

## A.4 Proof of equation (31)

The following two lemmas are required for the main proof.

**Lemma 3.** *Consider the regression model* (9)*. The marginal Gaussian loglikelihood* $\ell_d(y;\theta)$ *is given by*

$$\ell_d(y;\theta) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\log|\tilde{X}'\Sigma^{-1}\tilde{X}| - \frac{1}{2}e^{x\prime}\Sigma^{-1}e^x, \tag{44}$$

*where* $e^x$ *is the residual vector from a GLS regression on* $y$ *with covariate matrix* $\tilde{X}$ *and variance matrix* $\Sigma$*.*

*Proof.* See *e.g.* Harville (1974). □

**Lemma 4.** *Consider the linear regression model*

$$y_1 = X_1\beta + \varepsilon_1, \qquad \varepsilon_1 \sim N(0, \Omega_1), \tag{45}$$

$$y_2 = X_2\beta + \varepsilon_2, \qquad \varepsilon_2 \sim N(0, \Omega_2), \tag{46}$$

*where* $\varepsilon_1$ *and* $\varepsilon_2$ *are independent, the GLS estimator* $\hat{\beta}_{GLS}$ *of* $\beta$ *is given by*

$$\hat{\beta}_{GLS} = \hat{\beta}^2_{GLS} + VX_1'F^{-1}(y_1 - X_1\hat{\beta}^2_{GLS}), \qquad Var(\hat{\beta}_{GLS}) = V - VX_1'F^{-1}X_1V,$$

*where* $F = X_1VX_1' + \Omega_1$*,* $\hat{\beta}^2_{GLS}$ *is the GLS estimator of* $\beta$ *based on* $y_2$ *only and* $V$ *is the associated variance*

$$\hat{\beta}^2_{GLS} = (X_2'\Omega_2^{-1}X_2)^{-1}X_2'\Omega_2^{-1}y_2, \qquad V = (X_2'\Omega_2^{-1}X_2)^{-1}.$$

*Furthermore,*

$$\begin{array}{c}(y_1 - X_1\hat{\beta}_{GLS})'\Omega_1^{-1}(y_1 - X_1\hat{\beta}_{GLS}) + (y_2 - X_2\hat{\beta}_{GLS})'\Omega_2^{-1}(y_2 - X_2\hat{\beta}_{GLS}) \; = \\ (y_2 - X_2\hat{\beta}^2_{GLS})'\Omega_2^{-1}(y_2 - X_2\hat{\beta}^2_{GLS}) + (y_1 - X_1\hat{\beta}^2_{GLS})'F^{-1}(y_1 - X_1\hat{\beta}^2_{GLS}).\end{array} \tag{47}$$

*Proof.* The results follow from regression theory. □

*Proof of equation* (31). It can be verified from (44) that

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

Denote by $\hat{\beta}$, $\hat{\mu}^L$ and $\hat{\mu}^H$ the GLS estimators of respectively $\beta$, $\mu^L$ and $\mu^H$ based on $y$ then

$$\ell_d(\bar{y}^L, \bar{y}^H;\theta) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\log|\tilde{X}_A'\Sigma_A^{-1}\tilde{X}_A| - \frac{1}{2}\log|\Sigma_A| - \frac{1}{2}RSS(\hat{\beta}, \hat{\mu}^L, \hat{\mu}^H), \tag{48}$$

where $\Sigma_A = (I_T \otimes A)\Sigma(I_T \otimes A')$, $\tilde{X}_A = (I_T \otimes A)\tilde{X}$, suppressing the dependence on $\theta$, and

$$RSS(\beta, \mu^L, \mu^H) = \hat{e}^{L\prime}\Sigma^{-1}_{\bar{y}^L}\hat{e}^L + \hat{e}^{H\prime}\Sigma^{-1}_{\bar{y}^H}\hat{e}^H, \tag{49}$$

