Web-appendix for:
Empirical Bayes Methods for Dynamic Factor Models

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

[(a)] VU University Amsterdam, Tinbergen Institute and
CREATES, Aarhus University
[(b)] Universitat Pompeu Fabra, Barcelona GSE and
The Netherlands Institute for the Study of Crime and Law Enforcement

March 21, 2016

Abstract

We show the appendices and additional results that pertain to the paper "Empirical Bayes Methods for Dynamic Factor Models". We include: Appendix A: proof theorem 1, Appendix B: proof theorem 2, Appendix C: importance sampling representation loadings, Appendix D: importance sampling representation factors, Appendix E: importance sampling weights, Appendix F: adjustments for missing values, Appendix G: additional simulation results comparison to principal components, Appendix H: additional simulation results for different disturbance specifications, Appendix I: additional results for macro-forecasting study.

*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. Contact address: Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 2527, 08005 Barcelona, Spain, tel. nr. +34 672 124 872, email: geert.mesters@upf.edu, web: http://www.geertmesters.nl.

Appendix A

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

The observation equation for the dynamic factor model 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 $\boldsymbol{\alpha}_1^* = (\boldsymbol{\alpha}_1', \boldsymbol{0}, \ldots, \boldsymbol{0})'$, $\boldsymbol{\eta} = (\boldsymbol{0}, \boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_T')'$ and

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

It holds that

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

where

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

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 the dynamic factor model 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 $\boldsymbol{\lambda}$. More specifically, when fixing either $\boldsymbol{\alpha}$ or $\boldsymbol{\lambda}$ at a particular value $\tilde{\boldsymbol{\alpha}}$ or $\tilde{\boldsymbol{\lambda}}$ the densities $p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}})$ and $p(\boldsymbol{\alpha}|\boldsymbol{y}; \boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}})$ do not depend on the random variables $\boldsymbol{\alpha}$ or $\boldsymbol{\lambda}$. 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 & 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 & 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,

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

and

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

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 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 $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(s-1)}$ the dynamic factor model is a linear Gaussian state space model. Jungbacker & Koopman (2015) 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 $\boldsymbol{y}^L = (y_1^{L'}, \ldots, y_T^{L'})'$, with

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

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

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

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 & Koopman (2015) for additional details.

Given $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(s)}$ the dynamic factor model 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 & 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 & 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

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

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

$$
w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) = \frac{p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi})}. \tag{2}
$$

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

$$
1 = \frac{g(\boldsymbol{y}; \boldsymbol{\psi})}{p(\boldsymbol{y}; \boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \; \mathrm{d}\boldsymbol{\lambda}, \tag{3}
$$

and finally, when taking the ratio of (1) and (3) we obtain

$$\bar{\boldsymbol{f}} = \frac{\int_{\boldsymbol{\lambda}} \boldsymbol{f}(\boldsymbol{\lambda}) w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}}{\int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}}. \tag{4}$$

The expression in (4) only depends on the latent loading vectors. The latent factors are implicitly integrated out. The Monte Carlo estimate based on (4) is given in the paper.

Appendix D

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

$$\begin{aligned}
\bar{\boldsymbol{h}} &= \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) \; \frac{p(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})}{g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi})} g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda} \\
&= \frac{g(\boldsymbol{y}; \psi)}{p(\boldsymbol{y}; \boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) \; \frac{p(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi})}{g(\boldsymbol{y}|\boldsymbol{\lambda}; \boldsymbol{\psi})} g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda} \\
&= \frac{g(\boldsymbol{y}; \psi)}{p(\boldsymbol{y}; \boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) \; w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda},
\end{aligned} \tag{5}$$

where the weights $w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi})$ are given in (2). When we choose $\boldsymbol{h}(\boldsymbol{\alpha}) = 1$ we obtain

$$1 = \frac{g(\boldsymbol{y}; \psi)}{p(\boldsymbol{y}; \boldsymbol{\psi})} \int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}. \tag{6}$$

Finally, by taking the ratio of (5) and (6) we get

$$\bar{\boldsymbol{h}} = \frac{\int_{\boldsymbol{\lambda}} \mathrm{E}(h(\boldsymbol{\alpha})|\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) \; w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}}{\int_{\boldsymbol{\lambda}} w_{\boldsymbol{\lambda}}(\boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{\psi}) g(\boldsymbol{\lambda}|\boldsymbol{y}; \boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\lambda}}, \tag{7}$$

