TI 2014-061/III
Tinbergen Institute Discussion Paper

Empirical Bayes Methods for Dynamic Factor Models

Siem Jan Koopman
Geert Mesters

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

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

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

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

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Empirical Bayes Methods for Dynamic Factor Models

S.J. Koopman[(a,b,c)] and *G. Mesters*[(a,b,d)*]

[(a)] Department of Econometrics, VU University Amsterdam

[(b)] Tinbergen Institute

[(c)] CREATES, Aarhus University

[(d)] Netherlands Institute for the Study of Crime and Law Enforcement

May 22, 2014

Abstract

We consider the dynamic factor model where the loading matrix, the dynamic factors and the disturbances are treated as latent stochastic processes. We present empirical Bayes methods that enable the efficient shrinkage-based estimation of the loadings and the factors. We show that our estimates have lower quadratic loss compared to the standard maximum likelihood estimates. We investigate the methods in a Monte Carlo study where we document the finite sample properties. Finally, we present and discuss the results of an empirical study concerning the forecasting of U.S. macroeconomic time series using our empirical Bayes methods.

JEL classification: C32; C43

Some keywords: Importance sampling; Kalman filtering; Likelihood-based analysis; Posterior modes; Rao-Blackwellization; Shrinkage.

*S.J. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Corresponding author: G. Mesters, Department of Econometrics, VU University Amsterdam, The Netherlands. Address: De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, tel. nr. +31 (0) 6 455 86 453, email: g.mesters@vu.nl, web: http://www.geertmesters.nl.

1 Introduction

Consider the dynamic factor model for N time series and T time periods given by

$$y_{i,t} = \boldsymbol{\lambda}_i' \boldsymbol{\alpha}_t + \epsilon_{i,t}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{1}$$

where $y_{i,t}$ is the observation corresponding to variable i and time period t, $\boldsymbol{\lambda}_i$ is the $r \times 1$ vector of factor loadings, $\boldsymbol{\alpha}_t$ is the $r \times 1$ vector of dynamic factors and $\epsilon_{i,t}$ is the disturbance term. The aim is to decompose the vector of time series observations $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{N,t})'$ into two independent components: a common component that is driven by r common dynamic processes in the vector $\boldsymbol{\alpha}_t$ and an idiosyncratic component represented by the N time series processes $\epsilon_{i,t}$. Dynamic factor models are typically used for macroeconomic forecasting or structural analysis; see for example Stock en Watson (2002a) and Bernanke *et al.* (2005). The reviews of Bai en Ng (2008) and Stock en Watson (2011) provide more discussion and references.

We consider the dynamic factor model (1) for the case where the loading vectors $\boldsymbol{\lambda}_i$, the dynamic factor vectors $\boldsymbol{\alpha}_t$ and the disturbances $\epsilon_{i,t}$ are all treated as stochastic processes. More specifically, the loading vectors are assumed to be normally and independently distributed while the dynamic factors are specified as a stationary vector autoregressive process. The stochastic assumptions for both the loadings and factors have been considered earlier in Bayesian dynamic factor analysis; see for example Aguilar en West (2000). However, they contrast with most other specifications, where the elements of the loading vectors and possibly the factors are treated as deterministic unknown variables; see Stock en Watson (2011).

For the dynamic factor model (1) with stochastic components we develop empirical Bayesian methods for the estimation of the loadings and the factors. This approach combines the benefits of dimension reduction imposed by the structure of the dynamic factor model, see Forni *et al.* (2000) and Stock en Watson (2002b), with shrinkage-based parameter estimation, see James en Stein (1961), Efron en Morris (1973), Knox *et al.* (2004) and Efron (2010). In particular, we estimate the loadings and factors using filtering methods and the vector of unknown parameters, which is associated with the stochastic processes for

$\boldsymbol{\lambda}_i$, $\boldsymbol{\alpha}_t$ and $\epsilon_{i,t}$, using the maximum likelihood method. The resulting filtered estimates have lower expected quadratic loss compared to the standard maximum likelihood estimates for approximate dynamic factor models. The implementation of empirical Bayes methods for the dynamic factor model is non-trivial given the product of stochastic variables $\boldsymbol{\lambda}_i$ and $\boldsymbol{\alpha}_t$ in (1). Standard state space methods, as discussed in Durbin en Koopman (2012), for example, cannot be used and need to be modified. In particular, we provide three new results.

First, we apply the iterative conditional mode algorithm of Besag (1986) for obtaining the posterior modes of the loadings and the factors, simultaneously. The algorithm iterates between the updating of the loadings conditional on the factors and vice versa. We show that this algorithm can be implemented in a computationally efficient manner by exploiting the results of Jungbacker en Koopman (2014) and Mesters en Koopman (2014). We show that after convergence we have obtained the joint posterior mode of the loadings and factors.

Second, we develop a two-step estimation procedure for the vector of deterministic model parameters using likelihood-based methods. In the first step we treat the elements of the loading matrix deterministic and estimate the model using standard state space methods; see Jungbacker en Koopman (2014) and Doz *et al.* (2012). This gives maximum likelihood estimates for the loadings and the parameters that pertain to the distributions of the factors and the disturbances. In the second step we estimate the parameters that pertain to the distribution of the loadings using the maximum likelihood estimates for the loadings as observations.

Third, we consider the estimation of other posterior statistics, such as the posterior mean and the posterior variance of the loadings and the factors. We argue that analytical solutions are not available when both $\boldsymbol{\lambda}_i$ and $\boldsymbol{\alpha}_t$ are considered stochastic; see also the arguments provided in Bishop (2006, Chapters 8 and 12). We can resort to simulation methods which are a standard solution for the estimation of latent variables in nonlinear models. However, given the typical large dimensions of the dynamic factor model ($N, T > 100$), standard simulation methods converge slowly and are unreliable because they are subject to so-called infinite variance problems; see Geweke (1989). To solve this problem we factorize the estimation into two separate parts: one for the loadings and one for the factors. We show that the integral over the factors can be calculated analytically, while the integral over the remaining $\boldsymbol{\lambda}$-dependent terms can be calculated using basic simulation methods. The performance of

the "integrated" simulation-based estimation methods is more stable and has overall good properties.

The factors and loadings of the dynamic factor model are traditionally estimated by principal components and maximum likelihood methods; see for example Watson en Engle (1983), Stock en Watson (1989), Sargent (1989), Forni *et al.* (2000), Stock en Watson (2002b), Bai (2003), Bai en Ng (2008), Doz *et al.* (2012), Bai en Li (2012), and Jungbacker en Koopman (2014), where the loadings and, in some occasions, the factors are treated deterministic. More recently, a variety of two-step estimation methods have been proposed to combine the favorable elements of both traditional estimation methods. They typically rely on a principal components step to first approximate the factor space after which regression or state space methods are used to estimate the underlying factor dynamics; see for examples Stock en Watson (2005), Giannone *et al.* (2008), Doz *et al.* (2011), Breitung en Tenhofen (2011) and Bräuning en Koopman (2013). Alternatively, and more related to our model specification, Bayesian Markov Chain Monte Carlo (MCMC) methods for state space models can be used to jointly estimate the factors and loadings. In this approach, the loadings are typically treated as part of the parameter vector; examples are Kose *et al.* (2003) and Bernanke *et al.* (2005). Recent advances in Bayesian dynamic factor analysis that aim to model the loading matrix more sparsely include Bhattacharya en Dunson (2011), Kaufmann en Schumacher (2013) and Nakajima en West (2013).

The benefits of our model specification and estimation methods can be summarized as follows. First, we show that the parametric empirical Bayes joint posterior mode estimates for the loadings and the factors are more accurate in the mean squared error (MSE) sense when compared to the maximum likelihood estimates; see Hansen (2013). The differences are significant and large for both the loadings and the factors. The reductions in MSE for the loadings can be as large as 80% depending on the data generating process, number of factors and panel sizes. Under various forms of misspecification for the disturbances the differences remain the same. Second, given that the loadings and factors are estimated more accurate it follows that the resulting forecast errors are smaller when compared to those resulting from the maximum likelihood estimates. This is illustrated in our empirical study where we consider out-of-sample forecasting for a panel of macroeconomic and financial time series. Third, from a computational perspective, by computing several integrals analytically

we reduce the computational complexity when compared to hierarchical Bayesian MCMC methods that also aim to learn about the prior distributions. While we predominantly focus on empirical Bayes methods, the results can be adapted for full Bayesian inference methods as well. Finally, our methods produce posterior estimates that are not subject to errors and they do not depend on Taylor expansions or discrete function approximations. Therefore they compare favorably to other algorithms for nonlinear state space models such as the extended Kalman filter of Anderson en Moore (1979) and the unscented Kalman filter of Julier en Uhlmann (1997).

This paper develops and evaluates parametric empirical Bayes methods for dynamic factor models in finite sample settings. A full asymptotic analysis of the empirical Bayes posterior mode estimates is beyond the scope of the paper. Such an analysis involves developing a limit theory for the posterior mode estimates when $N, T \to \infty$. Similar analysis have been recently conducted by Doz *et al.* (2012) and Bai en Li (2012), who consider the analysis of maximum likelihood estimates for the dynamic factor model with deterministic loadings. We leave this for future work.

The remainder of this paper is organized as follows. In the next section we detail the specification of the dynamic factor model with stochastic loadings. In Section 3 we discuss the implementation of the empirical Bayesian estimation methods. The finite sample properties of the posterior mode estimates are studied in a Monte Carlo study that is presented in Section 4. The methods are evaluated for different panel sizes and different number of factors. In Section 5 we present the results from our empirical study for macroeconomic forecasting with many predictors (see Stock en Watson (2002b) and Stock en Watson (2012)). Section 6 concludes the paper and provides some directions for further research.

2 Dynamic factor model with stochastic loadings

We define the dynamic factor model with stochastic loadings for N variables which are indexed by $i = 1, \ldots, N$. The variables are observed over time for a span of T periods; each time period is indexed by $t = 1, \ldots, T$. The observation vector $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{N,t})'$ is

modeled by

$$\begin{array}{lcllll} \boldsymbol{y}_t & = & \boldsymbol{\Lambda}\boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t & \sim & NID(\boldsymbol{0}, \boldsymbol{\Omega}), \\ \boldsymbol{\Lambda} & = & \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)', & \boldsymbol{\lambda}_i & \sim & NID(\boldsymbol{\delta}, \boldsymbol{\Sigma}_\lambda), \\ \boldsymbol{\alpha}_{t+1} & = & \boldsymbol{H}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t & \sim & NID(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta), \qquad t = 1, \ldots, T, \end{array} \tag{2}$$

where $\boldsymbol{\Lambda}$ is the $N \times r$ loading matrix, with $r < N$, $\boldsymbol{\alpha}_t$ is the $r \times 1$ vector of common dynamic factors and $\boldsymbol{\epsilon}_t$ is the $N \times 1$ is the disturbance vector, with mean zero and variance matrix $\boldsymbol{\Omega}$. The loading vectors $\boldsymbol{\lambda}_i$ are normally and independently distributed with mean $\boldsymbol{\delta}$ variance $\boldsymbol{\Sigma}_\lambda$. The dynamic factors $\boldsymbol{\alpha}_t$ follow a vector autoregressive process of order one, with $r \times r$ coefficient matrix $\boldsymbol{H}$ and $r \times r$ variance matrix $\boldsymbol{\Sigma}_\eta$. The initial state vector $\boldsymbol{\alpha}_1$ is normally distributed with mean $\boldsymbol{a}_1$ and variance matrix $\boldsymbol{P}_1$. Our estimation methods are sufficiently general to consider more elaborate specifications for the factors. For notational and expositional convenience, we focus on the basic specification in (2).

The observation equation of model (2) can also be written as

$$\boldsymbol{y}_t = (\boldsymbol{\alpha}_t' \otimes \boldsymbol{I}_N)\boldsymbol{\lambda} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Omega}), \tag{3}$$

where $\boldsymbol{\lambda} = \text{vec}(\boldsymbol{\Lambda}) = (\acute{\boldsymbol{\lambda}}_1', \ldots, \acute{\boldsymbol{\lambda}}_r')'$ and the $N \times 1$ vector $\acute{\boldsymbol{\lambda}}_j$ is the j^{th} column of $\boldsymbol{\Lambda}$, for $j = 1, \ldots, r$. This alternative representation of the observation equation is convenient for our exposition below. To identify the space of the model components and the underlying parameters several strategies can be pursued; see Bai en Li (2012) for a recent discussion. The dynamic factor model with stochastic loadings is subject to the following identifying restrictions for the factor and loading spaces.

