Likelihood-based Analysis for Dynamic Factor Models

Borus Jungbacker
Siem Jan Koopman*

Faculty of Economics and Business Administration, VU University Amsterdam.

* Tinbergen Institute Amsterdam, the Netherlands.

Likelihood-based Dynamic Factor Analysis for Measurement and Forecasting

Borus Jungbacker[†] and Siem Jan Koopman[†‡]

[†] *Department of Econometrics, Vrije Universiteit Amsterdam,*
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[‡] *Tinbergen Institute, Amsterdam, The Netherlands*
CREATES, Aarhus University, Denmark
E-mail: `s.j.koopman@vu.nl`

Summary

We present new results for the likelihood-based analysis of the dynamic factor model. The latent factors are modeled by linear dynamic stochastic processes. The idiosyncratic disturbance series are specified as autoregressive processes with mutually correlated innovations. The new results lead to computationally efficient procedures for the estimation of the factors and for the parameter estimation by maximum likelihood methods. We also present the implications of our results for models with regression effects, for Bayesian analysis, for signal extraction, and for forecasting. An empirical illustration is provided for the analysis of a large panel of macroeconomic time series.

Keywords: *EM algorithm; Kalman Filter; Latent Factors; State Space Form; Maximum Likelihood.*

1. INTRODUCTION

The basic dynamic factor model is given by

$$y_t = \Lambda f_t + u_t, \qquad t = 1, \ldots, n, \tag{1.1}$$

where y_t denotes the observed $N \times 1$ vector of time series at time t, Λ is an $N \times r$ loading matrix, f_t is an $r \times 1$ vector of common latent factors an u_t is the $N \times 1$ idiosyncratic vector. The factors in f_t are modeled by linear dynamic processes while the idiosyncratic components in u_t are modeled by linear autoregressive processes. 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 in u_t and the common factors in f_t are Gaussian we can evaluate the likelihood function efficiently by the Kalman filter. When disturbances in the model are non-Gaussian, we treat the Gaussian likelihood as a quasi-likelihood. The Gaussian likelihood function can be numerically maximized to obtain maximum likelihood or quasi-maximum likelihood 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 et al. (1977) to find the optimum of the likelihood, see also 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 develop the results for the state space formulation of the dynamic factor model (1.1) and its various extensions leading to the generalized dynamic factor model. We consider both signal extraction and likelihood evaluation. Finding the maximum of a likelihood function is not straightforward if many parameters are present. We show that the new results also lead to an effective implementation of numerical optimization methods.

The key insight of this paper is that the observed time series can be split into a low-dimensional vector series and a high-dimensional vector series. For the measurement 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 in a likelihood-based analysis. A particular feature of a likelihood-based analysis is that the dynamic factors are explicitly modeled. It facilitates the estimation of the parameters of interest by accounting for the model specification. The dynamic factors can typically represent aspects of economic theory. As a result, hypothesis tests can be formulated and tested by statistical means. The resulting methods from the developments presented below allow the real-time estimation of the factors, the estimation of the past factors as well as the prediction of the factors and observations. The Kalman filter further produces mean squared errors of the factor estimates without an extra computational effort. When the data generating process can be represented by a Gaussian dynamic factor model, the standard results imply that parameter estimates are asymptotically efficient. Moreover, Doz et al. (2012) show, under mild conditions, that the factor estimates from the (quasi-)maximum likelihood procedure are consistent for the true factors when $n \to \infty$ and $N \to \infty$. Their estimation procedure is shown to be robust to misspecification of cross-sectional and time series correlation of the idiosyncratic components. They also present Monte Carlo evidence that such factor estimates can be more precise than principal component estimates.

The increasing availability of high-dimensional data sets in economics and finance has motivated much 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 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 et al. (2000) propose a frequency-domain estimation procedure that provides consistent estimates of the factors for a general class of dynamic factor models. A two-step method that partly relies on the Kalman filter applied to the dynamic factor model (1.1) is proposed by Doz et al. (2011).

We adopt an econometric analysis for the dynamic factor model as given in (1.1) from a classical perspective. However, our results are applicable and relevant more generally. Firstly, we will indicate that Bayesian approaches to dynamic factor models as they are pursued by Aguilar and West (2000) and Fiorentini et al. (2004) can benefit from the presented new results. Secondly, our focus is on dynamic factor analysis but related approaches such as the Bayesian vector autoregressions (BVAR) have similar aims including macroeconomic forecasting; see the review by Koop and Korobilis (2010). The application of the method of shrinkage in a large-dimensional BVAR analysis has shown

to produce comparable results in forecast precision as for a dynamic factor analysis; see Banbura et al. (2010). The presented results below can be used jointly with a shrinkage method in a BVAR analysis. An alternative use of the low-dimensional vector series is explored by Bräuning and Koopman (2014) and leads to the incorporation of principal components in a low-dimensional dynamic factor model. Our results are explored further for non-Gaussian and nonlinear extensions by Mesters and Koopman (2014).

The remainder of the paper is organized as follows. Section 2 presents the key results for a general state space analysis including parameter estimation, signal extraction and forecasting. Section 3 discusses the implications of the new results for a general class of dynamic factor models, possibly with autoregressive disturbances and with regression effects. Furthermore we briefly discuss the implications of our results for simulation methods in classical and Bayesian treatments. Section 4 presents an empirical illustration for a large panel of US macroeconomic time series. Section 5 concludes. Proofs and derivations are given in the Appendix.

2. MAIN RESULTS FOR STATE SPACE MODEL

The state space representation of a linear time series model is given by

$$y_t = Z\alpha_t + \varepsilon_t, \qquad \alpha_{t+1} = T\alpha_t + R\eta_t, \tag{2.2}$$

for $t = 1, \ldots, n$, where y_t is the $N \times 1$ vector of dependent variables at time t, α_t is the $p \times 1$ state vector, ε_t and η_t are disturbance vectors and Z, T and R are fixed matrices of appropriate dimensions. The two equations in (2.2) are referred to as the observation and state equation, respectively. The disturbance vectors ε_t and η_t are mutually and serially uncorrelated with zero means and variance matrices Σ_ε and Σ_η, respectively. The system matrices Z, T, and R, together with Σ_ε and Σ_η, are fixed and may (partly) depend on the parameter vector ψ. All results in this paper apply when the system matrices are known functions of the time index t and other covariates. Many linear time series models can be represented by (2.2); see, for example, Durbin and Koopman (2012, Ch. 3).