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

The integrated importance sampling weights $w_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \boldsymbol{y}; \boldsymbol{\psi})$ of Koopman & Mesters (2015, Section 3.3) need to have finite variance in order for the conditional mean function estimates in Koopman & Mesters (2015, Equations 11 and 14) 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, Shephard & Creal (2009) to empirically assess whether the integrated weights have finite variance. For the simulated observation vectors $\boldsymbol{y}(1)$ for different panel sizes and number of factors we estimate the parameter vector as discussed in Koopman & Mesters (2015, Section 4). 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 $w^{\min}$. The threshold is determined such to include the largest 1% to 50% of the weights 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.

Appendix F

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 & Koopman (2012, Section 4.10). The simulation methods

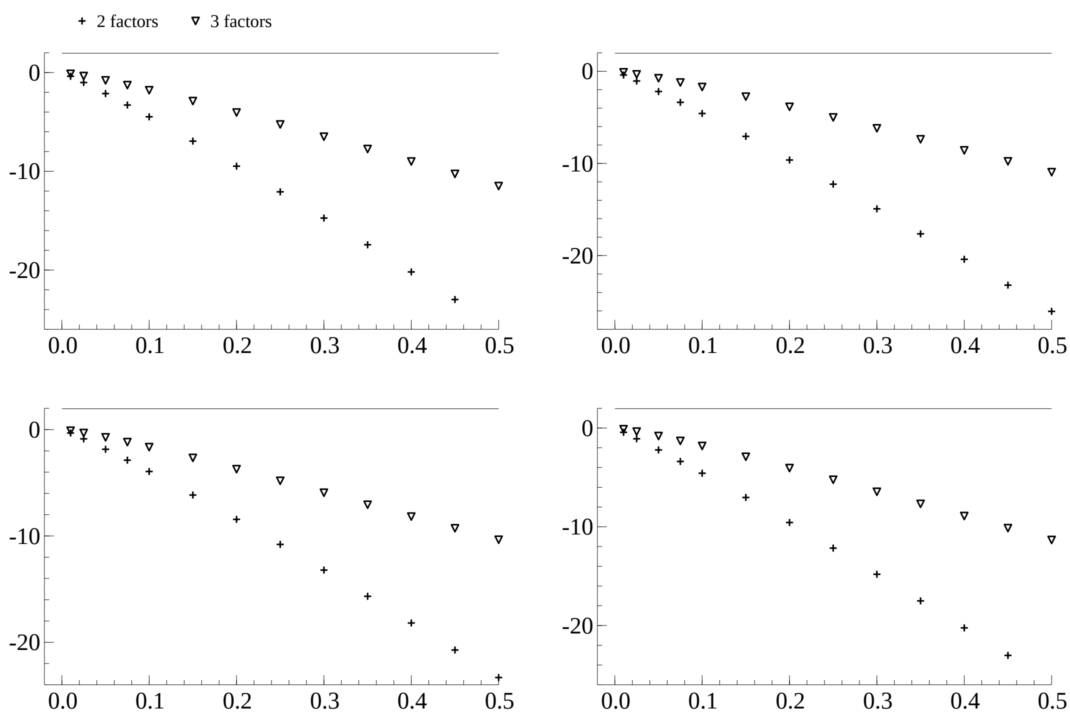

Figure 1: Importance sampling diagnostics for dynamic factor models with $r = 3$ and $r = 5$ factors, based on 100,000 simulations of weights $w_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}^{(j)}, \boldsymbol{y}; \hat{\boldsymbol{\psi}})$. We computed test statistics for different thresholds $w^{\min}$.

are adjusted similarly as discussed in Mesters & Koopman (2014, Section 4.4).