Assumption 1. *Consider the dynamic factor model in* (2)*. We assume the following.*

(a) ***Common factors***

The $r \times 1$ vector of common factors $\boldsymbol{\alpha}_t$ is stationary and restricted such that $\text{Var}(\boldsymbol{\alpha}_t) = \boldsymbol{I}_r$*; see Ansley en Kohn (1986). The initial state is given by $\boldsymbol{\alpha}_1 \sim N(\boldsymbol{0}, \boldsymbol{I}_r)$, that is $\boldsymbol{a}_1 = \boldsymbol{0}$ and $\boldsymbol{P}_1 = \boldsymbol{I}_r$. The common innovations $\boldsymbol{\eta}_t$ and the initial state $\boldsymbol{\alpha}_1$ are mutually independent and distributed independent of the loading vectors $\boldsymbol{\lambda}_i$ and the disturbances $\epsilon_{i,s}$, for all $i = 1, \ldots, N$ and $s, t = 1, \ldots, T$.*

(b) ***Loading vectors***

The loading vectors $\boldsymbol{\lambda}_i$ *in* (2)*, for* $i = 1, \ldots, N$*, are distributed normally and independently with mean* $\boldsymbol{\delta}$ *and positive definite variance* $\boldsymbol{\Sigma}_\lambda$*. The distribution of the loadings* $\boldsymbol{\lambda}_i$ *is independent of the disturbances* $\boldsymbol{\epsilon}_t$*, for all* $i = 1, \ldots, N$ *and* $t = 1, \ldots, T$*.*

(c) ***Disturbances***

The disturbance vectors $\boldsymbol{\epsilon}_t$ *in* (2) *are distributed normally and independently with mean zero and diagonal variance matrix* $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \ldots, \omega_N^2)$*.*

The assumptions 1.(a) and 1.(c) are standard for exact dynamic factor models and identify the factor space; see Doz *et al.* (2012). Assumption 1.(b) is novel but necessary. The value for $\boldsymbol{\Sigma}_\lambda$ determines the shrinkage for the loadings. Assumption 1.(c) rules out cross-section and serial correlation in the disturbances but allows for heteroskedasticity. Serial correlation can be handled by rewriting the model; see Stock en Watson (2005) and Jungbacker en Koopman (2014). To identify a particular rotation of the loadings and factors we need to impose hierarchical constraints on r vectors of the loading matrix $\boldsymbol{\Lambda}$; see Geweke en Zhou (1996). The choice for the restrictions will typically depend on the empirical application of interest and we consider different choices in Sections 4 and 5.

It follows from Brown en Rutemiller (1977) that under Assumption 1 the mean and variance for the observations $y_{i,t}$ are given by

$$\mathrm{E}(y_{i,t}) = \mathrm{E}(\boldsymbol{\lambda}_i'\boldsymbol{\alpha}_t) + \mathrm{E}(\epsilon_{i,t}) = \mathrm{E}(\boldsymbol{\lambda}_i')\mathrm{E}(\boldsymbol{\alpha}_t) + \mathrm{E}(\epsilon_{i,t}) = 0 \tag{4}$$

$$\begin{aligned}\mathrm{Var}(y_{i,t}) &= \mathrm{Var}(\boldsymbol{\lambda}_i'\boldsymbol{\alpha}_t) + \mathrm{Var}(\epsilon_{i,t}) \\ &= \mathrm{E}(\boldsymbol{\lambda}_i')\mathrm{Var}(\boldsymbol{\alpha}_t)\mathrm{E}(\boldsymbol{\lambda}_i) + \mathrm{E}(\boldsymbol{\alpha}_t')\mathrm{Var}(\boldsymbol{\lambda}_i)\mathrm{E}(\boldsymbol{\alpha}_t) + \mathrm{Tr}\left[\mathrm{Var}(\boldsymbol{\lambda}_i)\mathrm{Var}(\boldsymbol{\alpha}_t)\right] + \mathrm{Var}(\epsilon_{i,t}) \\ &= \mathrm{Tr}(\boldsymbol{\delta}\boldsymbol{\delta}' + \boldsymbol{\Sigma}_\lambda) + \omega_i^2, \end{aligned} \tag{5}$$

where Tr() denotes the trace operator. These moment expressions do not rely on the normality assumptions.

The equations in (2) define the dynamic factor model with stochastic loadings. The parameters pertaining to the model are collected in the vector $\boldsymbol{\psi}$. For our model (2) it contains the unrestricted elements of $\boldsymbol{\delta}$, $\boldsymbol{\Sigma}_\lambda$, $\boldsymbol{H}$, $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Omega}$. We notice that the number of

deterministic parameters is much smaller compared to the standard dynamic factor model with deterministic loadings as considered in for example Jungbacker en Koopman (2014) and Doz *et al.* (2012). When $\boldsymbol{\Sigma}_\lambda \to \infty$ the prior restriction on the loading matrix vanishes and when $\boldsymbol{H} \to \mathbf{0}$ and $\boldsymbol{\Sigma}_\eta \to \infty$ prior restriction on the factors vanishes.

In our exposition below we assume that the panel of observations is balanced and that there are no missing observations. This assumption is merely for notational convenience and in Appendix E we present the adjustments for the methods of Section 3 for the case where there are observations missing. The adjustments are minor.

3 Estimation of the loadings, factors and parameters

We develop the parametric empirical Bayesian methods for the estimation of the loadings, factors and the parameter vector for the dynamic factor model in (2) given the conditions in Assumption 1. The posterior modes of the loadings and factors are estimated using filtering methods and the vector of deterministic parameters is estimated by maximum likelihood. We provide an iterative algorithm for computing the joint posterior modes of the loadings and factors. The posterior mode estimates provide the point estimates and are the subject of our Monte Carlo study in Section 4. Further, we develop methods for the estimation of other posterior statistics for the loadings and the factors. These are necessary for the construction of finite sample confidence intervals.

3.1 Joint posterior mode analysis

The posterior modes for $\boldsymbol{\lambda} = \text{vec}(\boldsymbol{\Lambda})$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_T', \boldsymbol{\alpha}_{T+1}')'$ are defined by

$$\{\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}}\} = \underset{\boldsymbol{\lambda}, \boldsymbol{\alpha}}{\arg\max} \log p(\boldsymbol{\lambda}, \boldsymbol{\alpha} | \boldsymbol{y}; \boldsymbol{\psi}), \tag{6}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$. The direct optimization of $p(\boldsymbol{\lambda}, \boldsymbol{\alpha} | \boldsymbol{y}; \boldsymbol{\psi})$ with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ is complicated as the first order conditions for $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ depend on each other and solving analytically, or numerically, for either one is infeasible when N and T are large. The following theorem shows that we can separate the first order conditions of the posterior density into two parts. One part that can be used for calculating the posterior mode of $\boldsymbol{\lambda}$ and another

part that can be used for calculating the posterior mode of $\boldsymbol{\alpha}$.

Theorem 1. *For $\boldsymbol{y}$ defined by model (2) it holds, under Assumption 1, that*

$$\left.\frac{\partial \log p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\psi})}{\partial(\boldsymbol{\lambda}', \boldsymbol{\alpha}')'}\right|_{\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}}, \boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}} = \left.\frac{\partial \log \left[p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}, \boldsymbol{\psi}) p(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}}, \boldsymbol{\psi})\right]}{\partial(\boldsymbol{\lambda}', \boldsymbol{\alpha}')'}\right|_{\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}}, \boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}},$$

for all given $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\alpha}}$.

The proof is given in Appendix A. The decomposition of the score function in Theorem 1 enables the efficient computation of the posterior modes of both $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$. It holds that $\hat{\boldsymbol{\lambda}}$ can be found by maximizing $\log p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}, \boldsymbol{\psi})$ with respect to $\boldsymbol{\lambda}$. Also, $\hat{\boldsymbol{\alpha}}$ can be found by maximizing $\log p(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}, \boldsymbol{\psi})$ with respect to $\boldsymbol{\alpha}$. The modes, means and variances of $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}, \boldsymbol{\psi})$ and $p(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}, \boldsymbol{\psi})$ can be evaluated using standard methods. In particular, conditionally on $\boldsymbol{\alpha}$ the observation equation (3), together with the marginal density for $\boldsymbol{\lambda}$ forms a linear Gaussian regression model. Multivariate regression methods are used to evaluate the mean and variance of $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}, \boldsymbol{\psi})$. The mode follows from the equality between the mean and mode for the Gaussian density. Also, conditional on $\boldsymbol{\lambda}$ model (2) is a linear Gaussian state space model. The mean, and thus the mode, and variance of $p(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}, \boldsymbol{\psi})$ are evaluated by the Kalman filter smoother; see Durbin en Koopman (2012, Chapter 4).

However, we can only obtain the modes $\hat{\boldsymbol{\lambda}}$ or $\hat{\boldsymbol{\alpha}}$ when we have knowledge of either $\hat{\boldsymbol{\alpha}}$ or $\hat{\boldsymbol{\lambda}}$. The iterative conditional mode algorithm proposed in Besag (1986) provides a simple and stable solution for this problem. For the model (2) the following Theorem summarizes this algorithm.

Theorem 2. *Suppose that $p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\psi})$ is uni-modal in $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ with $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\alpha}}$ being the only stationary points. We assume that the conditions in Assumption 1 hold. For given arbitrary starting values $\boldsymbol{\lambda}^{(s)} \neq \mathbf{0}$, for $s = 0$, and using Theorem 1, the posterior modes $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\alpha}}$ in (6) can be obtained by iterating between*

(i) $\boldsymbol{\alpha}^{(s)} = \mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda}=\boldsymbol{\lambda}^{(s-1)}, \boldsymbol{\psi})$;

(ii) $\boldsymbol{\lambda}^{(s)} = \mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha}=\boldsymbol{\alpha}^{(s)}, \boldsymbol{\psi})$;

(iii) $s = s + 1$,

until convergence.

The proof is presented in Appendix B, together with the details for the efficient computation of $\mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)}, \boldsymbol{\psi})$ and $\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)}, \boldsymbol{\psi})$. The computational advances are based on Jungbacker en Koopman (2014) and Mesters en Koopman (2014). The algorithm can be viewed as the expectation conditional maximization (ECM) algorithm of Meng en Rubin (1993), where the E-step is unity. We notice that the steps (i) and (ii) in this algorithm are M-steps since given the Gaussian Assumption 1 the conditional expectations are equal to conditional maximization steps. The assumption of uni-modality is not different from the assumptions on the ECM algorithm and, as argued in Meng en Rubin (1993), not different from the assumptions for the EM algorithm. In practice, we use as a convergence criteria $||\lambda_{i,j}^{(s)}/\lambda_{i,j}^{(s-1)} - 1|| < 10^{-5}$ and $||\alpha_{j,t}^{(s)}/\alpha_{j,t}^{(s-1)} - 1|| < 10^{-5}$, for all $i = 1, \ldots, N$, $j = 1, \ldots, r$ and $t = 1, \ldots, T$.

While the conditional modes can be calculated iteratively up to any degree of accuracy, we should emphasize that the curvatures around $p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\psi})$ and $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}, \boldsymbol{\psi})p(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}, \boldsymbol{\psi})$ are different. The former takes into account the posterior dependence between $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$, while the latter does not. Nevertheless, the point estimates for the factors in (6) are interesting from a classical perspective as they can be compared to the standard maximum likelihood estimates that compute the posterior mean of the factors given the maximum likelihood estimate for the loading matrix. To operationalize the computation of the posterior mode estimates using Theorem 2, we require a choice for the parameter vector $\boldsymbol{\psi}$, which is the topic of the next section.

3.2 Likelihood evaluation

The parameter vector $\boldsymbol{\psi}$ contains the parameters that pertain to the distributions of the loadings, factors and the disturbances. The estimation of the parameter vector is based on maximum likelihood. For model (2) it is difficult to evaluate the marginal likelihood $L(\boldsymbol{\psi}; \boldsymbol{y}) = p(\boldsymbol{y}; \boldsymbol{\psi})$ analytically. This follows as the product of stochastic variables $\boldsymbol{\Lambda}$ and $\boldsymbol{\alpha}_t$ prohibits closed form solutions for the integral representation of the marginal likelihood

given by

$$p(\boldsymbol{y};\boldsymbol{\psi}) = \int_{\boldsymbol{\alpha}} \int_{\boldsymbol{\lambda}} p(\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{\alpha};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\lambda}\ \mathrm{d}\boldsymbol{\alpha}. \tag{7}$$

More specifically, both sequential methods, such as filtering, and iterative methods, such as the expectation-maximization procedures, require at some point the evaluation of the conditional mean function $\mathrm{E}(\boldsymbol{\Lambda}\boldsymbol{\alpha}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_s;\boldsymbol{\psi})$, for some $s \in \{1,\ldots,T\}$. Since closed form expressions for the conditional expectation of products of stochastic variables do not exist we cannot use these methods.

To solve this problem we rely on a two-step estimation method. We decompose the parameter vector $\boldsymbol{\psi} = \{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2\}$, where $\boldsymbol{\psi}_1 = \{\boldsymbol{\Omega}, \boldsymbol{H}, \boldsymbol{\Sigma}_\eta\}$ and $\boldsymbol{\psi}_2 = \{\boldsymbol{\delta}, \boldsymbol{\Sigma}_\lambda\}$. In the first step we consider the classical state space problem given by

$$\{\tilde{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}_1\} = \underset{\boldsymbol{\lambda},\boldsymbol{\psi}_1}{\arg\max} \log p(\boldsymbol{y}|\boldsymbol{\lambda},\boldsymbol{\psi}_1), \tag{8}$$

where $\boldsymbol{\lambda}$ is treated deterministic; see Jungbacker en Koopman (2014) and Doz *et al.* (2012). The conditional likelihood $p(\boldsymbol{y}|\boldsymbol{\lambda},\boldsymbol{\psi}_1)$ is evaluated via the prediction error decomposition given by

$$\log p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi}_1) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left(\log|\boldsymbol{F}_t| + \boldsymbol{v}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t\right), \tag{9}$$

where the quantities $\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{\Lambda}\boldsymbol{a}_t$ and $\boldsymbol{F}_t = \boldsymbol{\Lambda}\boldsymbol{P}_t\boldsymbol{\Lambda}' + \boldsymbol{\Omega}$ are computed by the Kalman filter. It holds that $\boldsymbol{a}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{t-1},\ldots,\boldsymbol{y}_1,\boldsymbol{\lambda})$ and $\boldsymbol{P}_t = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{t-1},\ldots,\boldsymbol{y}_1,\boldsymbol{\lambda})$, which are also functions of $\boldsymbol{\lambda}$ and are computed by the Kalman filter. The likelihood can be optimized with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}_1$ using numerical optimization methods. Computation efficiency can be improved by using the methods discussed in Jungbacker en Koopman (2014). We emphasize that $\tilde{\boldsymbol{\lambda}}$ is the maximum likelihood estimate for the loadings.