The key tool for a time series state space analysis is the Kalman filter. It produces the minimum mean square linear estimator (MMSLE) of the state vector α_t in (2.2) conditional on the set of past observations $y_1, \ldots, y_{t-1}$, together with its mean square error (MSE) matrix. We denote the MMSLE and MSE of the state α_t, conditional on observations and as a function of parameter vector ψ, by

$$a_{t|s} = \mathbb{E}(\alpha_t | y_1, \ldots, y_s; \psi), \qquad P_{t|s} = \text{Var}(a_{t|s} - \alpha_t | y_1, \ldots, y_s; \psi), \tag{2.3}$$

for $t, s = 1, \ldots, n$. The Kalman filter is a recursive procedure through the time-index t and is given by

$$
\begin{aligned}
v_t &= y_t - Za_{t|t-1}, \\
F_t &= ZP_{t|t-1}Z' + \Sigma_\varepsilon, \\
K_t &= TP_{t|t-1}Z', \\
a_{t+1|t} &= Ta_{t|t-1} + K_t F_t^{-1} v_t, \\
P_{t+1|t} &= TP_{t|t-1}T' - K_t F_t^{-1} K_t' + \Sigma_\eta,
\end{aligned}
\tag{2.4}
$$

for $t = 1, \ldots, n$. The derivation of recursion (2.4) is given by Anderson and Moore (1979, Chapter 2). The initial conditions $a_{1|0}$ and $P_{1|0}$ can be obtained analytically when the state vector is a stationary process. The MMSLE $a_{t|n}$, that is conditional on all available observations $y_1, \ldots, y_n$, together with its MSE matrix $P_{t|n}$, can be evaluated by a backwards recursion which we will refer to as a state smoothing algorithm. In our

study we have adopted the sequential Kalman filter implementation of Koopman and Durbin (2000) in which each element of y_t is updated separately. The Kalman filter plays a central role in smoothing as well as in the prediction error decomposition of a time series, in the treatment of missing observations and in forecasting; see, for example, Durbin and Koopman (2012, Ch. 4).

2.1. Model with high-dimensional observation vector

We treat the dynamic factor model as the state space model (2.2) with $N >> p$. A large panel of individual time series is represented by y_t and can be described by a small set of dynamic factors which are placed in the state vector α_t. The case of $N >> p$ with large N does not prohibit the use of the Kalman filter in a dynamic factor analysis as is illustrated by Engle and Watson (1981) and Quah and Sargent (1993). However, more recent applications of dynamic factor models have N as large as 250 and even larger. Such dimensions will lead to computational problems for the Kalman filter (2.4). In particular, the matrix F_t must be inverted and even with increasing computer power, the inversion of a high-dimensional matrix requires substantial computing time and compromises numerical precision. We notice that the inverse of F_t is required for $t = 1, \ldots, n$.

The direct computation of the inverse of F_t in (2.4) may be avoided via the exploitation of specific inversion lemmas. These inversion lemmas do not necessary lead to computational gains because the lemma expression and remaining Kalman filter expressions remain high-dimensional. In addition, the inversion lemmas cannot be applied generally as they require the inverse of $P_{t|t-1}$ in the Kalman filter (2.4) to exist. Ansley and Kohn (1985) have argued that various linear time series processes for the state vector lead to a singular matrix $P_{t|t-1}$. This problem occurs when the state equation in (2.2) represents any autoregressive moving average process. We therefore cannot rely on inversion lemmas to reduce the computational burden for the Kalman filter when N is very large. Instead we propose a new and general method based on transforming the observation vector.

2.2. Transformation of observation equation

Define $y_t^* = A y_t$ for any non-singular $N \times N$ matrix A and with $t = 1, \ldots, n$. The MMSLE $a_{t|s}$ as defined in (2.3) is not affected when the set $y_1, \ldots, y_s$ is replaced by $y_1^*, \ldots y_s^*$. We will show that for certain choices of A the MMSLE can be computed more efficiently based on $y_1^*, \ldots, y_n^*$ rather than $y_1, \ldots, y_n$. The results also apply to other computations that are related to the Kalman filter.

Suppose we partition matrix A and vector y_t^* as

$$A = \left[\begin{array}{c} A^L \\ A^H \end{array} \right], \qquad y_t^* = \left(\begin{array}{c} y_t^L \\ y_t^H \end{array} \right), \tag{2.5}$$

where

$$y_t^L = A^L y_t, \qquad y_t^H = A^H y_t,$$

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, where $0 < m \le p$ is the rank of matrix Z. We aim to choose matrix A such that y_t^L and y_t^H are not correlated with each other and that only y_t^L depends on α_t. More specifically, the model for y_t^* will be

of the form

$$y_t^L = A^L Z \alpha_t + e_t^L, \qquad y_t^H = e_t^H, \tag{2.6}$$

where $e_t^L = A^L \varepsilon_t$ and $e_t^H = A^H \varepsilon_t$. We have

$$\mathbb{E}(e_t^L) = 0, \qquad \mathbb{E}(e_t^H) = 0, \qquad \mathrm{Var}(e_t^L) = \Sigma^L, \qquad \mathrm{Var}(e_t^H) = \Sigma^H, \qquad \mathbb{E}(e_t^H e_t^{L\,\prime}) = 0,$$

for $t = 1, \ldots, n$, where $\Sigma^L = A^L \Sigma_\varepsilon A^{L\,\prime}$ and $\Sigma^H = A^H \Sigma_\varepsilon A^{H\,\prime}$.

Conditions   A suitable matrix A needs to fulfill the following conditions:

1. A is full rank,
2. $A^H \Sigma_\varepsilon A^{L\,\prime} = 0$,
3. $\mathrm{Row}\{A^H\} = \mathrm{Col}\{Z\}^\perp$,

where $\mathrm{Col}\{X\}$ and $\mathrm{Row}\{X\}$ denote the row and column spaces of any 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 (2.6) are uncorrelated and condition (iii) implies that the second equation in (2.6) does not depend on α_t. Condition (iii) is stronger than strictly necessary. The transformed model will still be of the form (2.6) 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 compromising the special form of (2.6). 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 Σ_ε since

$$|A|^2 = |\Sigma_\varepsilon|^{-1} |A \Sigma_\varepsilon A'| = |\Sigma_\varepsilon|^{-1} |A^L \Sigma_\varepsilon A^{L\,\prime}||A^H \Sigma_\varepsilon A^{H\,\prime}| = |\Sigma_\varepsilon|^{-1} |\Sigma^L|. \tag{2.7}$$

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

LEMMA 2.1. *Consider the state space representation* (2.2) *and partition matrix A as in* (2.5). *Suppose matrix A^H satisfies (iii), then matrix A satisfies (i)–(iii) if and only if*

$$A^L = Z^{\dagger\,\prime} \Sigma_\varepsilon^{-1}, \tag{2.8}$$

where the columns of the $N \times m$ matrix $Z^\dagger$ form a basis for the column space of Z.

Remarks

(a) It is easily verified that any matrix A with A^L given by (2.8) and A^H satisfying (iii), fulfills conditions (i)–(iii). We prove the necessity part of the lemma in Appendix A.1.