with $\hat{e}^L = (\bar{y}^L - X^L\beta - E_1\mu^L)$ and $\hat{e}^H = (\bar{y}^H - X^H\beta - E_2\mu^H)$, where $E_1 = i_T \otimes I_m$, $E_2 = i_T \otimes I_{N-m}$, with $i_T = (1, \ldots, 1)'$, $X^L = (X_1^{L\prime}, \ldots, X_T^{L\prime})'$, $X^H = (X_1^{H\prime}, \ldots, X_T^{H\prime})'$ and

$$\Sigma_{\bar{y}^L} = \mathrm{Var}(\bar{y}^L), \qquad \Sigma_{\bar{y}^H} = \mathrm{Var}(\bar{y}^H).$$

Denote $M_{\perp E_2} = I - E_2(E_2'\Sigma^{-1}_{\bar{y}^H}E_2)^{-1}E_2'\Sigma^{-1}_{\bar{y}^H}$, $\bar{y}^H_{\perp E_2} = M_{\perp E_2}\bar{y}^H$ and $X^H_{\perp E_2} = M_{\perp E_2}X^H$, we have

$$RSS(\beta, \mu^L, \hat{\mu}^H) = \hat{e}^{L\prime}\Sigma^{-1}_{\bar{y}^L}\hat{e}^L + (\bar{y}^L_{\perp E_2} - X^H_{\perp E_2}\beta)'\Sigma^{-1}_{\bar{y}^H}(\bar{y}^L_{\perp E_2} - X^H_{\perp E_2}\beta). \tag{50}$$

Using the result of equation (43) we have

$$\begin{aligned} X^{H\prime}_{\perp E_2}\Sigma^{-1}_{y^H}X^H_{\perp E_2} &= \sum_t (X_t - \bar{X})'A^{H\prime}\left(A^H\Sigma A^{H\prime}\right)^{-1}A^H(X_t - \bar{X}) \\ &= \sum_t (X_t - \bar{X})'M_\Lambda'\Sigma^{-1}M_\Lambda(X_t - \bar{X}) \\ &= \sum_t \tilde{X}_t'\Sigma^{-1}\tilde{X}_t = B, \end{aligned} \tag{51}$$

and similarly

$$X^{H\prime}_{\perp E_2}\Sigma^{-1}_{y^H}\hat{y}^H = \sum_t \tilde{X}_t'\Sigma^{-1}\tilde{y}_t = Bb.$$

From (47) we have

$$RSS(\hat{\beta}, \mu^L, \hat{\mu}^H) = (\bar{y}^L - E_1\mu^L - X^L b)' F^{-1} (\bar{y}^L - E_1\mu^L - X^L b) + (\bar{y}^H_{\perp E_2} - X^H_{\perp E_2} b)' \Sigma^{-1}_{\bar{y}^H} (\bar{y}^H_{\perp E_2} - X^H_{\perp E_2} b), \tag{52}$$

where $F = X^L B^{-1} X^{L\prime} + \Sigma_{\bar{y}^L}$. Minimizing (52) with respect to $\mu^L$, we find

$$\hat{\mu}^L = (E_1' F^{-1} E_1)^{-1} E_1' F^{-1} (\bar{y}^L - X^L b). \tag{53}$$

Note that $\hat{\mu}^L$ is identical to the GLS estimator of $\mu^L$ from model (35). It follows from the definitions in section 4.5 and expression (44) that

$$\frac{NT}{2} \log 2\pi + C_3 + C_4 - \frac{1}{2} RSS(\hat{\beta}, \hat{\mu}^L, \hat{\mu}^H) = L_d(\bar{y}^L; \theta) + L_d(\bar{y}^H; \theta), \tag{54}$$

where

$$C_1 = -\frac{N-m}{2} \log T - \frac{1}{2} \log |B|,$$
$$C_2 = -\frac{1}{2} \log |X^L B^{-1} X^{L'} + \Sigma_{\bar{y}^L}| - \frac{1}{2} \log |E_1' (X^L B^{-1} X^{L'} + \Sigma_{\bar{y}^L})^{-1} E_1|.$$