Appendix G

We compare the empirical Bayes and maximum likelihood estimates to the principal components estimates (PCA). We consider the same simulation design as that in Koopman & Mesters (2015, Section 4). A minor difference is that for the PCA estimator we need to standardize each time series prior to estimation. Based on the standardized time series we compute the principal components estimates for the factors and the loadings as in Stock & Watson (2002). Using the loadings and the factors we compute the inner-product and rescale this estimate to compute the mean-squared error statistics for the inner-product considered in Koopman & Mesters (2015, equation 17). Note that we only compute the inner product comparisons as the principal components estimates identify a different rotation of the loadings and the factors. This renders comparing the individual factors and loadings with the mean squared error statistics uninformative. We refer to Doz, Giannone & Reichlin (2012), Bai & Li (2012), Banbura & Modugno (2014) and Bai & Li (2015) for a more in debt comparisons of the PCA and MLE methods.

The results for the inner-product comparisons are shown in Table 1. In panel (i) we show the mean squared error of the empirical Bayes estimates relative to the principal components estimates. The empirical Bayes estimates are always more accurate when compared to the principal components estimates. The gains are large in magnitude and range between 30 and 50% in mean squared error accuracy. In panel (ii) we show the same statistics but now for the ratio between the maximum likelihood en principal components estimates. Also for the maximum likelihood estimates we find large gains relative to the principal components estimates. We notice that in line with Koopman & Mesters (2015, Table 1 panel (iii)) the relative gains of the maximum likelihood estimates over the principal components estimates are smaller when compared to the relative gains of the empirical Bayes estimates over the principal components estimates.

Appendix H

In this section we present some additional simulation results for models with different specifications for the error term. In particular, we modify the parameters in Koopman & Mesters (2015, equation 16) to obtain a model without serial correlation ($\gamma = 0$), a model without cross-sectional correlation ($\tau = 0$) and a model without serial and cross-sectional correlation ($\gamma = 0$ and $\tau = 0$). All other features of the simulation design are kept constant.

The results are shown in Table 2. Panels (i), (ii) and (iii) show the results for the model without serial correlation. Panels (iv), (v) and (vi) show the results for the model without cross-sectional correlation. Panels (vii), (viii) and (ix) show the results for the model without serial and cross-sectional correlation. Overall the results barely change when compared to Panels (i), (ii) and (iii) in Koopman & Mesters (2015, Table 1). The small differences that

			Panel (i): RMSE($\boldsymbol{\lambda}'\boldsymbol{\alpha}$)					Panel (ii): RMSE($\boldsymbol{\lambda}'\boldsymbol{\alpha}$)				
N	T	r	L=1	L=2	L=3	L=4	L=5	L=1	L=2	L=3	L=4	L=5
50	50	3	0.575	0.585	0.576	0.501	0.570	0.687	0.662	0.674	0.644	0.677
50	100	3	0.623	0.651	0.612	0.532	0.634	0.674	0.682	0.670	0.622	0.674
100	50	3	0.569	0.636	0.584	0.510	0.568	0.693	0.699	0.696	0.676	0.694
100	100	3	0.636	0.639	0.633	0.567	0.632	0.707	0.697	0.701	0.678	0.702
150	200	3	0.672	0.672	0.675	0.622	0.687	0.708	0.702	0.711	0.684	0.714
50	50	5	0.541	0.570	0.540	0.506	0.531	0.683	0.665	0.674	0.667	0.678
50	100	5	0.534	0.553	0.545	0.473	0.532	0.619	0.619	0.627	0.580	0.627
100	50	5	0.571	0.621	0.573	0.489	0.562	0.739	0.726	0.730	0.687	0.742
100	100	5	0.609	0.627	0.611	0.513	0.604	0.717	0.715	0.715	0.658	0.719
150	200	3	0.684	0.688	0.686	0.605	0.693	0.745	0.742	0.746	0.699	0.749

Table 1: Simulation results for the empirical Bayes and principal components estimates. The DGP and parameters are chosen as discussed in Section 4.1 Koopman and Mesters (2015). The code L indicates 1; normal(0,1), 2; normal(0,0.04), 3; tri-modal, 4; skewed, or 5; kurtotic distribution for the true-loadings. Panel (i) compares the inner-products of empirical Bayes vs principal components. Panel (ii) compares the inner-products of maximum likelihood vs principal components.

do occur are not systematic in any way and are most likely due to Monte Carlo simulation error. The results imply that the misspecification of the error term has similar influence on the maximum likelihood and empirical Bayes methods in a finite sample setting.