(b) In case the state space form (2.2) has a Z matrix that is not of full column rank, a suitable matrix $Z^\dagger$ can be obtained from the decomposition

$$Z = Z^\dagger C, \tag{2.9}$$

for any full rank $m \times p$ matrix C with $m \le p$. In many cases of interest, however,

matrix Z is of full column rank and $p = m$. We then have

$$Z^\dagger = ZC^{-1}, \tag{2.10}$$

for any non-singular $p \times p$ matrix C. Typical choices in the latter case are $C = I$ and $C = Z'\Sigma_\varepsilon^{-1}Z$.

(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 $Z^{\dagger\,\prime}$. 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 (2.5) and model (2.6). Although our results are more general and are developed for different purposes, a similar transformation as (2.5) for a different class of factor models is considered by Fiorentini et al. (2004, section 2.4.1).

2.3. Signal extraction

By considering a matrix A that satisfies the conditions (i)–(iv) in Section 2.2, we are able to efficiently compute the estimates of the state vector in (2.3), that is the MMSLE $a_{t|s}$ and its MSE $P_{t|s}$ for $t, s = 1, \ldots, n$ with typically $s \geq t$. Since matrix A has full rank, we can replace $y_1, \ldots, y_n$ by $y_1^*, \ldots, y_n^*$ in (2.3). Furthermore, from (2.6) it follows that y_t^L and y_t^H are uncorrelated and that y_t^H does not depend on α_t. Hence, we can replace $y_1, \ldots, y_n$ by $y_1^L, \ldots, y_n^L$ in (2.3). We therefore can evaluate $a_{t|s}$ and $P_{t|s}$ in (2.3) by applying the Kalman filter and related methods to the low-dimensional model

$$y_t^L = A^L Z \alpha_t + e_t^L, \qquad \mathbb{E}(e_t^L) = 0, \quad \mathrm{Var}(e_t^L) = \Sigma^L, \tag{2.11}$$

for $t = 1, \ldots, n$. The high-dimensional matrix A^H and vector y_t^H are not required for the estimation of α_t.

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 different t with $t = 1, \ldots, n$. This solution is also adopted in cases where the system matrices of the state space form (2.2) vary over time.

2.4. Forecasting observations

Forecasting in the state space framework is straightforward. We continue with the Kalman filter updating equations (2.4) after time $t = n$, and treat the observations y_{n+j} as missing so that v_{n+j} is missing with $F_{n+j} \to \infty$ in (2.4) for $j = 1, 2, \ldots$; see, for example, Durbin and Koopman (2012, Ch. 4). The Kalman filter equations reduce to

$$a_{n+j|n} = T a_{n+j-1|n}, \qquad P_{n+j|n} = T P_{n+j-1|n} T' + \Sigma_\eta, \qquad j = 2, 3, \ldots, \tag{2.12}$$

where $a_{n+1|n}$ and $P_{n+1|n}$ are obtained from (2.4). The MMSLE of the observation y_{n+j} and its MSE, conditional on all available observations $y_1, \ldots, y_n$, are denoted by $\hat{y}_{n+j}$ and $\mathrm{MSE}(\hat{y}_{n+j})$, respectively, and given by

$$\hat{y}_{n+j} = Z a_{n+j|n}, \qquad \mathrm{MSE}(\hat{y}_{n+j}) = Z P_{n+j|n} Z' + \Sigma_\varepsilon, \qquad j = 1, 2, \ldots. \tag{2.13}$$

By applying the Kalman filter to the low-dimensional model (2.11), we can still adopt (2.12) to obtain the state forecasts; the same arguments as those used for signal extraction

in Section 2.3 applies. Although we have computed the state MMSLE $a_{n+1|n}$ using the low-dimensional observations $y_1^L, \ldots, y_n^L$, we can still compute the observation forecasts as in (2.13).

2.5. Loglikelihood evaluation

Define the data vector $y = (y_1', \ldots, y_n')'$. The loglikelihood function $\ell(y; \psi)$ for the Gaussian density can be evaluated via the prediction error decomposition and is given by

$$\ell(y; \psi) = -\frac{Nn}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \log |F_t| - \frac{1}{2} \sum_{t=1}^{n} v_t' F_t^{-1} v_t, \tag{2.14}$$

where prediction error v_t and its MSE matrix F_t are evaluated by the Kalman filter (2.4); see Schweppe (1965) and Harvey (1989, section 3.4). In case $N >> p$, it is computationally more efficient to evaluate (2.14) by choosing a matrix A that satisfies the conditions (i)–(iv) in Section 2.2, by transforming y_t as in (2.5) and by considering model (2.6). We then have

$$\ell(y; \psi) = \ell(y^L; \psi) + \ell(y^H; \psi) + n \log |A|, \tag{2.15}$$

where $y^L = (y_1^{L\,\prime}, \ldots, y_n^{L\,\prime})'$ and $y^H = (y_1^{H\,\prime}, \ldots, y_n^{H\,\prime})'$. The first term $\ell(y^L; \psi)$ in (2.15) is evaluated by the Kalman filter applied to low-dimensional model (2.11). The second term is

$$\ell(y^H; \psi) = -\frac{(N-m)n}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} y_t^{H\,\prime} (\Sigma^H)^{-1} y_t^H, \tag{2.16}$$

since $|\Sigma^H| = 1$ from condition (iv) in Section 2.2. Lemma 2.2 shows that the last term in equation (2.16) can be calculated without the construction of matrix A^H. The proof is given in Appendix A.2.

LEMMA 2.2. *For the state space representation* (2.2), *transformation* (2.5) *and resulting model* (2.6), *with A^L given by* (2.8), *we have the identity*

$$y_t^{H\,\prime} (\Sigma^H)^{-1} y_t^H = e_t' \Sigma_\varepsilon^{-1} e_t, \tag{2.17}$$

where $e_t = \left[I_N - Z^\dagger \left(Z^{\dagger\,\prime} \Sigma_\varepsilon^{-1} Z^\dagger \right)^{-1} Z^{\dagger\,\prime} \Sigma_\varepsilon^{-1} \right] y_t$ *is the generalized least squares (GLS) residual vector for data vector y_t, covariate Z and variance matrix Σ_ε. The definition of e_t is valid for any decomposition of $Z^\dagger$ in* (2.9).

Given the expression for $|A|^2$ in (2.7), the loglikelihood function (2.15) becomes

$$\ell(y; \psi) = c + \ell(y^L; \psi) - \frac{n}{2} \log \frac{|\Sigma_\varepsilon|}{|\Sigma_L|} - \frac{1}{2} \sum_{t=1}^{n} e_t' \Sigma_\varepsilon^{-1} e_t, \tag{2.18}$$

where c is a constant independent of both y and ψ. It follows that for the evaluation of the loglikelihood, computation of matrix A^H and vectors y_t^H, for $t = 1, \ldots, n$, is not required. Expression (2.18) is instrumental for a computationally feasible approach to the likelihood based analysis of the dynamic factor model.