Using a well known determinant identity we have

$$|X^L B^{-1} X^{L'} + \Sigma_{\bar{y}^L}| = |B + X^{L'} \Sigma_{\bar{y}^L} X^L| |B^{-1}| |\Sigma_{\bar{y}^L}|, \tag{55}$$

and with the Woodbury matrix identity and results on determinants of block matrices, inverses of a block matrices and (43) we have

$$|E_1' (X^L B^{-1} X^{L\prime} + \Sigma_{\bar{y}^L})^{-1} E_1| = |E_1' \Sigma^{-1}_{\bar{y}^L} E_1 - E_1' \Sigma^{-1}_{\bar{y}^L} X^L (B + X^{L\prime} \Sigma^{-1}_{\bar{y}^L} X^L)^{-1} X^{L\prime} \Sigma^{-1}_{\bar{y}^L} E_1| \tag{56}$$

$$= \frac{1}{|D|} \begin{vmatrix} E_1' \Sigma^{-1}_{\bar{y}^L} E_1 & 0 & E_1' \Sigma^{-1}_{\bar{y}^L} X^L \\ 0 & T\Sigma_H^{-1} & \sum_t \Sigma_H^{-1} X_t^H \\ X^{L\prime} \Sigma^{-1}_{\bar{y}^L} E_1 & \sum_t X_t^{H\prime} \Sigma_H^{-1} & X^{L\prime} \Sigma_{\bar{y}^L} X^L + \sum_t X_t M_\Lambda' \Sigma_\varepsilon^{-1} M_\Lambda X_t \end{vmatrix}$$

$$= \frac{|\tilde{X}_A' \Sigma_A^{-1} \tilde{X}_A|}{|D|}, \tag{57}$$

where

$$|D| = \begin{vmatrix} T\Sigma_H^{-1} & \sum_t \Sigma_H^{-1} X_t^H \\ \sum_t X_t^{H\prime} \Sigma_H^{-1} & X^{L\prime} \Sigma_{\bar{y}^L} X^L + \sum_t X_t M_\Lambda' \Sigma_\varepsilon^{-1} M_\Lambda X_t \end{vmatrix} = |B + X^{L\prime} \Sigma_{\bar{y}^L} X^L| T^{N-m}.$$

Finally, we have

$$|\Sigma_{\bar{y}^L}| = |\Sigma_{\bar{y}^L}||\Sigma_{\bar{y}^H}| = |\Sigma_A|, \tag{58}$$

since $|\Sigma_{\bar{y}^H}| = 1$ and $\Sigma_A$ is block diagonal. Combining (48) and (54) with (55), (57) and (58) we obtain

$$L_d(\bar{y}^L;\theta) + L_d(\bar{y}^H;\theta) = \ell_d(\bar{y}^L, \bar{y}^H;\theta),$$

which concludes the proof.

□

## A.5 Proof of results in Section 4.6

We have proved in Appendix A.3 that the MMSLE of $\mu^L$ from model (35) equals the GLS estimator of $\mu^L$, see equation (53). Then,

$$\begin{aligned} P(\delta|\bar{y}^L) &= \mathbb{E}(\delta) + \mathrm{Cov}(\delta, \bar{y}^L)\mathrm{Var}(\bar{y}^L)^{-1}(y^L - E_1\mu_{GLS} - X^L\mathbb{E}(\delta)) \\ &= b + B^{-1}X_L'(X^L B^{-1} X^L + \Sigma_{\bar{y}^L})^{-1}(y^L - E_1\mu_{GLS} - X^L b) \end{aligned}$$

where $E_1$ is defined below equation (49) and the variances and covariances are evaluated assuming (35) is the true model. Consider the function $\mathrm{RSS}(\beta, \mu^L, \hat{\mu}^H)$ defined in (50). The GLS estimators of $\hat{\beta}$ and $\hat{\mu}^L$ can be obtained by minimizing $\mathrm{RSS}(\beta, \mu^L, \hat{\mu}^H)$ with respect to $\beta$ and $\mu^L$. Substituting $\hat{\mu}^L$ in $\mathrm{RSS}(\beta, \mu^L, \hat{\mu}^H)$ and minimizing with respect to $\beta$, equation (36) follows from Lemma 4.