In the second step we estimate the parameters $\boldsymbol{\psi}_2$ by solving

$$\hat{\boldsymbol{\psi}}_2 = \underset{\boldsymbol{\psi}_2}{\arg\max} \log p(\tilde{\boldsymbol{\lambda}};\boldsymbol{\psi}_2), \tag{10}$$

where the observations $\tilde{\boldsymbol{\lambda}}$ are obtained from the first step and the density $p(\tilde{\boldsymbol{\lambda}};\boldsymbol{\psi}_2)$ is implied by assumption 1.(b). It is easy to verify that the procedure yields consistent estimates for the parameter vector $\boldsymbol{\psi}$; see for example Newey en McFadden (1994).

3.3 Posterior Statistics for the Loadings

We consider the evaluation of the posterior means and variances of the loadings and the factors. For the mean and the variance analytical solutions are not available as the product of $\mathbf{\Lambda}$ and $\boldsymbol{\alpha}_t$ prohibits closed form solutions. Instead, we show that when evaluating the mean and variance of the loadings and the factors the factors can be integrated out analytically. The reverse is also possible but computationally much harder. To evaluate the resulting expressions, which only depend on the latent vector $\boldsymbol{\lambda}$, we develop adequate importance densities that are based on the Laplace approximation; see So (2003) and Jungbacker en Koopman (2007). Thus, we do half the work analytically and the other half by simulation methods.

Let $\boldsymbol{f}(\boldsymbol{\lambda})$ denote some arbitrary vector function of $\boldsymbol{\lambda}$. In the simplest case $\boldsymbol{f}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}$. We are interested in estimating the conditional mean function $\bar{\boldsymbol{f}} = \mathrm{E}(\boldsymbol{f}(\boldsymbol{\lambda})|\boldsymbol{y};\boldsymbol{\psi})$. It holds that

$$\bar{\boldsymbol{f}} = \int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}, \tag{11}$$

where the posterior density $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ is of unknown form. This can be seen by noticing that $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \propto p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})p(\boldsymbol{\lambda};\boldsymbol{\psi})$, where $p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})$ follows from the prediction error decomposition that is provided by the Kalman filter; see equation (9). The elements of the latent loading vector $\boldsymbol{\lambda}$ enter the log density (9) nonlinearly in both the mean vector and the variance matrix. Therefore the posterior density $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ has a complicated form.

To solve the problem of evaluating the integral in (11) we make use of the importance sampling technique. We rewrite the integral (11) using the importance density $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$. It holds that

$$\bar{\boldsymbol{f}} = \int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) \frac{p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})}{g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})} g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}. \tag{12}$$

In Appendix C we show that the Monte Carlo estimate of (12) is given by

$$\bar{\boldsymbol{f}} = \frac{\sum_{j=1}^{M} \boldsymbol{f}(\boldsymbol{\lambda}^{(j)}) w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})}{\sum_{j=1}^{M} w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})}, \qquad M \to \infty, \tag{13}$$

where $w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)};\boldsymbol{\psi}) = p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi}) \,/\, g(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})$ and where the samples $\boldsymbol{\lambda}^{(j)}$ are drawn from $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$. The importance density $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ targets $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$, which is the marginal pos-

terior density of the loadings that does not depend on $\boldsymbol{\alpha}$. An adequate importance density $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ that accurately approximates the marginal posterior density $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ needs to be obtained. An initial candidate that seems appropriate is $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}},\boldsymbol{\psi})$. While the location of this density is accurately chosen it can be verified that its variance is too small relative to $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ since $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}},\boldsymbol{\psi})$ does not account for the variance of the posterior mode estimate $\hat{\boldsymbol{\alpha}}$.

Instead, we choose $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ to follow a Gaussian distribution, where the mean and the variance are equal to the mode and the curvature around the mode of $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$. So (2003) and Jungbacker en Koopman (2007) argue that the mode can be obtained by maximizing $\log p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) = \log p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi}) + \log p(\boldsymbol{\lambda};\boldsymbol{\psi}) - \log p(\boldsymbol{y};\boldsymbol{\psi})$ with respect to $\boldsymbol{\lambda}$. We notice that this conditional mode is different from the conditional mode given in Theorem 2. From Jungbacker en Koopman (2007, Theorem 1) it follows that the mode and curvature can be found by the following Newton-Raphson steps:

(i) Initialize $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$;

(ii) Compute

$$\boldsymbol{x} = \boldsymbol{\lambda}^* - \{\ddot{p}(\boldsymbol{y}|\boldsymbol{\lambda}^*;\boldsymbol{\psi})\}^{-1}\dot{p}(\boldsymbol{y}|\boldsymbol{\lambda}^*;\boldsymbol{\psi}), \qquad \boldsymbol{C} = -\{\ddot{p}(\boldsymbol{y}|\boldsymbol{\lambda}^*;\boldsymbol{\psi})\}^{-1},$$

where

$$\dot{p}(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi}) = \frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})}{\partial \boldsymbol{\lambda}}, \qquad \ddot{p}(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi}) = \frac{\partial^2 \log p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'};$$

(iii) Update $\boldsymbol{\lambda}^*$ by computing $\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{x};\boldsymbol{\psi})$, where $\boldsymbol{x} = \boldsymbol{\lambda} + \boldsymbol{u}$, with $\boldsymbol{u} \sim N(0,\boldsymbol{C})$;

(iv) Iterate between (ii) and (iii) until convergence.

After convergence $\boldsymbol{\lambda}^*$ is equal to the mode of $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ and $\boldsymbol{C}$ is equal to the curvature around the mode. The main difficulty in the implementation of these steps is the computation of the derivatives in step (ii) since $\boldsymbol{\lambda}$ occurs in the log-determinant, the inverse and the predictive mean and variance of the density $p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})$ in (9). Given that the derivatives of the marginal likelihood are equivalent to the derivatives of the expected complete likelihood, Koopman en Shephard (1992) show that the derivatives can be computed in closed form.

Explicit expressions for the dynamic factor model with given loadings are given in Jungbacker en Koopman (2014).

The resulting importance density is chosen as $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \equiv N(\boldsymbol{\lambda}^*, \boldsymbol{C}^*)$, where $\boldsymbol{C}^* = \text{Var}(\boldsymbol{\lambda}) - \text{Var}(\boldsymbol{\lambda})(\text{Var}(\boldsymbol{\lambda}) + \boldsymbol{C})^{-1}\text{Var}(\boldsymbol{\lambda})$ evaluated at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$. The observation model for $\boldsymbol{x}$ in step (iii) can be used for sampling and to compute the weights $w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi})$. We typically ignore the off-diagonal elements in $\boldsymbol{C}$ as they do not affect the quality of the weights. In our Monte Carlo study we show that the weights that result from this choice for the importance density have finite variance for all model specifications that we consider. Given that the model for $\boldsymbol{x}$ is low dimensional and static by construction, sampling from $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$ can be implemented in a computationally efficient way.

By integrating out $\boldsymbol{\alpha}$ analytically via the Kalman filter we reduce the amount of simulation and the simulation variance. This is a direct application of the Rao-Blackwellization principal, which is discussed in general in Doucet *et al.* (2001, pp. 499-515) and Durbin en Koopman (2012, Section 12.7).

3.4 Posterior Statistics for the Factors

Similar results can be obtained for functions of the factors. In particular, define $\boldsymbol{h}(\boldsymbol{\alpha})$ as a vector function of $\boldsymbol{\alpha}$. We are interested in estimating the conditional mean function $\bar{\boldsymbol{h}} = \text{E}(\boldsymbol{h}(\boldsymbol{\alpha})|\boldsymbol{y};\boldsymbol{\psi})$. Similar as for the loadings it holds that

$$\bar{\boldsymbol{h}} = \int_{\boldsymbol{\alpha}} \boldsymbol{h}(\boldsymbol{\alpha}) p(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\psi}) \, \text{d}\boldsymbol{\alpha}, \tag{14}$$

where the marginal conditional density $p(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\psi})$ is unknown when $\boldsymbol{\lambda}$ is stochastic. We do not want to construct an additional importance density for the estimation of the factors for the following two reasons. First, when $\boldsymbol{\alpha}$ constitutes a persistent dynamic process, approximating $p(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\psi})$ accurately in high dimensions is complicated; see for example Koopman *et al.* (2014). Second, given that an accurate importance density is obtained, sampling from it is typically computationally demanding. Therefore we take a different approach and

rewrite the integral in (14) as

$$\begin{aligned}\bar{\boldsymbol{h}} &= \int_{\boldsymbol{\lambda}} \int_{\boldsymbol{\alpha}} \boldsymbol{h}(\boldsymbol{\alpha}) p(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\alpha} \, \mathrm{d}\boldsymbol{\lambda} \\ &= \int_{\boldsymbol{\lambda}} \int_{\boldsymbol{\alpha}} \boldsymbol{h}(\boldsymbol{\alpha}) p(\boldsymbol{\alpha}|\boldsymbol{\lambda}, \boldsymbol{y}; \boldsymbol{\psi}) p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\alpha} \, \mathrm{d}\boldsymbol{\lambda} \\ &= \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) \; p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}. \end{aligned} \tag{15}$$

We have rewritten the conditional mean function in terms of the marginal posterior density of the loadings. For a given value of $\boldsymbol{\lambda}$, we can evaluate $\mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi})$ using the Kalman filter smoother for many functions $h(\boldsymbol{\alpha})$. From Section 3.2 we have learned that analytic expressions for the moments of $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$ do not exist. In Appendix D we develop the Monte Carlo estimate for $\bar{\boldsymbol{h}}$ which is given by

$$\bar{\boldsymbol{h}} = \frac{M^{-1} \sum_{j=1}^{M} \; \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi}) \; w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi})}{M^{-1} \sum_{j=1}^{M} \; w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi})}, \qquad M \to \infty, \tag{16}$$

where the weights $w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi})$ are defined below (13) and the samples $\boldsymbol{\lambda}^{(j)}$ are drawn from $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$. Given that for every draw $\boldsymbol{\lambda}^{(j)}$ we can evaluate $\mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}^{(j)}; \boldsymbol{\psi})$ analytically, we only require the construction of the importance density $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$, as in Section 3.3, for the estimation of $\bar{\boldsymbol{h}}$.

4 Simulation study

In this section we study the finite sample properties of the methods that are discussed in Section 3. Our main interest is to compare the empirical Bayes estimates with the classical maximum likelihood estimates. The empirical Bayes point estimates that we consider are the posterior mode estimates from Section 3.1. The estimates are computed given the estimated parameter vector $\hat{\boldsymbol{\psi}}$ which is obtained as discussed in Section 3.2. The maximum likelihood estimates for the loadings are given by $\tilde{\boldsymbol{\lambda}}$ in (8) and the corresponding smoothed estimates for the factors are given by $\tilde{\boldsymbol{\alpha}}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}; \tilde{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}_1)$; see Jungbacker en Koopman (2014) and Doz *et al.* (2012). We study the differences between the empirical Bayes and maximum likelihood estimates for different data generating processes that are outlined below.

Further we examine the validity of the importance sampling methods that are discussed in Sections 3.3 and 3.4. We estimate the variances of the integrated importance sampling weight function $w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi})$. We apply the extreme value based tests that are discussed in Koopman *et al.* (2009) to determine whether the variance of the weights exists and hence whether a central limit theorem for the importance sampling estimate applies; see Geweke (1989).

4.1 Simulation design

We study the dynamic factor model for different cross-section and time series dimensions. We include combinations for $N, T = 50$ and 100. The number of factors is chosen to be equal to $r = 3$ or $r = 5$. During the simulation study we assume that the true number of factors is known. In empirical applications it is possible to rely on economic theory or information criteria to determine an appropriate number of factors; see Bai en Ng (2002).

We draw $L = 5000$ different panels of observations from model (2) for each combination of panel size and number of factors. We denote the sampled vectorized panels by $\boldsymbol{y}(l)$ for $l = 1, \ldots, L$. Each observation vector has its own "true" loadings and factors; $\boldsymbol{\lambda}(l)$ and $\boldsymbol{\alpha}(l)$. The loading vectors $\boldsymbol{\lambda}_i(l)$ are drawn from a variety of mixture distributions. The distributions are chosen such that they mimic the empirical distribution of the loadings that is found in the macroeconomic application of Stock en Watson (2012) (see Figure 4). In particular, we consider the elements of the loading vector to be given by

$$\lambda_{i,j} = k_1 N(\mu_1, \sigma_1^2) + \ldots + k_s N(\mu_s, \sigma_s^2), \tag{17}$$

where the values for μ_n, σ_n^2, k_n, for $n = 1, \ldots, s$ and s are taken such that the loadings have normal, tri-modal, skewed and outlier distributions. The values for the loading settings are given in Marron en Wand (1992, Table 1). The variety of sampling schemes for the "true" loadings ensures that our results do not depend on the normality assumption for the loadings and that we remain close to empirically relevant distributions. Additionally, we restrict the loading matrix $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1', \boldsymbol{\Lambda}_2')'$, such that the $r \times r$ matrix $\boldsymbol{\Lambda}_1$ is lower triangular with ones on the main diagonal. This identifies a particular rotation for the factors and allows us to calculate mean squared error statistics for the estimates.