2.6. Maximizing the loglikelihood function

The estimation of parameter vector ψ is based on maximizing the loglikelihood function $\ell(y; \psi)$ in (2.14) with respect to ψ. The number of parameters is typically very high, say 500 or 1,000. Section 2.5 has shown that the loglikelihood function can be evaluated efficiently for state space models with high-dimensional observation vectors. Numerical optimization procedures are used for the maximization of the loglikelihood function; for example, the quasi-Newton BFGS algorithm as described in Nocedal and Wright (1999). Many of such methods require the evaluation of the score vector. Since the number of parameters is high, evaluating the score vector numerically is infeasible in many cases, even if the results of Section 2 are used. In Appendix A.3 we show that the exact score function for the system and variance matrices in (2.2) can be obtained by a single Kalman filter and smoothing algorithm applied to the low-dimensional model (2.11). The exact score function is here key to the feasibility of maximum likelihood estimation of ψ. As an alternative, the expectation-maximization (EM) algorithm can be adopted to obtain the maximum likelihood estimates. We show in Appendix A.4 that for each EM step a single Kalman filter and smoothing algorithm based on the low-dimensional model (2.11) is required only.

2.7. Computational gains

We present possible gains in computing times that are achieved by our new methods based on the transformed observations $y_1^*, \ldots, y_n^*$ with $y_t^* = Ay_t$ for $t = 1, \ldots, n$. The gains are relative to the standard application of the Kalman filter based on $y_1, \ldots, y_n$. Also, computational gains depend primarily on the panel dimension N and state vector dimension p. We assume that matrix Z in (2.2) is full rank and $p = m$ in Lemma 2.1 To obtain some insights in the size of the gains, we calculate the loglikelihood function (2.14) for different values of N and p. The calculations are performed using the Kalman filter applied to the untransformed data and using the methods described in section 2.5. In all cases we have used the sequential version of the Kalman filter that updates each element of y_t (or y_t^L) separately as described in Koopman and Durbin (2000).

In 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 devices. 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.

3. FURTHER APPLICATIONS

The results in Section 2 are particularly useful for a likelihood-based dynamic factor analysis as is illustrated in Section 4. In this section we discuss a selection of further applications. The presence of a mean vector in model (2.2) is discussed and a method for its estimation together with the factors is presented. Further we show that our results are also applicable to state space models with regression effects. Finally, we indicate how the method can be used in a Bayesian treatment of the dynamic factor model.

Table 1. Computational Gains

We present the gains in computing time when evaluating the loglikelihood function for a state space model (2.2). The ratio d_1/d_2 is reported: d_1 is the CPU time for the Kalman filter, applied to $y_1, \ldots, y_n$, and d_2 is CPU time for the new devices of Section 2.5. The ratios are reported for different observation and state dimensions N and p, respectively.

$N \backslash p$	1	5	10	25	50
10	2.0	1.3	–	–	–
50	5.7	4.7	3.1	1.5	–
100	6.7	7.5	5.6	2.5	1.5
250	8.7	14.8	12.4	5.5	3.0
500	12.5	15.9	21.2	10.2	5.4

3.1. Generalized dynamic factor model in state space form

Consider the dynamic factor model (1.1) where disturbance u_t is a serially uncorrelated noise process with mean zero and variance matrix Σ_u and where the $r \times 1$ vector of factors f_t is modeled by a vector autoregressive moving average (VARMA) process. This model can be represented in state space form using the companion form of a VARMA process. For example, Shumway and Stoffer (2000) show that the VARMA process for f_t in state space has $f_t = G\alpha_t$ with $G = (I_p, 0, \ldots, 0)$ and where the block vector elements of the state vector α_t represents different components of the VARMA process. The dynamic factor model under consideration in state space form (2.2) has

$$Z = \Lambda G = (\Lambda, 0, \ldots, 0)', \qquad \Sigma_\varepsilon = \Sigma_u, \qquad (3.19)$$

and with matrices T, R and Σ_η given by Shumway and Stoffer (2000, Section 6.11). In relation to Lemma 2.1 in Section 2.2, $Z^\dagger = \Lambda$ when Λ has full column rank.

In more general versions of the dynamic factor model, the disturbance vector u_t can be subject to serial correlation. The assumption of a vector autoregressive (VAR) process for u_t is usually appropriate. We have

$$u_t = \Psi_1 u_{t-1} + \ldots + \Psi_{q_\Psi} u_{t-q_\Psi} + \varepsilon_t, \qquad (3.20)$$

with autoregressive coefficient matrices Ψ_j for $j = 1, \ldots, q_\Psi$ and where disturbance vector ε_t is defined below (2.2). In most practical applications, the coefficient matrix Ψ_j is diagonal. We can extend the state vector α_t to accommodate u_t but with $p > N$ our devices cannot be exploited and computations will be slow due to high-dimensional vectors y_t and α_t. Alternatively, we can transform y_t into

$$x_t = y_t - \Psi_1 y_{t-1} - \ldots - \Psi_{q_\Psi} y_{t-q_\Psi}, \qquad (3.21)$$

and then we consider the state space form (2.2) for the resulting model for x_t, that is

$$x_t = Z\alpha_t + \varepsilon_t, \qquad Z = (\Lambda, -\Psi_1\Lambda, \ldots, -\Psi_{q_\Psi}\Lambda), \qquad \alpha_t = (f'_t, \ldots, f'_{t-q_\Psi})'. \qquad (3.22)$$

The state equation for α_{t+1} can be adjusted accordingly for this composition of the state vector and where f_t follows a VARMA process. When q_Ψ is moderate and N is large, we have $N > p$ and the devices of Section 2 can be adopted. In Lemma 2.1, we have $Z^\dagger = Z$ when all matrices concerned are full rank. When the VARMA process contains

many lags, the state vector is extended by more appropriate components of the VARMA process of f_t and matrix Z is augmented accordingly with zero columns; see (3.19).

Finally, a generalized version of the dynamic factor model also allows for a lagged response of the factors on the observation vector y_t. In this case we have the observation equation

$$y_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \ldots + \Lambda_{q_\Lambda} f_{t-q_\Lambda} + u_t,$$

where Λ_j are loading coefficient matrices for $j = 0, 1, \ldots, q_\Lambda$. For example, with $q_\Psi = 2$ and $q_\Lambda = 1$, the state space form (2.2) for x_t in (3.21) has matrix Z and state vector given by

$$
\begin{aligned}
Z &= (\Lambda_0, \, \Lambda_1 - \Psi_1\Lambda_0, \, -\Psi_1\Lambda_1 - \Psi_2\Lambda_0, \, -\Psi_2\Lambda_1), \\
\alpha_t &= (f_t', f_{t-1}', f_{t-2}', f_{t-3}')'.
\end{aligned}
\tag{3.23}
$$

For moderate values of q_Ψ and q_Λ but with N large, we can still expect that $N > p$.