Appendix I

In this appendix we provide some additional results for the macroeconomic application of Koopman & Mesters (2015, Section 5). In particular, we show the estimated factors and loadings from the in-sample analysis and the forecasts from the principal components method.

Additional in-sample results

In Figure 2 we show the estimated factors for both for the empirical Bayes and maximum likelihood estimation methods from the 5 factor model. Note that the model is only identified up to an orthonormal rotation matrix, such that the large differences between the factors are not all attributable to differences between the methods. A large part of the differences is just because the methods identify a different rotation. This point is made clear in Figure 4 where we show the inner-product for 4 different series. The differences between the inner-product estimates are small.

In Figure 3 we show the correlations between the factors and the individual time series. We find that the first factor is associated with the price variables, whereas the second and third factors load more heavily on the real economic variables.

N	T	r	Panel (i): RMSE($\boldsymbol{\lambda}$) L=1	L=2	L=3	L=4	L=5	Panel (ii): RMSE($\boldsymbol{\alpha}$) L=1	L=2	L=3	L=4	L=5	Panel (iii): RMSE($\boldsymbol{\lambda}'\boldsymbol{\alpha}$) L=1	L=2	L=3	L=4	L=5
50	50	3	0.887	0.912	0.898	0.866	0.866	0.983	0.982	0.982	0.979	0.976	0.862	0.895	0.875	0.783	0.866
50	100	3	1.230	1.163	1.227	1.165	1.222	1.093	1.061	1.098	1.044	1.066	1.000	0.988	0.964	0.921	0.992
100	50	3	0.847	0.869	0.853	0.791	0.852	1.218	1.219	1.244	1.134	1.191	0.855	0.937	0.866	0.782	0.873
100	100	3	0.940	0.948	0.936	0.889	0.956	0.989	0.988	1.001	0.989	0.985	0.955	0.962	0.933	0.881	0.985
50	50	5	0.831	0.875	0.849	0.837	0.828	0.973	0.971	0.987	0.980	0.969	0.785	0.845	0.793	0.735	0.785
50	100	5	1.115	1.085	1.115	1.047	1.083	0.991	0.986	0.976	0.978	0.993	0.901	0.932	0.905	0.818	0.902
100	50	5	0.754	0.788	0.757	0.717	0.790	1.179	1.177	1.201	1.121	1.198	0.768	0.851	0.780	0.688	0.788