Finally, we prove that the augmented Kalman filter produces the correct variances and covariances for the GLS estimators $\hat{\beta}$ and $\hat{\mu}^L$. From (50) it follows that the GLS estimator of $\gamma_L = (\mu^{L\prime}, \beta')$ is given by $C^{-1}c$ where

$$C = \begin{pmatrix} E_1'\Sigma_{y^L}^{-1}E_1 & E_1'\Sigma_{y^L}^{-1}X^L \\ X^{L\prime}\Sigma_{y^L}^{-1}E_1 & X^{L\prime}\Sigma_{y^L}^{-1}X^L + B \end{pmatrix}, \quad c = \begin{pmatrix} E_1'\Sigma_{y^L}^{-1}\bar{y}^L \\ X^{L\prime}\Sigma_{y^L}^{-1}\bar{y}^L + Bb. \end{pmatrix},$$

Since the correct variance of the estimator is given by $C^{-1}$ we need to prove that

$$\mathbb{E}\left[(\hat{\gamma}_L - \gamma_L)(\hat{\gamma}_L - \gamma_L)'\right] = C^{-1}, \tag{59}$$

where $\hat{\gamma}_L = [P(\mu^L|\bar{y}^L)', P(\delta|\bar{y}^L)']'$ and the expectation is computed under the assumption

that (35) is the true model. Since we already showed that $\hat{\gamma}_L = C^{-1}c$, we have

$$\mathbb{E}((\hat{\gamma}_L - \gamma_L)(\hat{\gamma}_L - \gamma_L)') = C^{-1}\mathbb{E}\left[(c - C\gamma_L)(c - C\gamma_L)'\right]C^{-1}.$$

It can now be shown that $\mathbb{E}\left[(c - C\gamma_L)(c - C\gamma_L)'\right] = C$, which proves (59).

## A.6 The score function of Section 5.1

The derivatives of (39) with respect to the system matrices $Z$, $H$, $\Sigma_\varepsilon$ and $\Sigma_\zeta$ of the linear Gaussian state space model (7) – (6), with $\mu = 0$, $\beta = 0$ and $R = I$, are given by

$$\frac{\partial \ell(y)}{\partial Z} = \Sigma_\varepsilon^{-1}(\sum_{t=1}^{T}\{y_t - d_t\}a'_{t|T} - ZS^{(0)}_{1:T}), \qquad \frac{\partial \ell(y)}{\partial \Sigma_\varepsilon} = Q^*_\varepsilon\Sigma_\varepsilon^{-1} - \frac{1}{2}\mathrm{diag}(Q^*_\varepsilon\Sigma_\varepsilon^{-1}),$$
$$\frac{\partial \ell(y)}{\partial H} = \Sigma_\zeta^{-1}(S^{(1)}_{2:T} - HS^{(0)}_{1:T-1}), \qquad \frac{\partial \ell(y)}{\partial \Sigma_\zeta} = Q^*_\zeta\Sigma_\zeta^{-1} - \frac{1}{2}\mathrm{diag}(Q^*_\zeta\Sigma_\zeta^{-1}),$$

where $Q^*_\varepsilon = Q_\varepsilon - T$, $Q^*_\zeta = Q_\zeta - T - 1$, with $Q_\varepsilon$ and $Q_\zeta$ defined in (40),