3.2. Treatment of constant vector

It is common practice in many empirical studies to standardize the data before a dynamic factor analysis is carried out. However, in other studies the constant vector is of sufficient interest to include it in the model, see Diebold et al. (2006) for a recent illustration. The observation equation in (2.2) with a constant vector becomes $y_t = \mu + Z\alpha_t + u_t$ where μ is a $N \times 1$ vector of constants. In our transformation method of Section 2.2, a transformed constant will appear in both y_t^L and y_t^H while still we require an estimate of the original constant vector.

The models in (2.6) for the transformed data vectors from the observation equation in (2.2) with constant vector μ are given by

$$y_t^L = \mu^L + Z^L\alpha_t + \varepsilon_t^L, \qquad y_t^H = \mu^H + \varepsilon_t^H, \tag{3.24}$$

where $\mu^L = A^L\mu$ and $\mu^H = A^H\mu$. The low-dimensional constant vector μ^L only appears in the loglikelihood function $\ell(y^L; \psi)$ which can be evaluated by the Kalman filter to the state space model (2.2) with the observation equation replaced by the first equation in (3.24); see Section 3.3. As a consequence of the transformation, the loglikelihood function $\ell(y^L; \psi)$ does not depend on the high-dimensional vector μ^H but μ^H does appear in $\ell(y^H; \psi)$ which, apart from a constant, is given by

$$-\frac{1}{2}\sum_{t=1}^{n}(y_t^H - \mu^H)'\Sigma_H^{-1}(y_t^H - \mu^H) = -\frac{1}{2}\sum_{t=1}^{n}(e_t - M_Z\mu)'\Sigma_\varepsilon^{-1}(e_t - M_Z\mu), \tag{3.25}$$

where $M_Z = I_N - Z(Z'\Sigma_\varepsilon^{-1}Z)^{-1}Z'\Sigma_\varepsilon^{-1}$ and $e_t = M_Z y_t$. The equality in (3.25) is the result of Lemma 2.2. Given equation (1.32) in Appendix A.2, it follows that concentrating out μ^H is equivalent to concentrating out $M_Z\mu$ from the loglikelihood function $\ell(y^H; \psi)$. The GLS estimator of $M_Z\mu$ is denoted by $\hat{\mu}_{\perp Z}$ and is given by

$$\hat{\mu}_{\perp Z} = \frac{1}{n}\sum_{t=1}^{n} e_t. \tag{3.26}$$

The concentrated or profile loglikelihood function with respect to μ is given by (2.18) where e_t is replaced by the term $e_t - \hat{\mu}_{\perp Z}$ for $t = 1, \ldots, n$. The computation of the profile loglikelihood function for y^L is discussed in the next Section 3.3.

Finally, the GLS estimator of μ can be obtained via the identity

$$\mu = P_Z \mu^L + M_Z \mu, \qquad \text{where} \quad P_Z = Z(Z'\Sigma_\varepsilon^{-1}Z)^{-1}. \tag{3.27}$$

The GLS estimator of μ is denoted by $\hat{\mu}$ and is given by

$$\hat{\mu} = \hat{\mu}_{\perp Z} + P_Z \hat{\mu}^L, \tag{3.28}$$

where $\hat{\mu}^L$ is obtained from an augmented Kalman filter as described in the next section.

3.3. Regression effects in state vector

When regression effects become part of the state vector, we need to explicitly treat them within the Kalman filter and related methods. Rosenberg (1973) and de Jong (1991) propose to augment the Kalman filter by additional N-dimensional recursions, one for each regression coefficient. The number of time series in a dynamic factor model can be high and direct application of such methods can become infeasible. However, we can use the earlier results to overcome this computational burden.

We consider the state space model (2.2) with the following extension

$$y_t = \mu + Z\alpha_t + \varepsilon_t, \qquad \alpha_{t+1} = X_t\beta + T\alpha_t + R\eta_t,$$

where matrix X_t consists of covariates and vector β contains regression coefficients. The transformation of Section 2.2 affects the observation vector but not the state vector. After the transformation, we therefore obtain the model equations

$$y_t^L = \mu^L + Z^L\alpha_t + \varepsilon_t^L, \qquad y_t^H = \mu^H + \varepsilon_t^H, \qquad \alpha_{t+1} = X_t\beta + T\alpha_t + R\eta_t.$$

The generalized least squares (GLS) estimates of the coefficient vectors μ^L and β only rely on y_t^L and can be computed by the augmented Kalman filter for the low-dimensional model of y_t^L. The profile or concentrated loglikelihood function for y^L with respect to μ^L and β can also be computed from the augmented Kalman filter; see Rosenberg (1973) and de Jong (1991). The GLS estimate of the constant vector μ^H and the corresponding profile or concentrated loglikelihood function for y^H is obtained as described in Section 3.2. The profile loglikelihood function of y is simply the sum of those for y^L and y^H.

3.4. Bayesian inference and simulation smoothing

As an alternative to the maximum likelihood method of Section 2.6, we can treat parameter vector ψ as a variable, have a prior distribution $p(\psi)$ and perform a Bayesian analysis to obtain the posterior distribution $p(\psi|y)$ for a given dynamic factor model such as (1.1). Examples of Bayesian approaches to dynamic factor models are Aguilar and West (2000) and Fiorentini et al. (2004) in the context of modelling volatility in time series. We follow their approach by adopting Markov chain Monte Carlo (MCMC) methods because a closed form expression for $p(\psi|y)$ is usually not available. The MCMC method generate samples from a Markov chain that has the posterior density of the parameters as its stationary distribution. After a burn-in period, the samples can be used as correlated draws from the posterior distributions. Reviews of MCMC and related algorithms for Bayesian inference are given by, amongst others, Gilks et al. (1996) and Chib (2001).

Consider a dynamic factor model as represented by the state space model (2.2). The set of state vectors $\alpha_1, \ldots, \alpha_n$ is denoted by α. A typical MCMC algorithm for the state space model is given by: (i) initialize α and ψ; (ii) sample α from $p(\alpha|\psi, y)$; (iii) sample ψ from

$p(\psi|\alpha, y)$; (iv) goto (ii). The samples in step (iii) are ideally generated from well-defined densities. When they are not available, MCMC variants such as the Metropolis-Hastings methods can be considered. The main computational challenge is however step (ii) given the high-dimensional observation vector y. We propose to adopt the methods of Section 2 to perform this step in a computationally efficient way. We can sample the states α_t conditional on the observations, y, using such algorithms as developed by Fruhwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002). In particular, the following algorithm of Durbin and Koopman (2002) is fast and easy to implement while being able to fully exploit the results of Section 2.