The dynamic factors are simulated from the autoregressive process in (2) with autoregressive polynomial matrix $\boldsymbol{H}$ and variance matrix $\boldsymbol{\Sigma}_\eta = \boldsymbol{I}_r - \boldsymbol{H}\boldsymbol{H}'$, such that $\text{Var}(\boldsymbol{\alpha}_t) = \boldsymbol{I}_r$. The elements for the diagonal of $\boldsymbol{H}$ are drawn uniformly for each panel over the range (0.5,0.95). The off-diagonal elements are drawn from $N(0, 0.1)$. The transformations of Ansley en Kohn (1986) are applied to ensure that $\boldsymbol{H}$ and $\boldsymbol{\Sigma}_\eta$ form a stationary vector autoregressive process. For both the maximum likelihood and the empirical Bayes methods the factor process is correctly specified. The error term $\boldsymbol{\epsilon}_t$ is drawn from $N(\boldsymbol{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \ldots, \omega_N^2)$, where the elements ω_i are drawn uniformly over the range (0.1,0.9).

We also study the performance of our methods under some common forms of misspecification with respect to the error term $\boldsymbol{\epsilon}_t$. In particular, we evaluate our methods for cross-sectional dependent errors and serial correlated errors. Cross-section dependence is implemented by generating $\epsilon_{i,t}$ from

$$\epsilon_{i,t} = (1 + b^2)\zeta_{i,t} + b\zeta_{i-1,t} + b\zeta_{i+1,t}, \qquad \zeta_{i,t} \sim N(0, \omega_i^2), \tag{18}$$

where we choose $b = 0.5$. Serial correlation is generated by

$$\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \zeta_{i,t}, \qquad \zeta_{i,t} \sim N(0, \omega_i^2), \tag{19}$$

where the autoregressive coefficients ρ_i are drawn from the uniform distribution between 0.5 and 0.9. We may also combine the equations (18) and (19) to have both serial correlation and cross-section dependence in the errors. Similar sources of misspecification are studied in Stock en Watson (2002b) and Doz *et al.* (2012) for principal components and maximum likelihood methods. We emphasize that we have implemented our methods of Section 3 under the conditions of Assumptions 1.

4.2 Comparing empirical Bayes and maximum likelihood

We study the accuracy of the empirical Bayes estimates and compare these to the maximum likelihood estimates. We define the accuracy of the estimates for the loadings and the factors

by computing the average mean squared error statistics. In particular, we compute

$$\begin{aligned} \text{MSE}(\check{\boldsymbol{\lambda}}) &= L^{-1} \textstyle\sum_{l=1}^{L} (\boldsymbol{\lambda}(l) - \check{\boldsymbol{\lambda}}(l))'(\boldsymbol{\lambda}(l) - \check{\boldsymbol{\lambda}}(l)), \\ \text{MSE}(\check{\boldsymbol{\alpha}}) &= L^{-1} \textstyle\sum_{l=1}^{L} (\boldsymbol{\alpha}(l) - \check{\boldsymbol{\alpha}}(l))'(\boldsymbol{\alpha}(l) - \check{\boldsymbol{\alpha}}(l)), \end{aligned} \tag{20}$$

where the averaging is over the L samples and $\check{\boldsymbol{\lambda}}$ and $\check{\boldsymbol{\alpha}}$ may denote the empirical Bayes estimates ($\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\alpha}}$) or the maximum likelihood estimates ($\tilde{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\alpha}}$). For each data generating process that is discussed in Section 4.1 we present the relative mean squared error statistics $\text{MSE}(\tilde{\boldsymbol{\lambda}})/\text{MSE}(\hat{\boldsymbol{\lambda}})$ and $\text{MSE}(\tilde{\boldsymbol{\alpha}})/\text{MSE}(\hat{\boldsymbol{\alpha}})$.

In Table 1 we present the results from which we deduct the following six conclusions. First, the relative statistics are smaller then one in most cases. It shows that the empirical Bayes estimates are on average more accurate when compared to the maximum likelihood estimates. Second, the relative statistics for the loadings are smaller when compared to the relative statistics for the factors. This is not surprising since the factors are specified the same for both the empirical Bayes and the maximum likelihood methods. Third, the relative statistics depend on the panel dimensions. For panels with $N \geq T$ the relative performance of the empirical Bayes estimates is much better. When $N < T$ the results are mixed. Fourth, the relative performance of the empirical Bayes estimates improves for most cases when the number of factors r increases. Fifth, the results depend on the sampling scheme for the loadings. Perhaps somewhat counter intuitive, but the empirical Bayes estimates improve relative to the maximum likelihood estimates when the sampling scheme for the loadings deviates from the standard normal distribution. Sixth and finally, the different sampling schemes for the error terms do not affect the results. This is not surprising since both the empirical Bayes and the maximum likelihood estimates are affected similarly by misspecification in the error term and the relative statistics.

When we study the results for the loadings more carefully we find that the standard normal specification for the loadings ($L = 1$) is relatively unfavorable for the empirical Bayes estimates. When the standard normal distribution is replaced by the normal distribution with standard deviation 0.2 ($L = 2$) the performance of the empirical Bayes estimates becomes much better. Some further experiments (not shown) have confirmed that the relative

performance of the empirical Bayes estimates increases under a variety of circumstances when the variance of the true loadings decreases. In other words, when the sampling space of the loadings becomes smaller the empirical Bayes estimates show more relative improvement. Also, under multi-modal, skewed and outlier distributions for the true loadings, the empirical Bayes estimates show large improvements over the maximum likelihood estimates. We emphasize that these cases are empirically relevant for macroeconomic applications; see Figure 4. Only when $N < T$ and when the loadings are sampled from the standard normal distribution the maximum likelihood estimates outperform the empirical Bayes estimates. The magnitude of the gain depends on the data generating process. The gains for the empirical Bayes estimates of the loadings can be as large as 82% while the largest loss has only been 7%.

For the factors the results are closer. Nevertheless, only for cases where $N < T$ we sometimes find that the maximum likelihood estimates outperform the empirical Bayes estimates. When $N \geq T$ the relative gains for the empirical Bayes estimates are around 20%. While both the maximum likelihood and the empirical Bayes estimates are based on vector autoregressive process for the factors, the empirical Bayes estimates are on average more accurate.

Overall we may conclude that the empirical Bayes estimates are more accurate when compared to the maximum likelihood estimates for many data generating processes. The magnitude of the gains depend on the sampling scheme for the loadings and the panel dimensions. For all computations in this study, we have written the code in the `Ox` programming language version 7.00 of Doornik (2007).

4.3 Importance sampling weights

The integrated importance sampling weights $w_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \boldsymbol{y}; \boldsymbol{\psi})$, defined below (13), need to have finite variance in order for the conditional mean function estimates in (13) and (16) to have a $\sqrt{M}$ convergence rate; see Geweke (1989). Failure of this condition leads to slow and unstable convergence.

In this section we use the diagnostic tests of Koopman *et al.* (2009) to empirically assess whether the integrated weights have finite variance. For the simulated observation vectors

				MSE($\hat{\boldsymbol{\lambda}}$)/MSE($\tilde{\boldsymbol{\lambda}}$)					MSE($\hat{\boldsymbol{\alpha}}$)/MSE($\tilde{\boldsymbol{\alpha}}$)				
N	T	r	E	L=1	L=2	L=3	L=4	L=5	L=1	L=2	L=3	L=4	L=5
50	50	3	n	0.59	0.35	0.54	0.38	0.36	0.78	0.84	0.84	0.86	0.87
50	100	3	n	0.96	0.48	0.60	0.48	0.42	1.10	0.94	0.98	0.95	0.94
100	50	3	n	0.58	0.32	0.37	0.33	0.29	0.81	0.84	0.83	0.86	0.85
100	100	3	n	0.71	0.39	0.52	0.42	0.34	0.80	0.89	0.89	0.91	0.88
50	50	5	n	0.52	0.39	0.39	0.38	0.38	0.81	0.88	0.86	0.91	0.87
50	100	5	n	0.92	0.43	0.51	0.45	0.44	1.05	0.94	0.96	0.94	0.92
100	50	5	n	0.52	0.28	0.32	0.29	0.29	0.83	0.91	0.85	0.87	0.91
100	100	5	n	0.73	0.33	0.40	0.32	0.31	0.85	0.89	0.89	0.90	0.94
50	50	3	c	0.59	0.38	0.42	0.27	0.33	0.73	0.91	0.81	0.79	0.92
50	100	3	c	1.03	0.46	0.60	0.38	0.40	1.13	1.01	1.03	1.02	1.02
100	50	3	c	0.71	0.37	0.43	0.40	0.29	0.97	0.89	0.89	0.90	0.88
100	100	3	c	0.60	0.39	0.50	0.44	0.40	0.72	0.91	0.83	0.83	0.95
50	50	5	c	0.50	0.24	0.28	0.22	0.26	0.70	0.93	0.80	0.83	0.95
50	100	5	c	0.90	0.29	0.40	0.25	0.33	1.10	1.03	1.02	1.09	1.04
100	50	5	c	0.57	0.21	0.31	0.17	0.23	0.90	0.82	0.87	0.68	0.93
100	100	5	c	0.55	0.26	0.29	0.27	0.23	0.79	0.91	0.79	0.85	0.97
50	50	3	s	0.60	0.43	0.48	0.40	0.35	0.66	0.87	0.78	0.78	0.89
50	100	3	s	1.04	0.52	0.69	0.50	0.43	1.12	0.97	1.04	1.04	0.99
100	50	3	s	0.66	0.37	0.48	0.46	0.35	0.94	0.84	0.80	0.77	0.86
100	100	3	s	0.64	0.47	0.54	0.46	0.43	0.67	0.87	0.77	0.79	0.94
50	50	5	s	0.54	0.29	0.36	0.27	0.26	0.68	0.88	0.77	0.83	0.90
50	100	5	s	1.00	0.35	0.51	0.33	0.35	1.05	1.00	1.00	1.09	1.01
100	50	5	s	0.57	0.27	0.30	0.23	0.28	0.89	0.79	0.85	0.72	0.87
100	100	5	s	0.55	0.22	0.28	0.30	0.31	0.87	0.89	0.83	0.81	0.93
50	50	3	sc	0.54	0.42	0.41	0.33	0.62	0.65	0.89	0.76	0.78	0.80
50	100	3	sc	1.07	0.56	0.59	0.39	0.93	1.09	0.96	0.96	1.00	1.05
100	50	3	sc	0.50	0.33	0.38	0.32	0.56	0.81	0.85	0.76	0.72	0.78
100	100	3	sc	0.57	0.43	0.45	0.39	0.71	0.66	0.88	0.75	0.79	0.82
50	50	5	sc	0.45	0.31	0.27	0.18	0.50	0.65	0.89	0.74	0.79	0.81
50	100	5	sc	0.78	0.37	0.39	0.26	0.81	0.97	0.97	0.93	1.03	1.00
100	50	5	sc	0.98	0.21	0.33	0.18	0.48	0.87	0.81	0.72	0.66	0.80
100	100	5	sc	0.62	0.23	0.25	0.20	0.63	0.83	0.84	0.88	0.82	0.82

Table 1: Simulation results for the empirical Bayes and maximum likelihood estimates. The DGP and parameters are chosen as discussed in Section 4.1. The code L indicates 1; normal(0,1), 2; normal(0,0.04), 3; tri-modal, 4; skewed, or 5; outlier distribution for the true-loadings. The codes n, c, s and sc indicate : n; normal disturbances, s; serial correlated disturbances, c; cross-section correlated disturbances and sc; serial and cross-section correlated disturbances.

$\boldsymbol{y}(1)$ for different panel sizes and number of factors we estimate the parameter vector as discussed in Section 3.2. Next, given the estimated parameter vector we generate $100,000$ importance sampling weights $w_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}^{(j)}, \boldsymbol{y}; \hat{\boldsymbol{\psi}})$ using the importance density $g(\boldsymbol{\lambda}|\boldsymbol{y}; \hat{\boldsymbol{\psi}})$. The choice for the data vector $\boldsymbol{y}(1)$ does not affect the results.

For each set of weights we consider s exceedence sampling weights, denoted by $x_1, \ldots, x_s$, which are larger than some threshold $w^{\min}$ and are assumed to come from the generalized Pareto distribution with logdensity function $f(a, b) = -\log b - (1 + a^{-1}) \log(1 + ab^{-1}x_i)$ for $i = 1, \ldots, s$, where unknown parameters a and b determine the shape and scale of the density, respectively. When $a \leq 0.5$, the variance of the importance sampling weights exists and a $\sqrt{M}$ convergence rate can be assumed. A Wald type test statistic is computed as follows. Estimate a and b by maximum likelihood and denote the estimates by $\tilde{a}$ and $\tilde{b}$, respectively. Compute the t-test statistic $t_w = \tilde{b}^{-1}\sqrt{s\,/\,3}(\tilde{a} - 0.5)$ to test the null hypothesis $H_0 : a = 0.5$. We reject the null hypothesis when the statistic is positive and significantly different from zero, that is, when it is larger than 1.96 with 95% confidence.

We compute the test statistics for different thresholds $w^{\min}$, such that between 1% and 50 % of the largest weights are included. This ensures that we capture sufficiently the tail of the distribution and that the results do not depend on the choice of the threshold. In Figure 1 we present the test statistics for different thresholds and for correctly specified models. The horizontal line at 1.96 indicates the rejection area. For the integrated weights the test statistics are always very negative. This even holds for samples of weights from the end of the tail of the distribution. It provides evidence that the variance in the sampled weights is likely to exist. Hence we may conclude that the constructed importance density $g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})$, from which $\boldsymbol{\alpha}$ is integrated out, can be used to obtain reliable importance sampling estimates. For misspecified models the importance sampling weights are the same since the misspecification affects the original density and the importance density in the same way.