100	100	5	0.897	0.911	0.905	0.861	0.907	0.996	0.996	0.995	0.989	0.996	0.880	0.902	0.877	0.796	0.894
			Panel (iv): RMSE($\boldsymbol{\lambda}$)					Panel (v): RMSE($\boldsymbol{\alpha}$)					Panel (vi): RMSE($\boldsymbol{\lambda}'\boldsymbol{\alpha}$)				
N	T	r	L=1	L=2	L=3	L=4	L=5	L=1	L=2	L=3	L=4	L=5	L=1	L=2	L=3	L=4	L=5
50	50	3	0.891	0.909	0.898	0.861	0.886	0.964	0.963	0.980	0.977	1.228	0.839	0.877	0.841	0.752	0.842
50	100	3	1.154	1.113	1.124	1.052	1.150	0.994	0.988	0.985	0.961	0.981	0.949	0.972	0.909	0.857	0.962
100	50	3	0.837	0.859	0.838	0.793	0.838	1.252	1.251	1.277	1.199	0.946	0.828	0.919	0.844	0.751	0.826
100	100	3	0.940	0.945	0.934	0.883	0.912	0.989	0.983	0.997	0.986	0.985	0.920	0.933	0.911	0.851	0.923
50	50	5	0.845	0.880	0.852	0.844	0.850	1.156	1.372	1.030	1.107	1.111	0.755	0.829	0.771	0.709	0.788
50	100	5	1.061	1.046	1.074	0.992	1.055	0.924	0.924	0.922	0.916	0.920	0.858	0.892	0.877	0.796	0.884
100	50	5	0.780	0.811	0.787	0.747	0.809	1.250	1.250	1.268	1.215	1.248	0.767	0.854	0.785	0.694	0.796
100	100	5	0.903	0.915	0.904	0.851	0.907	0.993	0.993	0.993	0.993	0.992	0.854	0.881	0.854	0.775	0.870
			Panel (vii): RMSE($\boldsymbol{\lambda}$)					Panel (viii): RMSE($\boldsymbol{\alpha}$)					Panel (ix): RMSE($\boldsymbol{\lambda}'\boldsymbol{\alpha}$)				
N	T	r	L=1	L=2	L=3	L=4	L=5	L=1	L=2	L=3	L=4	L=5	L=1	L=2	L=3	L=4	L=5
50	50	3	0.911	0.929	0.908	0.861	0.912	0.983	0.986	0.995	0.994	0.972	0.895	0.923	0.884	0.791	0.925
50	100	3	1.192	1.148	1.168	1.073	1.197	1.069	1.068	1.047	1.033	1.072	0.988	0.996	0.950	0.901	0.983
100	50	3	0.861	0.880	0.858	0.784	0.885	1.290	1.288	1.333	1.227	1.264	0.863	0.954	0.859	0.762	0.893
100	100	3	0.944	0.951	0.944	0.882	0.960	0.990	0.988	1.002	0.987	0.980	0.944	0.953	0.931	0.870	0.975
50	50	5	0.845	0.885	0.843	0.821	0.843	0.997	0.996	0.998	1.742	0.990	0.792	0.860	0.792	0.713	0.794
50	100	5	1.121	1.092	1.125	1.031	1.115	0.985	0.984	0.989	0.982	0.990	0.919	0.950	0.916	0.851	0.943
100	50	5	0.772	0.802	0.767	0.719	0.772	1.209	1.205	1.233	1.174	1.189	0.780	0.868	0.773	0.684	0.779
100	100	5	0.905	0.920	0.908	0.857	0.907	0.992	0.992	0.995	0.999	0.985	0.887	0.907	0.882	0.806	0.898