Suppose the vectors $\alpha^+ = (\alpha_1^{+\prime}, \ldots, \alpha_n^{+\prime})'$ and $y^+ = (y_1^{+\prime}, \ldots, y_n^{+\prime})'$ are samples from the unconditional density $p(\alpha, y|\psi)$. Denote $\widetilde{\alpha}$ as the set $\widetilde{\alpha}_1, \ldots, \widetilde{\alpha}_n$ where $\widetilde{\alpha}_t$ is a sample from the density $p(\alpha_t|\psi, y)$ for $t = 1, \ldots, n$. A simulated state vector $\widetilde{\alpha}_t$ is then computed by

$$\widetilde{\alpha}_t = \alpha_t^+ - \mathbb{E}(\alpha_t|\psi, y_1^+, \ldots, y_n^+) + \mathbb{E}(\alpha_t|\psi, y_1, \ldots, y_n), \qquad t = 1, \ldots, n, \qquad (3.29)$$

where $\mathbb{E}(\alpha_t|\psi, y_1, \ldots, y_n)$ is effectively the MMSLE $a_{t|n}$ as defined in (2.3) and is computed by the Kalman filter and smoother. In the same way $\mathbb{E}(\alpha_t|\psi, y_1^+, \ldots, y_n^+)$ is computed but with data y replaced by the sampling realisation y^+.

By choosing a matrix A such that it fulfills the conditions of Section 2.2, we transform the data vector y into y^L and y^H. As a result, we have $p(\alpha|\psi, y) = p(\alpha|\psi, y^L)$ and we can obtain $\widetilde{\alpha}_t$ from $p(\alpha^+, y^{L+}|\psi)$ from the state space model where the observation equation is the first equation of (2.6). It follows that we can compute $\widetilde{\alpha}_t$ by

$$\widetilde{\alpha}_t = \alpha_t^+ - \mathbb{E}(\alpha_t|\psi, y_1^{L+}, \ldots, y_n^{L+}) + \mathbb{E}(\alpha_t|\psi, y_1^L, \ldots, y_n^L), \qquad t = 1, \ldots, n. \qquad (3.30)$$

This is computationally more efficient than the original algorithm since all computations are based on the low-dimensional vectors y_t^L and y_t^{L+}. Similar advances based on the results in Section 2.2 can be exploited for other simulation smoothing methods.

4. ILLUSTRATION

We present an illustration of the likelihood-based treatment of the dynamic factor model for a large panel of US macroeconomic time series using state space methods. We analyse the data set of Stock and Watson (2005) with monthly US macroeconomic time series from 1960:1 through 2003:12. The data is transformed and differenced to obtain a stationary panel of $N = 132$ time series with $n = 528$; the details of each series and its transformation are given in their Appendix A. The 132 series are categorized into 15 sectors, with number or series in each sector between paranthesses: 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). For all series, observations larger than 6 times the standard deviation of the series, σ, (in absolute value) are set to $\pm 6\sigma$. In total, 46 (out of $69,696$) observations are *winsorized* in this way (0.066%). Then each time series is standardized such that its sample mean equals zero and its sample variance equals one.

Our empirical illustration is close in spirit to the likelihood-based analyses of Bernanke et al. (2005) and Boivin and Giannoni (2006). The estimation of parameters is based

on maximizing the likelihood function. The dynamic properties of the factors can be analyzed by investigating the estimated coefficients. 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

$$y_t = \Lambda f_t + u_t, \qquad f_t = \Phi_1 f_{t-1} + \zeta_t, \qquad u_t = \Psi_1 u_{t-1} + \varepsilon_t, \tag{4.31}$$

where Λ is the $N \times r$ loading matrix, f_t is the $r \times 1$ vector of factors, u_t is the autoregressive disturbance vector, and Φ_1 and Ψ_1 are autoregressive coefficient matrices. The disturbance vectors ε_t and ζ_t have mean zero and variance matrices Σ_ε and Σ_ζ, respectively, and are serially and mutually uncorrelated at all times. The model is represented in its state space form as discussed in Section 3.1. For identification purposes, we let the r top rows of the loading matrix Λ form a lower-triangular matrix with its diagonal elements restricted to one and we let the variance matrix Σ_ζ be diagonal. We further assume that both the variance matrix Σ_ε and the autoregressive coefficient matrix Ψ_1 are diagonal matrices.

The illustration considers two model specifications: Model I has $r = 7$ and $\Psi_1 = 0$ such that $u_t = \varepsilon_t$; Model II has $r = 4$ and a non-zero diagonal matrix Ψ_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. In Model II we therefore set $r = 4$ and obtain a state vector of dimension $m = 8$ when introducing autoregressive disturbances of order 1, that is $q_\Psi = 1$ in (3.20). The dimensions of the two model specifications are reported in Table 2.

Table 2. Dynamic factor model specifications

The table reports dimensions for two specifications of the dynamic factor model (4.31), for its state space form (2.2) and for its parameter vector ψ. In model I, the vector autoregressive process for disturbance vector u_t has $\Phi_1 = 0$ such that $u_t = \varepsilon_t$ with a diagonal variance matrix Σ_ε. In model II, we have a non-zero diagonal autoregressive coefficient matrix Ψ_1. Further details of the model are given below model equation (4.31).

	model (4.31)			state space			parameter vector ψ				
	r	q_Λ	q_Ψ	N	p	m	Λ	Φ_1	Ψ_1	Σ_ε	total
I	7	0	0	132	7	7	903	49	0	132	1084
II	4	0	1	132	8	8	522	16	132	132	802

4.1. Parameter estimates

Table 3 presents the maximum likelihood estimates of all the elements in the vector autoregressive (VAR) coefficient matrix Φ_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 Φ_1. We learn from Table 3 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 Φ_1.

Table 3. Maximum likelihood estimates of VAR coefficients

We report the QML estimates of the coefficients in the $r \times r$ matrix Φ_1 for Models I ($r = 7$) and II ($r = 4$). The eigenvalues of the estimates of Φ_1 are reported in descending order. For complex eigenvalues we present both the real and imaginary (img) components.
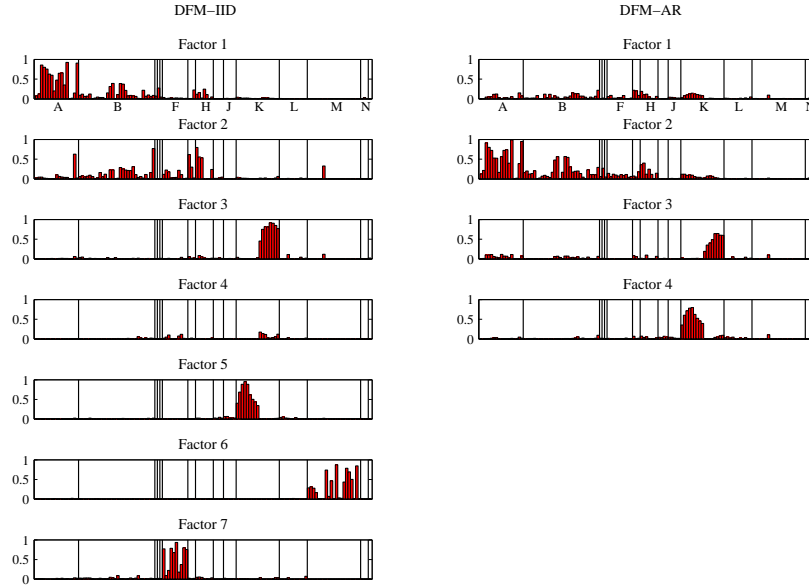
Model I