5 Macroeconomic forecasting

Our empirical study is concerned with the comparison of the empirical Bayes and maximum likelihood approaches for dynamic factor models based on a quarterly U.S. macroeconomic dataset. The key question is whether and to what extent the empirical Bayes methods

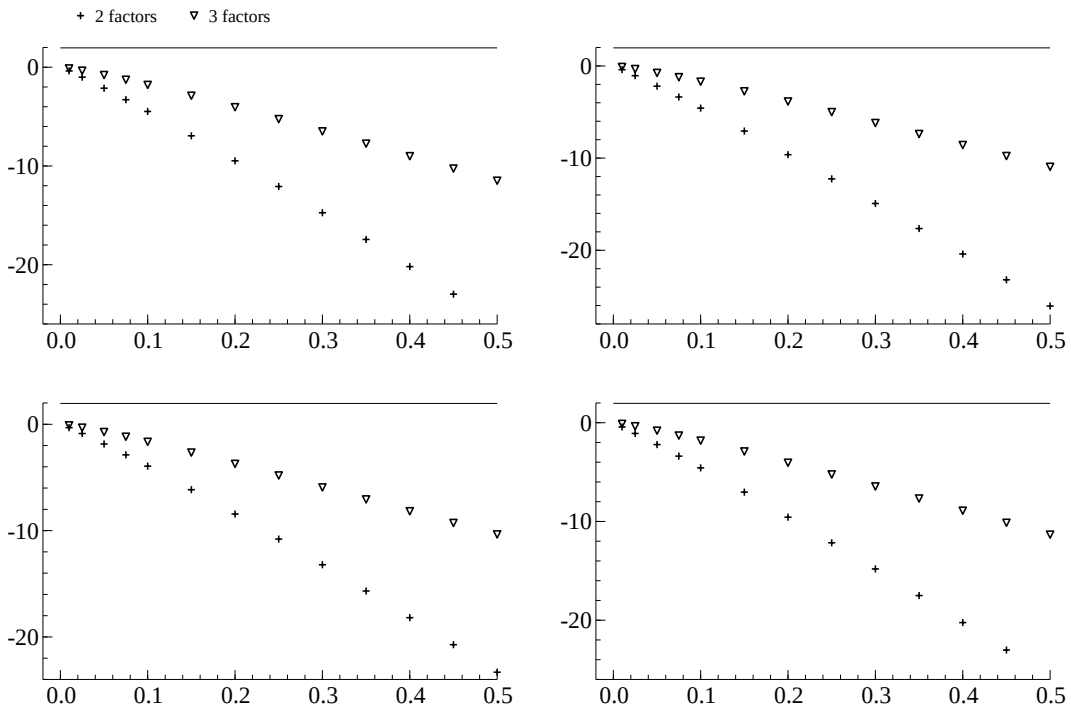

Figure 1: Importance sampling diagnostics for dynamic factor models with $r = 2$ and $r = 3$ factors, based on $100,000$ simulations of weights $w_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}^{(j)}, \boldsymbol{y}; \hat{\boldsymbol{\psi}})$. We computed test statistics for different thresholds $w^{\min}$, by procedures explained in Section 4.3.

improve out-of-sample forecasts when compared to the maximum likelihood methods. We consider the data set of Stock en Watson (2012), which includes $N = 144$ macro economic and financial time series. These series capture a large part of the available disaggregated macroeconomic and financial time series. Table 2 summarizes the categories for which the time series are included. There are 13 different blocks including large blocks for employment series and prices. Each block is indexed by a letter. From this data set we construct stationary quarterly time series following the guidelines in the web appendix of Stock en Watson (2012). The resulting panel ranges from 1959-1 until 2008-4, with $T = 200$. The panel is unbalanced and the estimation methods of Section 3 are adjusted accordingly as discussed in Appendix E.

We consider the dynamic factor model with stochastic loadings in (2) with the conditions of Assumption 1. Additionally, we partially restrict the loading matrix $\mathbf{\Lambda} = (\mathbf{\Lambda}_1', \mathbf{\Lambda}_2')'$, where the $r \times r$ upper block matrix $\mathbf{\Lambda}_1$ is lower triangular. The loadings and factors are identified up to a sign change and this is sufficient for forecasting applications. We have taken this particular rotation for the factors to facilitate the comparison of mean squared error statistics for the different estimates. The same restrictions are imposed on the dynamic factor model with deterministic loadings.

We consider the dynamic factor model with $r = 5$ factors and a diagonal variance matrix for the disturbances. Similar results can be obtained for models with $r = 6$ factors; see also Stock en Watson (2012). We first discuss the full sample parameter estimation results after which we discuss the forecasting exercise and its results.

5.1 Estimation results

In Table 3 we present the parameter estimates together with the distributions of the loadings and the factors. The methods of Section 3 are used to obtain the estimates. In the top panel we show the estimates for the autoregressive matrix $\boldsymbol{H}$. From the eigenvalues it follows that all the factors are estimated as stationary processes. The factors are persistent since the largest eigenvalue is 0.822. Additionally, we find some evidence for a persistent cyclical behavior in the factors because one conjugate pair of complex eigenvalues is obtained with its real part equal to 0.809. The remaining two eigenvalues are relatively small. Since the

	Category	number of series (144)
A	GDP components	16
B	Industrial production	14
C	Employment	20
D	Unemployment rate	7
E	Housing starts	6
F	Inventories	6
G	Prices	37
H	Wages	6
I	Interest rates	13
J	Money	8
K	Exchange rates	5
L	Stock prices	5
M	Consumer expectations	1

Table 2: Summary of the time series that are included in the empirical application

VAR process of the factors impose a zero mean and an identity variance matrix, we can relate the individual coefficients in $\boldsymbol{H}$ to each other. However, as for any VAR analysis, individual coefficients in $\boldsymbol{H}$ do not have a clear interpretation.

In the bottom panel of Table 3 we show the estimates for the mean and variance of the loading matrix. The mean coefficients are small and centered around zero. The diagonal elements of the variance matrix indicate that the variance in the loadings is small. This indicates that relatively large gains in accuracy are likely for the empirical Bayes estimates as Hansen (2013) shows that shrinkage estimates gain most when the estimates are close to the restriction imposed; see also Section 4.

In Figures 2 and 3 we show the estimates for the loadings and the factors, both for the empirical Bayes and maximum likelihood estimation methods. The differences between the two are small and hard to visualize. The real economic time series, such as the industrial production and employment time series, load mainly on the first and third factors. The first factor is noisy, whereas the third factor displays some slowly varying business cycle indicator around a long term declining trend. The fourth factor evolves around zero but becomes negative during recession times.

In Figure 4 we show the empirical distributions of the demeaned estimates for the columns of the loading matrix. The first loading column displays multiple modes, whereas the second is slightly skewed to the left. Their empirical distributions are similar to the normal distribu-

	Autoregressive coefficients $\boldsymbol{H}$					Eigenvalues	
#	1	2	3	4	5	real	img
1	0.308	-0.278	-0.055	-0.160	0.268	0.822	0.000
2	-0.191	0.355	0.266	-0.029	-0.029	0.809	0.174
3	0.371	-0.041	0.822	-0.074	0.119	0.809	-0.174
4	-0.148	0.069	-0.183	0.299	0.520	0.029	0.000
5	0.125	-0.157	0.053	0.427	0.557	-0.129	0.000

	Distribution loadings $\boldsymbol{\Sigma}_\lambda$					$\boldsymbol{\delta}$
#	1	2	3	4	5	
1	0.079	-0.021	0.028	-0.002	0.001	0.229
2	-0.020	0.050	-0.006	0.002	-0.007	-0.070
3	0.028	-0.006	-0.006	-0.006	0.005	0.153
4	-0.002	0.002	-0.007	0.032	-0.009	0.033
5	0.001	-0.007	0.005	-0.009	0.035	0.019

Table 3: Parameter estimates for the vector autoregressive coefficients and the mean and variance of the loadings. The full sample of observations is used ($N = 144$ and $T = 200$) and we estimate the model parameters using the methods developed in Section 3. The columns indicated by Eigen summarize the real and imaginary eigenvalues of the matrix $\boldsymbol{H}$ in decreasing order.

tion, with standard deviations of approximately 0.25. This also holds for the third column of the loading matrix. The distributions of the fourth and fifth columns of the loading matrix are centered around zero but show outliers. The tails of the distributions are heavier when compared to the normal distribution. In the lower right of Figure 4 we show the empirical distribution for all loadings. This distribution is skewed to the left and has heavy tails.

5.2 Forecasting study

The out-of-sample forecasting study for the panel of macroeconomic time series is designed as follows. We forecast each time series 1, 2 and 4 quarters ahead for 1985-1 until 2008-4. In total we compute $m = 96$ out-of sample predictions for each horizon. In particular, let the integer n denote the sample split point (1984-4 for $h = 1$, 1984-3 for $h = 2$ and 1984-1 for $h = 4$). The forecasts are computed for $n + 1, \ldots, n + m$ based on subsamples of the observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n+j-h}$, for $j = 1, \ldots, m$. We estimate the parameter vector $\boldsymbol{\psi}$ for each subsample using the methods of Section 3. Based on the estimated parameter vector we

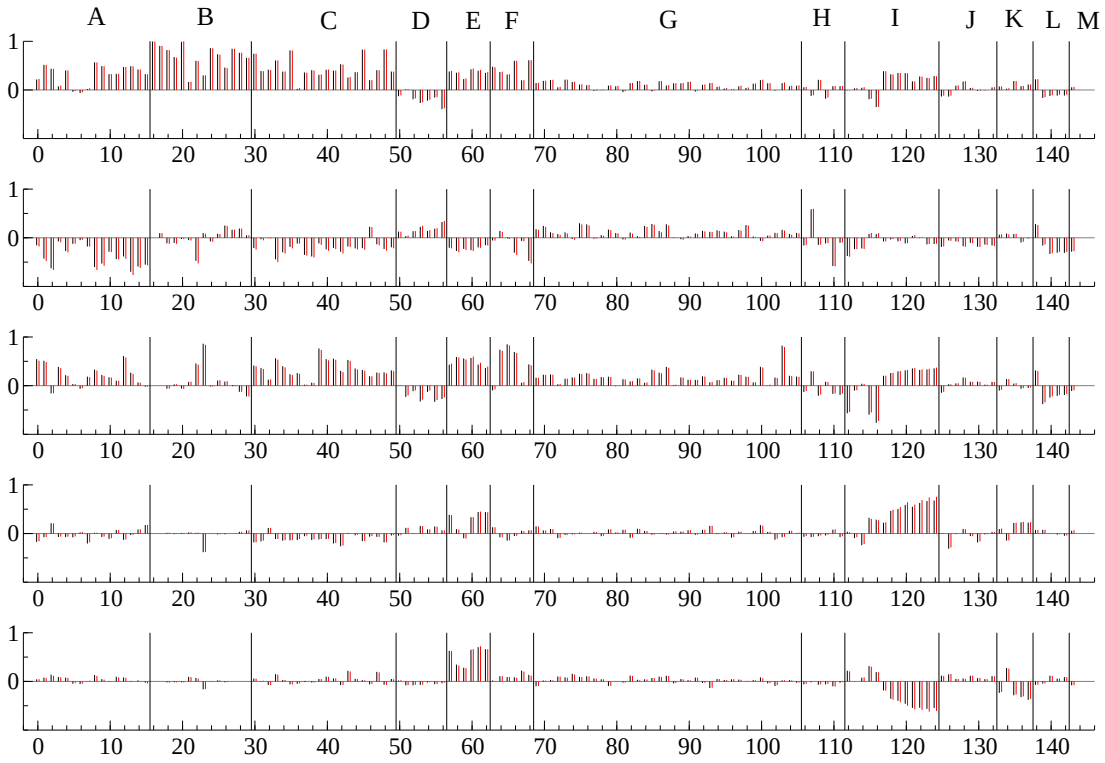

Figure 2: Posterior mode empirical Bayes estimates and maximum likelihood estimates for the loadings. The left bars pertain to the empirical Bayes estimates and the right bars pertain to the maximum likelihood esitimates.)