Table 2: Simulation results for the empirical Bayes and maximum . The DGP and parameters are chosen as discussed in Section 4.1 Koopman and Mesters (2015). The code L indicates 1; normal(0,1), 2; normal(0,0.04), 3; tri-modal, 4; skewed, or 5; kurtotic distribution for the true-loadings. Panels (i), (ii) and (iii) show the results for the model without serial correlation ($\gamma = 0$). Panels (iv), (v) and (vi) show the results for the model without cross-sectional correlation ($\tau = 0$). Panels (vii), (viii) and (ix) show the results for the model without serial and cross-sectional correlation ($\gamma = 0$ and $\tau = 0$).

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

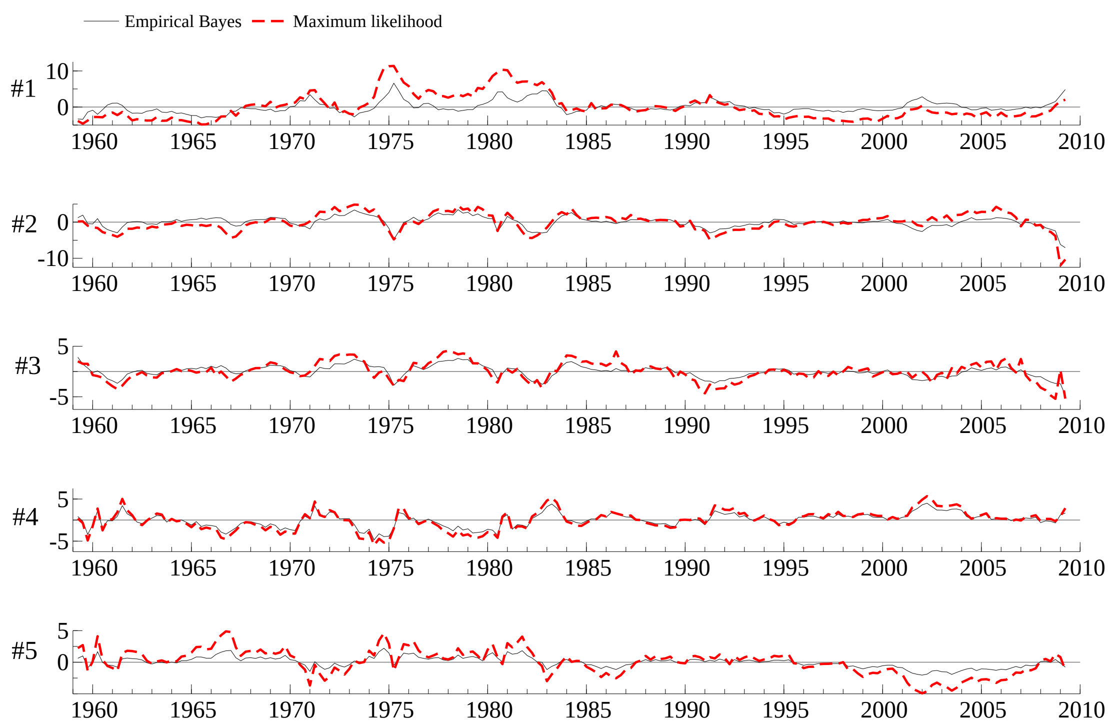

Figure 2: Posterior mode empirical Bayes estimates and maximum likelihood estimates for the factors.

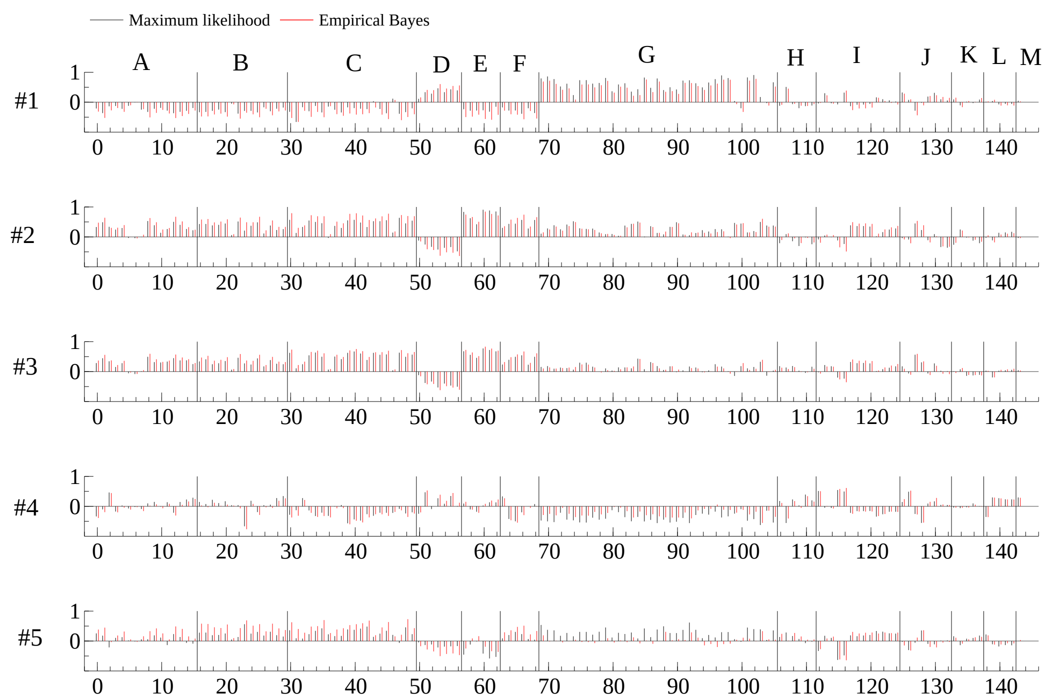

Figure 3: Posterior mode empirical Bayes estimates and maximum likelihood estimates for the loadings. The left bars pertain to the empirical Bayes estimates and the right bars pertain to the maximum likelihood esitimates.