Factor	VAR coefficients							Eigenvalues	
	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

Factor	VAR coefficients				Eigenvalues	
	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 factor loading estimates for Λ 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 factor estimate. The series of N regressions can be repeated for each factor and the resulting N dimensional series of R^2 statistics can be displayed as an index plot for each factor. We present the R^2 statistics for the seven factors of Model I in the left panel of Figure 1. 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 sector (G) real inventories and sector (H) but also correlated with variables in the sector (F) housing starts & sales and sector (B). The two individual variables production and unemployment in sectors (A) and (B) are particularly highly correlated with the second factor. The third and fifth factors are associated with the variables interest rates and spreads, respectively, from sector (K). The R^2 statistics for the four factors in Model II are presented in the right panel of Figure 1. The third and fourth factors in Model II are

Figure 1. 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 $\bar{\Lambda}$, Φ_1 and Σ_ϵ) and in the right-panel for the four factors in Model II (with QML estimates for $\bar{\Lambda}$, Φ_1, Ψ_1 and Σ_ϵ).

comparable to the third and fifth factors of Model I, respectively. The first two factors of Model II are a mix of the first two factors of Model I which are associated with the "real" sectors (A), (B), (F) and (H).

4.2. Diagnostic checking

An attractive feature of a model-based analysis is that model misspecification tests and diagnostic statistics concerning normality, heteroskedasticity and serial correlation can be considered. In time series analysis, diagnostic tests are usually applied to standardized one-step ahead prediction errors $F_t^{-1/2} v_t$ and can be obtained from the Kalman filter (2.4). When the model is correctly specified, the prediction errors are serially uncorrelated. 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 in comparison with other specifications.

The devices in Section 2 allow us to compute the prediction errors for all 132 series in a few seconds. To illustrate residual diagnostics in the context of dynamic factor analysis, we compute for each residual series the serial correlation portmanteau χ^2 test of Ljung and Box (1978); it is based on the sum of the first q sample residual autocorrelations. The Box-Ljung statistics for the 132 time series are graphically presented as index plots in Figure 2 for $q = 5$. The upper panel presents the residual statistics for Model I while the

Figure 2. Box-Ljung statistics

We present the Box-Ljung statistics based on the first 5 residual autocorrelations for the dynamic factor models I and II.
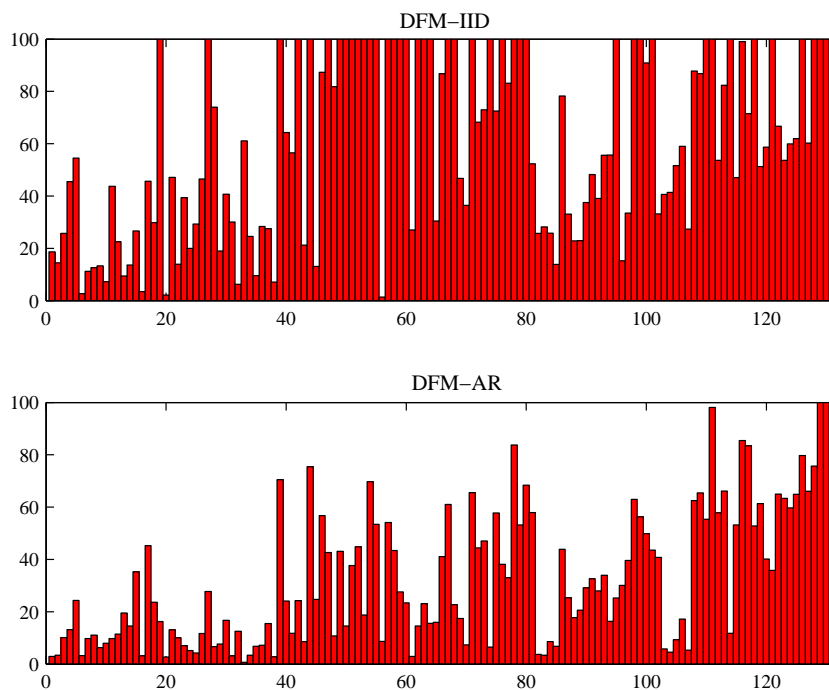
lower panel presents those for Model II. The displayed Box-Ljung values are truncated at 100. It is evident that for many series the null hypothesis of no serial correlation in the residuals is rejected. The current dynamic factor models are therefore not highly satisfactory for this panel of macroeconomic time series. Model II is most successful in capturing the collective dynamics in the data set for this comparison.

Although we acknowledge that both dynamic factor model specifications are incorrect, an important purpose of the presented illustration is to show that our likelihood-based methodology makes it computationally feasible to distinguish models that are closer to the correct specification than other models. The presented methods have facilitated parameter estimation and computation of diagnostic tests in such a way that they become a matter of routine for even high-dimensional dynamic factor models.

5. CONCLUSIONS

We have presented a number of new results which are instrumental for an effective likelihood-based analysis of dynamic factor models. It is shown that a high-dimensional dynamic factor model can be reduced to a low-dimensional state space model. This insight leads to substantial computational savings when estimating factors and evaluating the loglikelihood function. These developments extend to nonlinear and non-Gaussian state space models for simulation-based methods but also for generalisations such as

the extended Kalman filter and the unscented Kalman filter as described in Julier and Uhlmann (1997). Our empirical illustration is concerned with macroeconomic forecasting as discussed in Stock and Watson (2002b) where they advocate a two-step approach: (i) extract a sufficient number of factors 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 as in Bräuning and Koopman (2014). The likelihood-based methods may become a viable alternative to the two-step approaches when employing the results of this paper. Future work may 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.

ACKNOWLEDGEMENTS

A. APPENDICES

A.1. Proof of Lemma 2.1

From conditions (i), (ii) and (iii) in Section 2.2 and using that $\Sigma_\varepsilon A^{L\prime}$ has full column rank, we obtain

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

Define $Z^\dagger = \Sigma_\varepsilon A^{L\prime}$ then $A^L = Z^{\dagger\prime}\Sigma_\varepsilon^{-1}$ and $\mathrm{Col}\{Z^\dagger\} = \mathrm{Col}\{Z\}$. This proves the necessity part of Lemma 2.1.