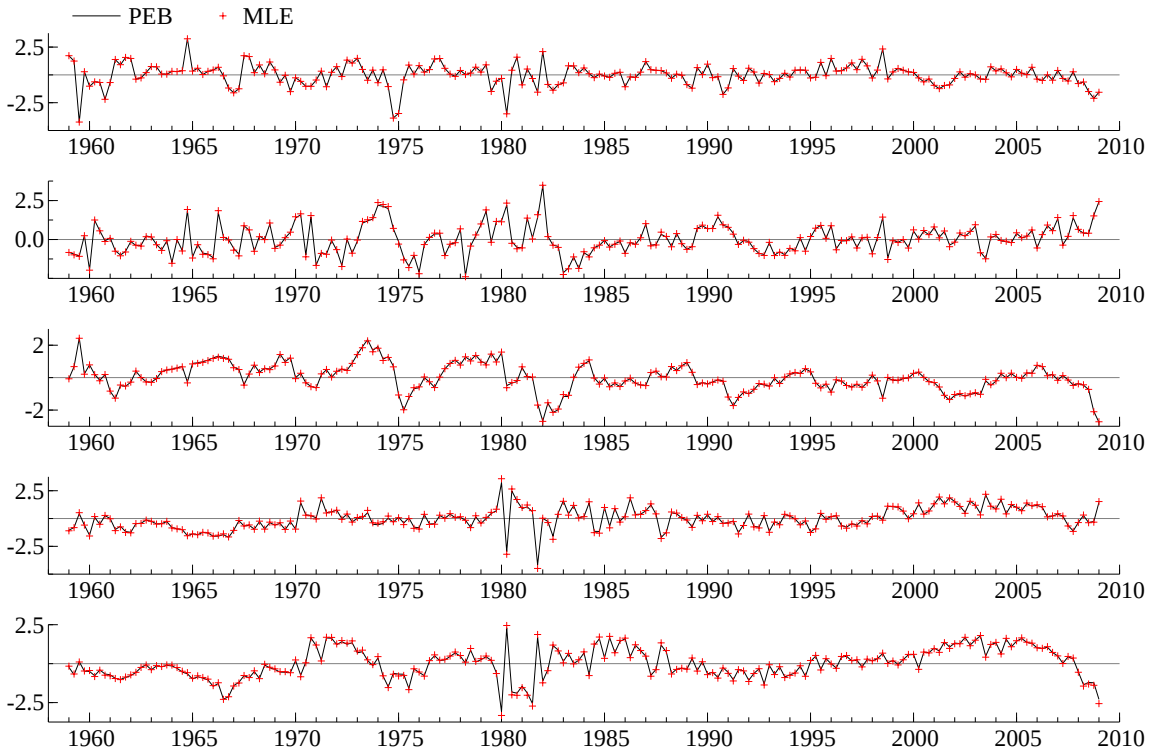

Figure 3: Posterior mode empirical Bayes estimates and maximum likelihood estimates for the factors.

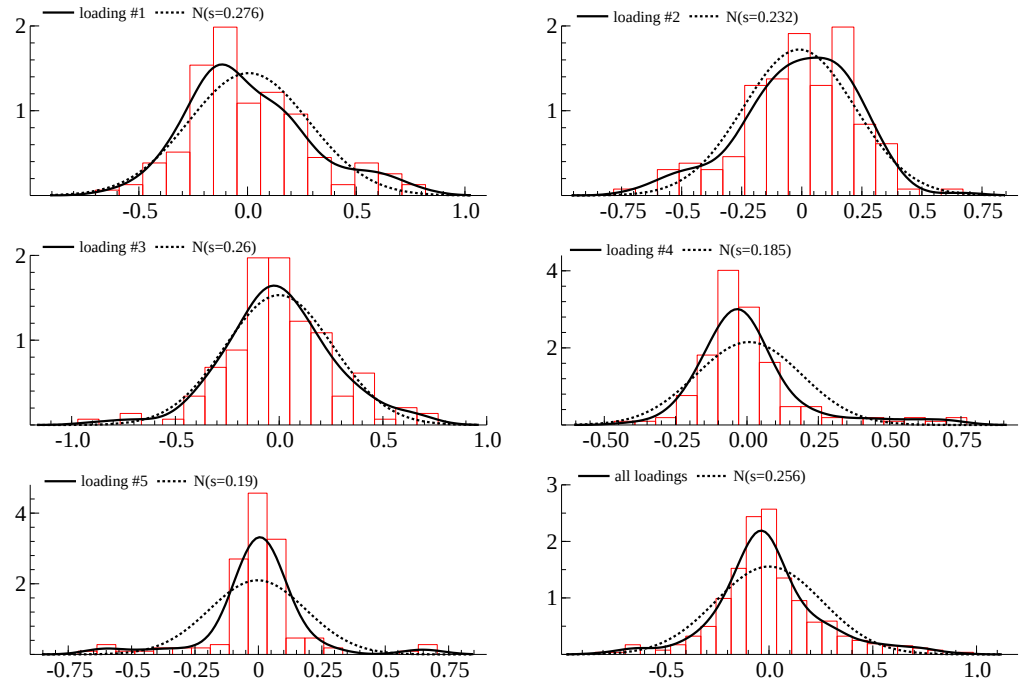

Figure 4: Empirical distribution for the posterior modes of the loadings. We present the results per factor and overall. The dotted line indicates the normal approximation.

compute the empirical Bayes posterior mode and maximum likelihood forecasts by

$$\hat{\boldsymbol{y}}_{n+j} = \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\alpha}}_{n+j} \quad \text{and} \quad \tilde{\boldsymbol{y}}_{n+j} = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\alpha}}_{n+j}, \tag{21}$$

where $\hat{\boldsymbol{\alpha}}_{n+j} = \mathrm{E}(\boldsymbol{\alpha}_{n+j}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n+j-h}; \hat{\boldsymbol{\Lambda}}; \hat{\boldsymbol{\psi}})$ is the posterior mode forecast for the factors and $\tilde{\boldsymbol{\alpha}}_{n+j} = \mathrm{E}(\boldsymbol{\alpha}_{n+j}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n+j-h}; \tilde{\boldsymbol{\Lambda}}; \hat{\boldsymbol{\psi}}_1)$ is the forecast for the factors based on the maximum likelihood estimates. These forecasts are computed for all horizons $h = 1, 2, 4$.

As a measure of accuracy we consider the mean squared error (MSE) of the out-of-sample forecasts. In particular we compute for each time series

$$\mathrm{MSE}_i^{\mathrm{PEB}} = m^{-1}\sum_{j=1}^{m}(y_{i,n+j} - \hat{y}_{i,n+j})^2 \quad \text{and} \quad \mathrm{MSE}_i^{\mathrm{MLE}} = m^{-1}\sum_{j=1}^{m}(y_{i,n+j} - \tilde{y}_{i,n+j})^2, \tag{22}$$

where $\hat{y}_{i,n+j}$ and $\tilde{y}_{i,n+j}$ are the elements of $\hat{\boldsymbol{y}}_{n+j}$ and $\tilde{\boldsymbol{y}}_{n+j}$. In this way we compute mean squared error statistics for 144 time series for all forecasting horizons.

5.3 Forecasting results

In Table 4 we present summary statistics for the relative mean squared error statistics; $\mathrm{MSE}_i^{\mathrm{PEB}}/\mathrm{MSE}_i^{\mathrm{MLE}}$. In the first row the average statistics over all series are displayed. For the one quarter ahead forecasts, the empirical Bayes estimates are almost 10% more accurate, this decreases to 8% for the two quarters ahead forecasts and to 4% for the four quarters ahead forecasts. We have obtained substantial forecasting gains for shorter forecast horizons. When we increase the number of factors, the relative gains become larger (not shown here). The relative improvement in accuracy of empirical Bayes forecasts declines as we forecast further into the future. This is not surprising since both the empirical Bayes and maximum likelihood estimates are based on the same vector autoregressive process for the factors. The out-of sample forecasting variance for a larger forecast horizon is dominated by the contribution of the factors given that the loadings are time-invariant.

We further present a selection of quantiles of the distributions. For the 0.05 quantile, we find gains between 50% and 27% depending on the forecast horizon. This indicates that for a modest number of time series the gains are very large. On the other side of the distribution, the relative accuracy is somewhat more in favor of the maximum likelihood

estimates. However, the differences are as small as between 6% and 4%.

Finally, we summarize the relative mean squared error statistics per category. We find that the largest gains are obtained for the real economic categories such as Housing (this includes series related to housing starts), Inventories and Industrial Production. For the category of Wages and Prices we do not obtain much improvements for the empirical Bayes methods. We notice that the largest gains are found for the time series that load strongly on the factors.

	$h = 1$	$h = 2$	$h = 4$
All series			
Mean	0.904	0.928	0.965
Quantiles			
0.05	0.498	0.586	0.736
0.25	0.849	0.895	0.959
0.50	0.969	0.985	0.995
0.75	1.004	1.008	1.006
0.95	1.061	1.050	1.044
Components (*Mean*)			
GDP components	0.855	0.898	0.971
Industrial Production	0.838	0.899	0.943
Employment	0.888	0.923	0.962
Unemployment rate	0.891	0.920	0.970
Housing	0.581	0.647	0.799
Inventories	0.715	0.867	1.002
Prices	0.996	0.997	1.000
Wages	0.969	0.947	0.981
Interest rates	0.796	0.800	0.859
Money	0.914	0.949	0.974
Exchanges rates	0.991	0.982	1.001
Stock prices	0.993	1.003	0.999
Consumer Expectations	0.999	1.005	1.009

Table 4: Relative mean squared error statistics for out-of-sample forecasting using empirical Bayes and maximum likelihood methods. The results summarize the distribution of the statistics $\text{MSE}_i^{\text{PEB}}/\text{MSE}_i^{\text{MLE}}$, for $i = 1, \ldots, 144$ and forecast horizons $h = 1, 2, 4$.

6 Conclusion

We have developed parametric empirical Bayes methods for the estimation of the dynamic factor model. The loadings, factors and disturbances of the model are treated as latent stochastic variables, which follow Gaussian distributions. For the estimation of the loadings and factors we have developed a posterior mode algorithm which relies on standard methods for linear time series and regression models. The parameter vector is estimated by likelihood based methods using a simple two-step implementation. The posterior means and variances of the loadings are estimated by standard simulations methods, where we circumvent the infinite variance problem by calculating the integral over the factors analytically. We emphasize that the computational effort for the empirical Bayes methods is only modestly larger when compared to standard maximum likelihood methods; see Jungbacker en Koopman (2014) and Doz *et al.* (2012). The methods are evaluated in a Monte Carlo study for dynamic factor models with different dimensions and different numbers of factors. The empirical Bayes estimates for both the loadings and the factors outperform the maximum likelihood estimates for the data generating processes considered. This holds even when the data generating process for the loadings is highly non-Gaussian. We have further illustrated our methods in an empirical application for forecasting macroeconomic time series. The empirical Bayes approach again dominates the maximum likelihood approach.

Appendix A

In order to provide parsimonious proofs for the results of this paper we first write model (2) in matrix notation. Details for this can be found in Durbin en Koopman (2012, Section 4.13).

The observation equation for model (2) can be written as

$$\boldsymbol{y} = \boldsymbol{\Lambda}^* \boldsymbol{\alpha} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}^*),$$

or, alternatively,

$$\boldsymbol{y} = \boldsymbol{A}^* \boldsymbol{\lambda} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}^*),$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_T', \boldsymbol{\alpha}_{T+1}')'$, $\boldsymbol{\lambda} = (\acute{\boldsymbol{\lambda}}_1', \ldots, \acute{\boldsymbol{\lambda}}_r')$, where $\acute{\boldsymbol{\lambda}}_j$ denotes the jth column of $\boldsymbol{\Lambda}$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_T')'$. Further

$$\boldsymbol{\Lambda}^* = \begin{bmatrix} \boldsymbol{\Lambda} & & \boldsymbol{0} & \boldsymbol{0} \\ & \ddots & & \vdots \\ \boldsymbol{0} & & \boldsymbol{\Lambda} & \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{A}^* = \begin{bmatrix} (\boldsymbol{\alpha}_1' \otimes \boldsymbol{I}_N) \\ \vdots \\ (\boldsymbol{\alpha}_T' \otimes \boldsymbol{I}_N) \end{bmatrix},$$

$$\boldsymbol{\Omega}^* = \begin{bmatrix} \boldsymbol{\Omega} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{\Omega} \end{bmatrix}.$$

The state equation takes the form

$$\boldsymbol{\alpha} = \boldsymbol{H}^* (\boldsymbol{\alpha}_1^* + \boldsymbol{\eta}), \qquad \boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta^*),$$

with $\boldsymbol{\alpha}_1^* = (\boldsymbol{\alpha}_1', \mathbf{0}, \ldots, \mathbf{0})'$, $\boldsymbol{\eta} = (\mathbf{0}, \boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_T')'$ and

$$\boldsymbol{H}^* = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \boldsymbol{H} & \boldsymbol{I} & \mathbf{0} & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \boldsymbol{H}^2 & \boldsymbol{H}^2 & \boldsymbol{I} & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \boldsymbol{H}^3 & \boldsymbol{H}^2 & \boldsymbol{H} & \boldsymbol{I} & & \mathbf{0} & \mathbf{0} \\ & & & & \ddots & & \vdots \\ \boldsymbol{H}^{T-1} & \boldsymbol{H}^{T-2} & \boldsymbol{H}^{T-3} & \boldsymbol{H}^{T-4} & & \boldsymbol{I} & \mathbf{0} \\ \boldsymbol{H}^T & \boldsymbol{H}^{T-1} & \boldsymbol{H}^{T-2} & \boldsymbol{H}^{T-3} & \ldots & \boldsymbol{H} & \boldsymbol{I} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_\eta^* = \begin{bmatrix} \boldsymbol{\Sigma}_\eta & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{\Sigma}_\eta \end{bmatrix}.$$

It holds that

$$\boldsymbol{\alpha} \sim N(\boldsymbol{H}^* \boldsymbol{a}_1^*, \boldsymbol{H}^* (\boldsymbol{P}_1^* + \boldsymbol{\Sigma}_\eta^*) \boldsymbol{H}^{*'}),$$

where

$$\boldsymbol{a}_1^* = \begin{bmatrix} \boldsymbol{a}_1 \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \qquad \boldsymbol{P}_1^* = \begin{bmatrix} \boldsymbol{P}_1 & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \vdots & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{bmatrix}.$$

For the loadings it holds that

$$\boldsymbol{\lambda} \sim N\left((\boldsymbol{\delta} \otimes \boldsymbol{\iota}_N), (\boldsymbol{\Sigma}_\lambda \otimes \boldsymbol{I}_N)\right).$$