$$S^{(0)}_{j:k} = \sum_{t=j}^{k} a_{t|T}a'_{t|T} + Q_{t|T}, \qquad S^{(1)}_{j:k} = \sum_{t=j}^{k} a_{t|T}a'_{t-1|T} + Q_{t,t-1|T}, \tag{60}$$

for $j, k = 1, \ldots, T$ $(j \leq k)$, where $a_{t|T}$, $Q_{t|T}$ and $Q_{t-1,t|T} = Q'_{t,t-1|T}$ are evaluated by the KFS methods of Appendix A.1. Matrices $Q_\varepsilon$ and $Q_\zeta$ depend on the smoothed disturbances $\hat{\varepsilon}_t = y_t - d_t - Za_{t|T}$ and $\hat{\zeta}_t = a_{t|T} - Ha_{t-1|T}$ together with their variances which depend on $Q_{t|T}$, $Q_{t,t-1|T}$ and the system matrices. The derivatives given above are evaluated at $\psi = \psi^*$. The system matrices are functions of coefficient vector $\psi$. 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 5.2

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 $\Psi_1$. 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

$$
\begin{aligned}
Z_i^+ &= -\sum_{t=2}^{T}(y_{it} - \Psi_{ii}y_{i,t-1})(a_{t|T} - \Psi_{ii}a_{t-1|T})'[\Psi_{ii}^2 S_{1:T-1}^{(0)} - \Psi_{ii}S_{2:T}^{(1)} - \Psi_{ii}S_{2:T}^{(1)\prime} + S_{2:T}^{(0)}]^{-1}, \\
H^+ &= S_{2:T}^{(1)} S_{2:T}^{(0)\,-1}, \\
\Psi_{ii}^+ &= \sum_{t=2}^{T} Z_i P_{t,t-1|T} Z_i' - \hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-1} \;/\; (Z_i S_{1:T-1}^{(0)} Z_i' + \sum_{t=2}^{T} y_{it}\{\hat{\varepsilon}_{it} - Z_i a_{t|T}\}),
\end{aligned}
\tag{61}
$$

where $Z_i$ is the $i$th row of $Z$ and $\Psi_{ii}$ is the $i$th diagonal element of $\Psi_1$ for $i = 1, \ldots, N$. Expressions for $\Sigma_\varepsilon^+$ and $\Sigma_\zeta^+$ are obtained as solutions of $Q_\varepsilon = 0$ and $Q_\zeta = 0$, respectively. The system matrices are evaluated at $\psi = \psi^*$. The new coefficients for $\lambda_i$ are distilled from $Z_i^+$ for $i = 1, \ldots, N$ while new coefficients for $\Phi_i$ $(i = 1, \ldots, q_\Phi)$ are distilled from $H^+$.

The equations for $Z_i^+$ and $\Psi_{ii}^+$ in (61) can not be solved separately. Keeping $S_{jk}^{(0)}$, $S_{jk}^{(1)}$, $a_{t|T}$, $Q_{t,t-1|T}$ and $\hat{\varepsilon}_{it}$ $(i = 1, \ldots, N,\ t = 2, \ldots, T)$ 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


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 factors in approximating factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *J. Business and Economic Statist. 25*, 52–60.

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika 70*, 343–65.

Bellhouse, D. (1990). On the equivalence of marginal and approximate conditional likelihoods for correlation parameters under a normal model. *Biometrika 77*, 743–746.

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.

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

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.

Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Royal Statistical Society B 49*, 1–39.

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.

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.

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.

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.

Hannan, E. J., W. T. M. Dunsmuir, and M. Deistler (1980). Estimation of vector ARMAX models. *Journal of Multivariate Analysis 10*(3), 275–295.

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

Harville, D. A. (1974). Bayesian inference for variance components using only error contrast. *Biometrika 61*, 383–85.

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.

Lawley, D. N. and A. E. Maxwell (1971). *Factor analysis as a statitical method*. London: Butterworths. 2nd edition.

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.

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.

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.