A.2. Proof of Lemma 2.2

We have

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

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

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

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

$$J^{H\prime} = \Sigma_\varepsilon A^{H\prime}(A^H\Sigma_\varepsilon A^{H\prime})^{-1}A^H = I - Z^\dagger(Z^{\dagger\prime}\Sigma_\varepsilon^{-1}Z^\dagger)^{-1}Z^{\dagger\prime}\Sigma_\varepsilon^{-1} \overset{def.}{=} M_Z. \qquad (1.32)$$

The proof of (2.17) is completed by the identity $J^H \Sigma_\varepsilon^{-1} = J^H \Sigma_\varepsilon^{-1} J^{H\prime}$ and the definition $e_t \stackrel{def.}{=} M_Z y_t$ where e_t is the GLS residual from data vector y_t with covariate $Z^\dagger$ and variance matrix Σ_ε.

A.3. 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

$$\frac{\partial \ell(y; \psi)}{\partial \psi}\bigg|_{\psi=\psi^*} = \frac{\partial Q(\psi^*|\psi)}{\partial \psi}\bigg|_{\psi=\psi^*}, \tag{1.33}$$

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

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

with $p(y, \alpha; \psi)$ as the joint density of y and $\alpha = (\alpha_1', \ldots, \alpha_n')'$. For time series models in the state space form (2.2), $Q(\psi^*|\psi)$ can be expressed by

$$Q(\psi^*|\psi) = c - \frac{n}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\text{tr } Q_\varepsilon - \frac{n-1}{2}\log|\Sigma_\eta| - \frac{1}{2}\text{tr } Q_\eta$$
$$- \frac{1}{2}\log|P| - \frac{1}{2}\text{tr}[P^{-1}\{(a_{1|n} - a)(a_{1|n} - a)' + P_{1|n}\}], \tag{1.34}$$

where c is a constant independent of ψ and

$$Q_\varepsilon = \Sigma_\varepsilon^{-1}\sum_{t=1}^{n}\{\hat{\varepsilon}_t \hat{\varepsilon}_t' + \text{Var}(\varepsilon_t|y; \psi)\}, \qquad Q_\eta = \Sigma_\eta^{-1}\sum_{t=2}^{n}\{\hat{\eta}_t \hat{\eta}_t' + \text{Var}(\eta_t|y; \psi)\}, \tag{1.35}$$

where $\hat{\varepsilon}_t = \mathbb{E}(\varepsilon_t|y; \psi)$, $\text{Var}(\varepsilon_t|y; \psi)$, $\hat{\eta}_t = \mathbb{E}(\eta_t|y; \psi)$ and $\text{Var}(\eta_t|y; \psi)$ can be expressed in terms of $a_{j|n}$ and $P_{j|n}$ for $j = 1, \ldots, n$, and which are evaluated using the Kalman filter and smoother of Section 2.

The derivatives of (1.34) with respect to the system matrices Z, T, Σ_ε and Σ_η in (2.2), for the case of $R = I$, are given by

$$\frac{\partial \ell(y)}{\partial Z} = \Sigma_\varepsilon^{-1}\left(\sum_{t=1}^{n} y_t\, a_{t|n}' - Z S_{1:n}^{(0)}\right), \qquad \frac{\partial \ell(y)}{\partial \Sigma_\varepsilon} = Q_\varepsilon^* \Sigma_\varepsilon^{-1} - \frac{1}{2}\text{diag}(Q_\varepsilon^* \Sigma_\varepsilon^{-1}),$$

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

where $Q_\varepsilon^* = Q_\varepsilon - n$, $Q_\eta^* = Q_\eta - n - 1$, with Q_ε and Q_η defined in (1.35),

$$S_{j:k}^{(0)} = \sum_{t=j}^{k} a_{t|n} a_{t|n}' + P_{t|n}, \qquad S_{j:k}^{(1)} = \sum_{t=j}^{k} a_{t|n} a_{t-1|n}' + P_{t,t-1|n}, \tag{1.36}$$

for $j, k = 1, \ldots, n$, with $j \leq k$, where $a_{t|n}$, $P_{t|n}$ and $P_{t-1,t|n} = P_{t,t-1|n}'$ can be evaluated by Kalman filter and smoother algorithms. Since the system matrices are functions of ψ, the score vector with respect to ψ is obtained via the chain rule for any $\psi = \psi^*$.

It is shown that the Kalman filter and smoother are key in the computation of the score vector. In Section 2.3 we have shown that these methods can be based on the low-dimensional observation equation (2.11) while the computation of matrix A^H and the

time series y_t^H are not required. Hence the computational gains reported in Section 2.7 apply here.

A.4. The EM algorithm

The EM algorithm is developed in the seminal article of Dempster et al. (1977). It is an iterative algorithm that repeatedly performs two types of calculations: (E)xpectation and (M)aximization. For a given parameter vector $\psi = \psi^*$, the E and M steps are given by: (E) evaluate the expected complete loglikelihood function $Q(\psi^*|\psi)$ as given by (1.34); (M) maximize $Q(\psi^*|\psi)$ with respect to ψ. The M step produces a vector ψ^+ with the property $\ell(y; \psi^+) \geq \ell(y; \psi^*)$. If the EM steps are continuously repeated, convergence to a (local) optimum of $\ell(y; \psi)$ is guaranteed, see Wu (1983) for a more detailed discussion. Shumway and Stoffer (1982) and Watson and Engle (1983) have proposed the use of the EM algorithm in the context of time series models in state space form. A feasible EM algorithm for high-dimensional dynamic factor models is obtained by applying the devices of Section 2 in the E step. The evaluation of $Q(\psi^*|\psi)$ relies on the Kalman filter and smoother which can be based on the low-dimensional observations y_t^L; see Appendix A.3.

REFERENCES

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

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

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

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

Bai, J. and S. Ng (2002). Determining the number of 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.

Banbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *J. Applied Econometrics 25*, 71–92.

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.

Bräuning, F. and S. J. Koopman (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International J. of Forecasting 30*, forthcoming.

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

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

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

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

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

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

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

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

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

Diebold, F., S. Rudebusch, and S. Aruoba (2006). The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach. *J. Econometrics 131*, 309–338.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *J. Econometrics 164*, 188–205.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *Rev. Economics and Statistics 94*, 1014–1024.

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

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). 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.

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

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

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

Julier, S. J. and J. K. Uhlmann (1997). A new extension of the Kalman filter to nonlinear systems. In I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition VI*, Volume 3068, pp. 182–193.

Koop, G. and D. Korobilis (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*, 267–358.

Koopman, S. J. and J. Durbin (2000). Fast filtering and smoothing for multivariate state space models. *J. Time Series Analysis 21*, 281–96.

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

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

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

Mesters, G. and S. J. Koopman (2014). Generalized dynamic panel data models with random effects for cross-section and time. *J. Econometrics*, forthcoming.

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. W. Watson (Eds.), *Business cycles, indicators and forecasting*, pp. 285–306. Chicago: University of Chicago Press.

Rosenberg, B. (1973). Random coefficients models: the analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement 2*, 399–428.

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

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

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

Shumway, R. H. and D. S. Stoffer (2000). *Time Series Analysis and Its Applications*. New York: Springer-Verlag.

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

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

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

Stock, J. H. and M. W. 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.

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.