Now we are ready to proceed with the proof of Theorem 1. We suppress the dependence on the parameter vector $\boldsymbol{\psi}$ for notational convenience. All densities can be considered given $\boldsymbol{\psi}$. For model (2) under **Assumption 1** it follows from Bayes rule that

$$\log p(\boldsymbol{\lambda}, \boldsymbol{\alpha} | \boldsymbol{y}) = \log p(\boldsymbol{y} | \boldsymbol{\lambda}, \boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha}) + \log p(\boldsymbol{\lambda}) - \log(\boldsymbol{y}),$$

where assumption (C) implies $p(\boldsymbol{y} | \boldsymbol{\lambda}, \boldsymbol{\alpha}) \equiv N(\boldsymbol{\Lambda}^* \boldsymbol{\alpha}, \boldsymbol{\Omega}^*)$ and the densities for $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\lambda})$

are given above. When we consider only the terms that depend on $\boldsymbol{\alpha}$ or $\boldsymbol{\lambda}$ we obtain

$$\begin{aligned}
\log p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}) &\propto \log p(\boldsymbol{y}|\boldsymbol{\lambda}, \boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha}) + \log p(\boldsymbol{\lambda}) \\
&\propto -\frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{\Lambda}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\Lambda}^*\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\Lambda}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y} \\
&\quad -\frac{1}{2}\boldsymbol{\alpha}'(\boldsymbol{H}^*(\boldsymbol{P}_1^* + \boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha} + \boldsymbol{a}_1^{*'}\boldsymbol{H}^{*'}(\boldsymbol{H}^*(\boldsymbol{P}_1^* + \boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha} \\
&\quad -\frac{1}{2}\boldsymbol{\lambda}'(\boldsymbol{\Sigma}_\lambda \otimes \boldsymbol{I}_N)^{-1}\boldsymbol{\lambda} + (\boldsymbol{\delta} \otimes \boldsymbol{\iota}_N)'(\boldsymbol{\Sigma}_\lambda \otimes \boldsymbol{I}_N)^{-1}\boldsymbol{\lambda},
\end{aligned}$$

which can be alternatively written as

$$\begin{aligned}
\log p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}) &\propto \log p(\boldsymbol{y}|\boldsymbol{\lambda}, \boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha}) + \log p(\boldsymbol{\lambda}) \\
&\propto -\frac{1}{2}\boldsymbol{\lambda}'\boldsymbol{A}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{A}^*\boldsymbol{\lambda} + \boldsymbol{\lambda}'\boldsymbol{A}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y} \\
&\quad -\frac{1}{2}\boldsymbol{\alpha}'(\boldsymbol{H}^*(\boldsymbol{P}_1^* + \boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha} + \boldsymbol{a}_1^{*'}\boldsymbol{H}^{*'}(\boldsymbol{H}^*(\boldsymbol{P}_1^* + \boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha} \\
&\quad -\frac{1}{2}\boldsymbol{\lambda}'(\boldsymbol{\Sigma}_\lambda \otimes \boldsymbol{I}_N)^{-1}\boldsymbol{\lambda} + (\boldsymbol{\delta} \otimes \boldsymbol{\iota}_N)'(\boldsymbol{\Sigma}_\lambda \otimes \boldsymbol{I}_N)^{-1}\boldsymbol{\lambda}.
\end{aligned}$$

Next we calculate the first order conditions for $p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y})$. Using the first representation for $p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y})$ we find that

$$\begin{aligned}
\frac{\partial \log p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y})}{\partial \boldsymbol{\alpha}} &= -\boldsymbol{\Lambda}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\Lambda}^*\boldsymbol{\alpha} + \boldsymbol{\Lambda}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y} \\
&\quad -(\boldsymbol{H}^*(\boldsymbol{P}_1^* + \boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha} + (\boldsymbol{H}^*(\boldsymbol{P}_1^* + \boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{H}^*\boldsymbol{a}_1^*.
\end{aligned}$$

When using the second representation for $p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y})$ we find that

$$\begin{aligned}
\frac{\partial \log p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y})}{\partial \boldsymbol{\lambda}} &= -\boldsymbol{A}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{A}^*\boldsymbol{\lambda} + \boldsymbol{A}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y} \\
&\quad -(\boldsymbol{\Sigma}_\lambda \otimes \boldsymbol{I}_N)^{-1}\boldsymbol{\lambda} + (\boldsymbol{\Sigma}_\lambda \otimes \boldsymbol{I}_N)^{-1}(\boldsymbol{\delta} \otimes \boldsymbol{\iota}_N).
\end{aligned}$$

Next, we show that the first order conditions for $\log p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}})p(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}})$ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ are the same when both are evaluated at $\tilde{\boldsymbol{\alpha}}$ and $\tilde{\boldsymbol{\lambda}}$. It holds that $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}})$

is independent of $\boldsymbol{\alpha}$ and that $p(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}})$ is independent of $\boldsymbol{\lambda}$. Thus,

$$\begin{aligned}\log p(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}}) &\propto \log p(\boldsymbol{y}|\boldsymbol{\alpha};\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}})+\log p(\boldsymbol{\alpha})\\ &\propto -\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y}-\frac{1}{2}\boldsymbol{\alpha}'\tilde{\boldsymbol{\Lambda}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\tilde{\boldsymbol{\Lambda}}^*\boldsymbol{\alpha}+\boldsymbol{\alpha}'\tilde{\boldsymbol{\Lambda}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y}\\ &\quad -\frac{1}{2}\boldsymbol{\alpha}'(\boldsymbol{H}^*(\boldsymbol{P}_1^*+\boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha}+\boldsymbol{a}_1^{*'}\boldsymbol{H}^{*'}(\boldsymbol{H}^*(\boldsymbol{P}_1^*+\boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}})}{\partial\boldsymbol{\alpha}} &= -\tilde{\boldsymbol{\Lambda}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\tilde{\boldsymbol{\Lambda}}^*\boldsymbol{\alpha}+\tilde{\boldsymbol{\Lambda}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y}\\ &\quad -(\boldsymbol{H}^*(\boldsymbol{P}_1^*+\boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{\alpha}+(\boldsymbol{H}^*(\boldsymbol{P}_1^*+\boldsymbol{\Sigma}_\eta^*)\boldsymbol{H}^{*'})^{-1}\boldsymbol{H}^*\boldsymbol{a}_1^*.\end{aligned}$$

Similarly, for the loadings it holds that

$$\begin{aligned}\log p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}) &\propto \log p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}})+\log p(\boldsymbol{\lambda})\\ &\propto -\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y}-\frac{1}{2}\boldsymbol{\lambda}'\tilde{\boldsymbol{A}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\tilde{\boldsymbol{A}}^*\boldsymbol{\lambda}+\boldsymbol{\lambda}'\tilde{\boldsymbol{A}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y}\\ &\quad -\frac{1}{2}\boldsymbol{\lambda}'(\boldsymbol{\Sigma}_\lambda\otimes\boldsymbol{I}_N)^{-1}\boldsymbol{\lambda}+(\boldsymbol{\delta}\otimes\boldsymbol{\iota}_N)'(\boldsymbol{\Sigma}_\lambda\otimes\boldsymbol{I}_N)^{-1}\boldsymbol{\lambda},\end{aligned}$$

where the first order condition is given by

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}})}{\partial\boldsymbol{\lambda}} &= -\tilde{\boldsymbol{A}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\tilde{\boldsymbol{A}}^*\boldsymbol{\lambda}+\tilde{\boldsymbol{A}}^{*'}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y}\\ &\quad -(\boldsymbol{\Sigma}_\lambda\otimes\boldsymbol{I}_N)^{-1}\boldsymbol{\lambda}+(\boldsymbol{\Sigma}_\lambda\otimes\boldsymbol{I}_N)^{-1}(\boldsymbol{\delta}\otimes\boldsymbol{\iota}_N).\end{aligned}$$

It is easy to see that the first order conditions for both $p(\boldsymbol{\lambda},\boldsymbol{\alpha}|\boldsymbol{y})$ and $p(\boldsymbol{\lambda}|\boldsymbol{y};\tilde{\boldsymbol{\alpha}})p(\boldsymbol{\alpha}|\boldsymbol{y};\tilde{\boldsymbol{\lambda}})$ are the same when $\boldsymbol{\lambda}$ is evaluated at $\tilde{\boldsymbol{\lambda}}$ and $\boldsymbol{\alpha}$ is evaluated at $\tilde{\boldsymbol{\alpha}}$.

Appendix B

In order to proof Theorem 2 we check whether the general conditions of Meng en Rubin (1993) hold. This amounts to proving that 1. the restrictions that we iteratively impose on the vector $\boldsymbol{z}=(\boldsymbol{\lambda}',\boldsymbol{\alpha}')'$ in order to maximize $p(\boldsymbol{\lambda},\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\psi})$ are "space filling" and 2. that each iteration in Theorem 2 leads to a unique maximum. Under these conditions, the regularity

conditions in Wu (1983) (equations 5-9), and the assumption that $p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\psi})$ is uni-modal the iterations in Theorem 2 converge and lead to a unique maximum, which is an immediate consequence of Theorem 3 and Corollary 3 in Meng en Rubin (1993)

Given $\boldsymbol{z}^{(s-1)} = (\boldsymbol{\lambda}^{(s-1)'}, \boldsymbol{\alpha}^{(s-1)'})'$, the optimization problem in iteration (s) for step (i) is given by

$$\max_{\boldsymbol{\alpha}} \log p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\psi}) \quad \text{given constraint} \quad g_{(i)}(\boldsymbol{z}) = g_{(i)}(\boldsymbol{z}^{(s-1)}),$$

where $g_{(i)}(\boldsymbol{z}) = \boldsymbol{\lambda}$. We can denote the output from this first step by $\boldsymbol{z}^{(s-\frac{1}{2})} = (\boldsymbol{\lambda}^{(s-1)'}, \boldsymbol{\alpha}^{s'})'$. For step (ii) the problem is given by

$$\max_{\boldsymbol{\lambda}} \log p(\boldsymbol{\lambda}, \boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\psi}) \quad \text{given constraint} \quad g_{(ii)}(\boldsymbol{z}) = g_{(ii)}(\boldsymbol{z}^{(s-\frac{1}{2})}),$$

where $g_{(ii)}(\boldsymbol{z}) = \boldsymbol{\alpha}$. We denote the output corresponding by $\boldsymbol{z}^{(s)} = (\boldsymbol{\lambda}^{(s)'}, \boldsymbol{\alpha}^{s'})'$.

It follows

$$\log p(\boldsymbol{z}^{(s)}|\boldsymbol{y}; \boldsymbol{\psi}) \geq \log p(\boldsymbol{z}^{(s-\frac{1}{2})}|\boldsymbol{y}; \boldsymbol{\psi}) \geq \log p(\boldsymbol{z}^{(s-1)}|\boldsymbol{y}; \boldsymbol{\psi})$$

If the sequence $p(\boldsymbol{z}^{(s)}|\boldsymbol{y}; \boldsymbol{\psi})$ is bounded from above then it converges monotonically to some value, say p^*. The constraints are space filling whenever

$$J(\boldsymbol{z}) = J_{(i)}(\boldsymbol{z}) \cap J_{(ii)}(\boldsymbol{z}) = \{0\},$$

where $J_{(i)}(\boldsymbol{z})$ and $J_{(ii)}(\boldsymbol{z})$ are the column spaces of the scores,

$$J_{(i)}(\boldsymbol{z}) = \{\frac{\partial g_{(i)}(\boldsymbol{z})}{\partial \boldsymbol{z}}\boldsymbol{\gamma} = (\boldsymbol{\iota}'_{Nr}, \boldsymbol{0}'_{Tr})'\boldsymbol{\gamma};\ \gamma \in \boldsymbol{R}^{(N+T)r}\}$$

and

$$J_{(ii)}(\boldsymbol{z}) = \{\frac{\partial g_{(i)}(\boldsymbol{z})}{\partial \boldsymbol{z}}\boldsymbol{\gamma} = (\boldsymbol{0}'_{Nr}, \boldsymbol{\iota}'_{Tr})'\boldsymbol{\gamma};\ \gamma \in \boldsymbol{R}^{(N+T)r}\}.$$

Since $J_{(i)}(\boldsymbol{z})$ is orthogonal to $J_{(ii)}(\boldsymbol{z})$ for all $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ it follows that our constraints are space filling. Given either $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s-1)}$, or $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s)}$ the model (2) is equal to a linear Gaussian model that is identified by under **Assumption 1**. It follows from the equality of the mean and the mode for Gaussian models that the expectations $\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)}, \boldsymbol{\psi})$ and $\mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)}, \boldsymbol{\psi})$ are the unique maximizers of the conditional maximization steps (i)

and (ii).

Next, we discuss the implementation details for the fast computation of $\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)},\boldsymbol{\psi})$ and $\mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)},\boldsymbol{\psi})$. Given $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)}$ model (2) is a linear Gaussian state space model. Jungbacker en Koopman (2014) show that $\mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)},\boldsymbol{\psi}) = \mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y}^L;\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)},\boldsymbol{\psi})$, where $\boldsymbol{y}^L = (y_1^{L\prime},\ldots,y_T^{L\prime})'$, with

$$\boldsymbol{y}_t^L = \boldsymbol{C}^{(s-1)\prime}\boldsymbol{\Lambda}^{(s-1)\prime}\boldsymbol{\Omega}^{-1}\boldsymbol{y}_t, \qquad \text{with} \quad \boldsymbol{C}^{(s-1)}\boldsymbol{C}^{(s-1)\prime} = (\boldsymbol{\Lambda}^{(s-1)\prime}\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}^{(s-1)})^{-1},$$

where $\boldsymbol{C}^{(s-1)}$ is lower triangular. The model for the transformed $r \times 1$ observation vector $\boldsymbol{y}_t^L$ is given by

$$\begin{aligned} \boldsymbol{y}_t^L &= (\boldsymbol{C}^{(s-1)})^{-1}\boldsymbol{\alpha}_t + \boldsymbol{e}_t, & \boldsymbol{e}_t &\sim NID(\boldsymbol{0},\boldsymbol{I}_r), \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{H}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim NID(\boldsymbol{0},\boldsymbol{\Sigma}_\eta), \qquad t = 1,\ldots,T. \end{aligned}$$

The Kalman filter smoother can be applied to the model for $\boldsymbol{y}_t^L$ in order to compute $\mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)},\boldsymbol{\psi})$. The transformation step collapses the large cross-section of the original model and speeds up the evaluation of $\mathrm{E}(\boldsymbol{\alpha}|\boldsymbol{y};\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)},\boldsymbol{\psi})$ by a factor 10; see Jungbacker en Koopman (2014) for additional details.