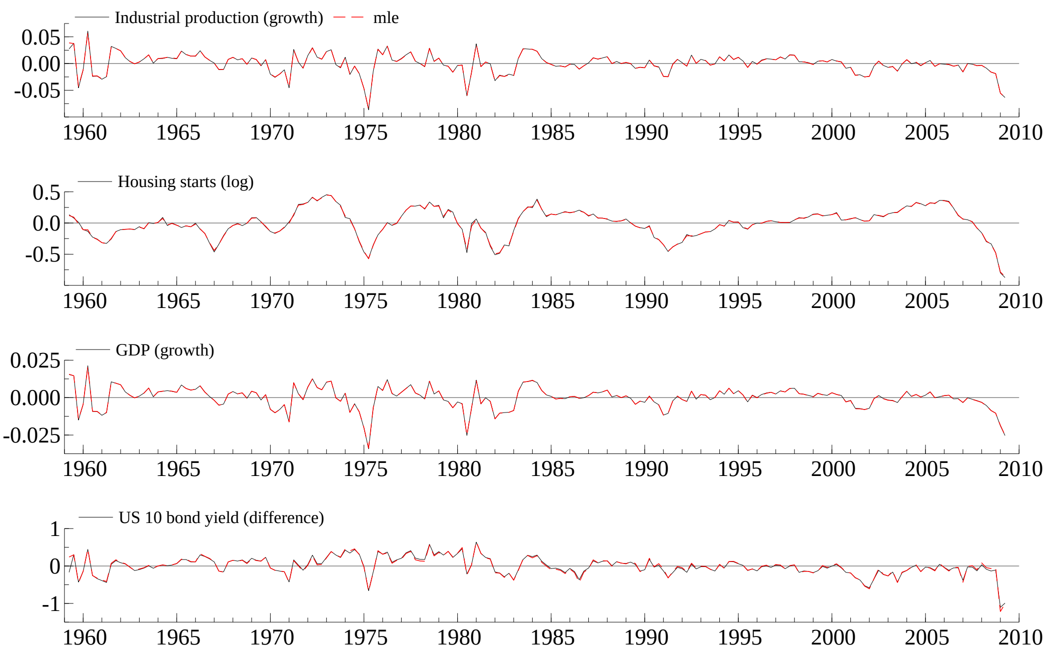

Figure 4: Posterior mode empirical Bayes estimates and maximum likelihood estimates for the inner-products for 4 time series.

Principal components forecasts

Next, we discuss the forecasting results based on the principal components method. The set-up for the forecasting study for the principal components method is similar to Koopman & Mesters (2015, Section 5.2). A minor difference is that prior to estimation we standardize each time series to have variance one. The forecasts are rescaled before computing the mean squared error as in Koopman & Mesters (2015, equation 19). The forecasts are computed following Stock & Watson (2002). Missing values are handled using the EM algorithm detailed in Stock & Watson (2002).

The results for the 5 factor model are presented in Table 4. The results are clearly in favor of the likelihood based methods. The relative gain is 50% for the first horizon and still 30% for the four quarter ahead forecasts. The results for the 7 factor model are presented in Table 5. The findings are again in favor of the empirical Bayes and maximum likelihood estimates.

	EB vs PCA			MLE vs PCA		
	$h=1$	$h=2$	$h=4$	$h=1$	$h=2$	$h=4$
All series						
Mean	0.499	0.580	0.683	0.547	0.593	0.687
Quantiles						
0.05	0.165	0.149	0.149	0.174	0.146	0.138
0.25	0.316	0.402	0.461	0.358	0.392	0.498
0.50	0.676	0.672	0.736	0.723	0.699	0.755
0.75	0.879	0.893	0.911	0.917	0.912	0.924
0.95	1.069	1.041	1.051	1.088	1.089	1.130
Components (*Mean*)						
GDP components	0.460	0.466	0.530	0.510	0.504	0.569
Industrial Production	0.720	0.818	0.681	0.751	0.840	0.695
Employment	0.770	0.754	0.810	0.929	0.910	0.924
Unemployment rate	0.730	0.741	0.841	0.739	0.765	0.863
Housing	0.528	0.616	0.748	0.618	0.653	0.755
Inventories	0.329	0.419	0.497	0.469	0.482	0.548
Prices	0.517	0.589	0.804	0.486	0.551	0.755
Wages	0.587	0.460	0.501	0.581	0.463	0.506
Interest rates	0.959	0.856	0.989	0.986	0.852	0.983
Money	1.315	1.001	1.004	1.010	1.005	1.003
Exchanges rates	0.926	0.936	0.912	0.897	0.927	0.902
Stock prices	1.074	1.101	1.034	1.027	1.131	1.024
Consumer Expectations	0.724	0.913	0.856	0.779	0.935	0.847

Table 4: Relative mean squared error statistics for out-of-sample forecasting of the empirical Bayes and maximum likelihood methods versus the principal components method for the model with $r = 5$ factors. The results summarize the distribution of the statistics $\text{MSE}_i^{\text{PEB}}/\text{MSE}_i^{\text{PCA}}$ and $\text{MSE}_i^{\text{MLE}}/\text{MSE}_i^{\text{PCA}}$, for $i = 1, \ldots, 144$ and forecast horizons $h = 1, 2, 4$.