Given $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)}$ model (2) is a multivariate Gaussian regression model. We define the $Nr \times 1$ dimensional vector $\bar{\boldsymbol{y}}^L = (A^{(s)*\prime}(\boldsymbol{\Omega}^*)^{-1}A^{(s)*})^{-1}A^{(s)*\prime}(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{y}$, where $\boldsymbol{\Omega}^*$ and $A^{(s)*}$ are given in Appendix A, with α replaced by $\boldsymbol{\alpha}^{(s)}$. Mesters en Koopman (2014) sow that $\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)},\boldsymbol{\psi}) = \mathrm{E}(\boldsymbol{\lambda}|\bar{\boldsymbol{y}}^L;\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)},\boldsymbol{\psi})$, which can be calculated by standard methods applied to the model given by

$$\begin{aligned} \bar{\boldsymbol{y}}^L &= \boldsymbol{\lambda} + \bar{\boldsymbol{e}}, & \bar{\boldsymbol{e}} &\sim N(\boldsymbol{0},(A^{(s)*\prime}(\boldsymbol{\Omega}^*)^{-1}A^{(s)*})^{-1}), \\ \boldsymbol{\lambda} &= (\boldsymbol{\lambda}_1',\ldots,\boldsymbol{\lambda}_N')', & \boldsymbol{\lambda}_i &\sim NID(\boldsymbol{\delta},\boldsymbol{\Sigma}_\lambda), \end{aligned}$$

When $\boldsymbol{\Omega}$ is diagonal $\mathrm{E}(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)},\boldsymbol{\psi})$ can be computed separately for each $\boldsymbol{\lambda}_i$; see Mesters en Koopman (2014) for additional details.

Appendix C

The conditional mean function $\bar{\boldsymbol{f}} = \mathrm{E}(\boldsymbol{f}(\boldsymbol{\lambda})|\boldsymbol{y};\boldsymbol{\psi})$ can be expressed in terms of the importance density $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$, that is

$$\bar{\boldsymbol{f}} = \int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) \frac{p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})}{g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})} g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}.$$

By adopting the Bayes' rule, we obtain

$$\bar{\boldsymbol{f}} = \frac{g(\boldsymbol{y};\boldsymbol{\psi})}{p(\boldsymbol{y};\boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}, \tag{23}$$

where the "integrated" weights $w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi})$ are given by

$$w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) = \frac{p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})}. \tag{24}$$

When choosing $\boldsymbol{f}(\boldsymbol{\lambda}) = 1$ we obtain

$$1 = \frac{g(\boldsymbol{y};\boldsymbol{\psi})}{p(\boldsymbol{y};\boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}, \tag{25}$$

and finally, when taking the ratio of (23) and (25) we obtain

$$\bar{\boldsymbol{f}} = \frac{\int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}}{\int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}}. \tag{26}$$

The expression in (26) only depends on the latent loading vectors. The latent factors are implicitly integrated out. The Monte Carlo estimate based on (26) is given by (13).

Appendix D

Here we develop an importance sampling estimate of $\bar{\boldsymbol{h}}$ in 15, that is

$$\bar{\boldsymbol{h}} = \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) \; p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}.$$

We choose an adequate importance density that avoids sampling from $p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$. For this purpose, we can rewrite $\bar{\boldsymbol{h}}$ as

$$\begin{aligned}\bar{\boldsymbol{h}} &= \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi})\ \frac{p(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})}{g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})} g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\lambda} \\ &= \frac{g(\boldsymbol{y};\psi)}{p(\boldsymbol{y};\boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi})\ \frac{p(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\lambda};\boldsymbol{\psi})} g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\lambda} \\ &= \frac{g(\boldsymbol{y};\psi)}{p(\boldsymbol{y};\boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi})\ w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\lambda}, \end{aligned} \tag{27}$$

where the weights $w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi})$ are given in (24). When we choose $\boldsymbol{h}(\boldsymbol{\alpha}) = 1$ we obtain

$$1 = \frac{g(\boldsymbol{y};\psi)}{p(\boldsymbol{y};\boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\lambda}. \tag{28}$$

Finally, by taking the ratio of (27) and (28) we get

$$\bar{\boldsymbol{h}} = \frac{\int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi})\ w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\lambda}}{\int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})\ \mathrm{d}\boldsymbol{\lambda}}, \tag{29}$$

for which a Monte Carlo estimate is given by

$$\bar{\boldsymbol{h}} = \frac{M^{-1}\sum_{j=1}^{M} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y},\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})\ w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})}{M^{-1}\sum_{j=1}^{M} w_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda}^{(j)};\boldsymbol{\psi})}, \qquad M \to \infty,$$

where the samples $\boldsymbol{\lambda}^{(j)}$ are drawn from $g(\boldsymbol{\lambda}|\boldsymbol{y};\boldsymbol{\psi})$.

Appendix E

In this appendix we summarize the modifications for the methods of Section 3 that occur when a selection of observations is missing. The steps in the posterior mode algorithm in Section 3 rely on multivariate regression methods and the Kalman filter smoother. These methods can be adjusted to deal with missing values by using the methods in Wooldridge (2010, Chapter 19) and Durbin en Koopman (2012, Section 4.10). The simulation methods are adjusted similarly as discussed in Mesters en Koopman (2014, Section 4.4).

References

Omar Aguilar en Mike West, 2000. Bayesian Dynamic Factor Models and Portfolio Allocation. *Journal of Business and Economic Statistics*, 18 (2000), 338–357.

Brian D. O. Anderson en John B. Moore, 1979. *Optimal Filtering*. Englewood Cliffs, Prentice-Hall.

Craig F. Ansley en Robert Kohn, 1986. A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *Journal of Statistical Computation and Simulation*, 24 (1986), 99–106.

Jushan Bai, 2003. Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71 (2003), 135–171.

Jushan Bai en Kunpeng Li, 2012. Statistical Analysis of Factor Models of High Dimension. *The Annals of Statistics*, 40 (2012), 436–465.

Jushan Bai en Serena Ng, 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70 (2002), 191–221.

Jushan Bai en Serena Ng, 2008. Large Dimensional Factor Analysis. *Foundations and Trends in Econometrics*, 3 (2008), 89–163.

Ben S. Bernanke, Jean Boivin, en Piotr Eliasz, 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics*, 120 (2005), 387–422.

Julian. E. Besag, 1986. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society, Series B*, 48 (1986), 259–302.

Anirban Bhattacharya en David B. Dunson, 2011. Sparse Bayesian infinite factor models. *Biometrika*, 98 (2011), 291–306.

Christopher M. Bishop, 2006. *Pattern Recognition and Machine Learning*. New York, Springer.

Falk Bräuning en Siem Jan Koopman, 2013. Forecasting Macroeconomic Variables using Collapsed Dynamic Factor Analysis. *International Journal of Forecasting*, 29 (2013), forthcoming.

Jörg Breitung en Jörn Tenhofen, 2011. GLS Estimation of Dynamic Factor Models. *Journal of the American Statistical Association*, 106 (2011), 1150–1166.

Gerald G. Brown en Herbert C. Rutemiller, 1977. Means and Variances of Stochastic Vector Products with Applications to Random Linear Models. *Management Science*, 24 (1977), 210–216.

Jurgen A. Doornik, 2007. *Object-Oriented Matrix Programming Using Ox*. London, Timberlake Consultants Press.

Arnaud N. Doucet, Nando de Freitas, en Neil J. Gordon, 2001. *Sequential Monte Carlo Methods in Practice*. New York, Springer Verlag.

Catherine Doz, Domenico Giannone, en Lucrezia Reichlin, 2011. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164 (2011), 188–205.

Catherine Doz, Domenico Giannone, en Lucrezia Reichlin, 2012. A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models. *Review of Economics and Statistics*, 94 (2012), 1014–1024.

James Durbin en Siem Jan Koopman, 2012. *Time Series Analysis by State Space Methods*. Oxford, Oxford University Press.

Brad Efron, 2010. *Large Scale Inference, Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge, Cambridge University Press.

Brad Efron en Carl Morris, 1973. Stein's Estimation Rule and Its Competitors. *Journal of the American Statistical Association*, 68 (1973), 117–130.

Mario Forni, Marc Hallin, Marco Lippi, en Lucrezia Reichlin, 2000. The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics*, 82 (2000), 540–554.

John F. Geweke, 1989. Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57 (1989), 1317–1339.

John F. Geweke en Guofu Zhou, 1996. Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, 9 (1996), 557–587.

Domenico Giannone, Lucrezia Reichlin, en David Small, 2008. Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55 (2008), 665–676.

Bruce E. Hansen, 2013. Efficient Shrinkage in Parametric Models. (2013). Working paper.

Willard James en Charles Stein, 1961. Estimation with Quadratic Loss. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, redactie, p. 361–379. University of California Press, Berkeley.

Simon J. Julier en Jeffrey K. Uhlmann, 1997. A new extension of the Kalman filter to nonlinear systems. In *Signal Processing, Sensor Fusion, and Target Recognition VI*, I. Kadar, redactie, deel 3068, p. 182–193. SPIE.

Boris Jungbacker en Siem Jan Koopman, 2007. Monte Carlo Estimation for Nonlinear Non-Gaussian State Space Models. *Biometrika*, 94 (2007), 827–839.

Boris Jungbacker en Siem Jan Koopman, 2014. Likelihood-based Analysis for Dynamic Factor Analysis for Measurement and Forecasting. *Econometrics Journal*, 17 (2014), forthcoming.

Sylvia Kaufmann en Christian Schumacher, 2013. Bayesian estimation of sparse dynamic factor model with order-independent identification. (2013). Working paper.

Thomas Knox, James H. Stock, en Mark W. Watson, 2004. Empirical Bayes Regression With Many Regressors. (2004). Working paper.

Siem Jan Koopman, Andre Lucas, en Marcel Scharth, 2014. Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models. *Journal of Business and Economic Statistics*, 32 (2014), forthcoming.

Siem Jan Koopman en Neil Shephard, 1992. Exact score for time series models in state space form. *Biometrika*, 79 (1992), 823–826.

Siem Jan Koopman, Neil Shephard, en Drew D. Creal, 2009. Testing the assumptions behind importance sampling. *Journal of Econometrics*, 149 (2009), 2–11.

M. Ayhan Kose, Christopher Otrok, en Charles H. Whiteman, 2003. International Business Cycles: World, Region, and Country-Specific Factors. *American Economic Review*, 93 (2003), 1216–1239.

J. Steve Marron en Matt P. Wand, 1992. Exact Mean Integrated Squared Error. *The Annals of Statistics*, 20 (1992), 712–736.

Xiao-Li Meng en Donald B. Rubin, 1993. Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika*, 80 (1993), 267–278.

Geert Mesters en Siem Jan Koopman, 2014. Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time. *Journal of Econometrics*, 154 (2014), 127–140.

Jouchi Nakajima en Mike West, 2013. Bayesian Analysis of Latent Threshold Dynamic Models. *Journal of Business and Economic Statistics*, 31 (2013), 151–164.

Whitney K. Newey en Daniel L. McFadden, 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, R. F. Engle en D. L. McFadden, redactie, p. 2111–2245. Elsevier, Amsterdam.

Thomas J. Sargent, 1989. Two Models of Measurements and the Investment Accelerator. *Journal of Political Economy*, 97 (1989), 251–287.

Mike K. P. So, 2003. Posterior mode estimation for nonlinear and non-Gaussian state space models. *Statistica Sinica*, 13 (2003), 255–274.

James H. Stock en Mark W. Watson, 1989. A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates. *Journal of Business and Economic Statistics*, 76 (1989), 774–781.

James H. Stock en Mark W. Watson, 2002a. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 220 (2002), 147–162.

James H. Stock en Mark W. Watson, 2002b. Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97 (2002), 1167–1179.

James H. Stock en Mark W. Watson, 2005. Implications of Dynamic Factor Models for VAR Analysis. (2005). manuscript.

James H. Stock en Mark W. Watson, 2011. Dynamic Factor Models. In *Oxford Handbook of Economic Forecasting*, M. P. Clements en D. F. Hendry, redactie. Oxford University Press, Oxford.

James H. Stock en Mark W. Watson, 2012. Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business and Economic Statistics*, 30 (2012), 481–493.

Mark W. Watson en Robert F. Engle, 1983. Alternative Algorithms for the Estimation of Dynamic Factor MIMIC and Varying Coefficient Models. *Journal of Econometrics*, 23 (1983), 385–400.

Jeffrey M. Wooldridge, 2010. *Econometric analysis of cross section and panel data - 2nd edition*. Cambridge, Massachusetts, The MIT Press.

Jeff C. F. Wu, 1983. On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11 (1983), 95–103.