	EB vs PCA			MLE vs PCA		
	$h = 1$	$h = 2$	$h = 4$	$h = 1$	$h = 2$	$h = 4$
All series						
Mean	0.591	0.548	0.645	0.662	0.576	0.660
Quantiles						
0.05	0.125	0.130	0.132	0.133	0.116	0.113
0.25	0.307	0.385	0.423	0.349	0.377	0.466
0.50	0.643	0.608	0.708	0.727	0.664	0.698
0.75	0.871	0.875	0.896	0.910	0.873	0.909
0.95	1.135	1.019	1.079	1.195	1.128	1.107
Components (*Mean*)						
GDP components	0.425	0.448	0.480	0.510	0.463	0.505
Industrial Production	0.651	0.742	0.670	0.777	0.830	0.736
Employment	0.770	0.721	0.785	0.921	0.842	0.907
Unemployment rate	0.648	0.698	0.817	0.629	0.702	0.826
Housing	0.473	0.600	0.773	0.651	0.712	0.823
Inventories	0.335	0.395	0.488	0.507	0.455	0.542
Prices	0.546	0.556	0.728	0.524	0.544	0.691
Wages	0.566	0.439	0.464	0.589	0.458	0.472
Interest rates	1.013	0.871	1.004	1.107	0.929	1.049
Money	1.476	1.010	1.010	1.012	1.006	1.004
Exchanges rates	0.938	0.921	0.899	0.891	0.890	0.879
Stock prices	1.083	1.099	1.059	0.988	1.115	1.034
Consumer Expectations	0.681	0.882	0.790	0.739	0.899	0.794

Table 5: Relative mean squared error statistics for out-of-sample forecasting of the empirical Bayes and maximum likelihood methods versus the principal components method for the model with $r = 7$ factors. The results summarize the distribution of the statistics $\text{MSE}_i^{\text{PEB}}/\text{MSE}_i^{\text{PCA}}$ and $\text{MSE}_i^{\text{MLE}}/\text{MSE}_i^{\text{PCA}}$, for $i = 1, \ldots, 144$ and forecast horizons $h = 1, 2, 4$.

References

Bai, J. & Li, K. (2012), 'Statistical Analysis of Factor Models of High Dimension', *The Annals of Statistics* **40**, 436–465.

Bai, J. & Li, K. (2015), 'Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension', *Review of Economics and Statistics* . forthcoming.

Banbura, M. & Modugno, M. (2014), 'Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data', *Journal of Applied Econometrics* **29**, 133–160.

Doz, C., Giannone, D. & Reichlin, L. (2012), 'A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models', *Review of Economics and Statistics* **94**, 1014–1024.

Durbin, J. & Koopman, S. J. (2012), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

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

Jungbacker, B. & Koopman, S. J. (2015), 'Likelihood-based Analysis for Dynamic Factor Analysis for Measurement and Forecasting', *Econometrics Journal* **18**, C1–C21.

Koopman, S. J. & Mesters, G. (2015), 'Empirical Bayes Methods for Dynamic Factor Models'. Tinbergen Institute Working Paper: TI 2014-061/III.

Koopman, S. J., Shephard, N. & Creal, D. D. (2009), 'Testing the assumptions behind importance sampling', *Journal of Econometrics* **149**, 2–11.

Meng, X.-L. & Rubin, D. B. (1993), 'Maximum Likelihood Estimation via the ECM Algorithm: A General Framework', *Biometrika* **80**, 267–278.

Mesters, G. & Koopman, S. J. (2014), 'Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time', *Journal of Econometrics* **154**, 127–140.

Stock, J. H. & Watson, M. W. (2002), 'Forecasting Using Principal Components From a Large Number of Predictors', *Journal of the American Statistical Association* **97**, 1167–1179.

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

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