Bayesian estimation of sparse dynamic factor models with order-independent identification

Sylvia Kaufmann and Christian Schumacher

Working Paper 13.04

Bayesian estimation of sparse dynamic factor models with order-independent identification

Sylvia Kaufmann* and Christian Schumacher†‡

January 2013

Abstract

The analysis of large panel data sets (with N variables) involves methods of dimension reduction and optimal information extraction. Dimension reduction is usually achieved by extracting the common variation in the data into few factors (k, where $k \ll N$). In the present project, factors are estimated within a state space framework. To obtain a parsimonious representation, the $N \times k$ factor loading matrix is estimated under a sparse prior, which assumes that either many zeros may be present in each column of the matrix, or many rows may contain zeros. The significant factor loadings in columns define the variables driven by specific factors and offer an explicit interpretation of the factors. Zeros in rows indicate irrelevant variables which do not add much information to the inference. The contribution includes a new way of identification which is independent of variable ordering and which is based on semi-orthogonal loadings.
JEL classification: C11,C32
Key words: Dynamic factor model, identification, sparsity.

1 Introduction

Sparse factor models greatly condense information in large cross-section or panel data sets. So far they have particularly been used in gene expression analysis, where only few out of potentially tens of thousands of genes may be responsible for some physiological outcome of interest. Individual gene expressions may thus be influenced by common biological factors, each of which involves only a subgroup of genes. A sparse loading matrix arises naturally in this context, in which many zero rows indicate that only a small share of all genes determine the biological factors of interest and zeros in columns indicate that

*Study Center Gerzensee, Foundation of the Swiss National Bank, Dorfstrasse 2, P.O. Box 21, CH-3115 Gerzensee, sylvia.kaufmann@szgerzensee.ch

†Deutsche Bundesbank, Wilhelm-Epstein-Str.14, 60431 Frankfurt/Main, Germany, christian.schumacher@bundesbank.de

‡The paper contains the views of the authors and not necessarily those of the Bundesbank.

genes usually determine one or only a few of the biological factors (West 2003, Lucas et al. 2006). This framework is also of interest for economic analysis. In recent times the practice of including as much data as available or using the highest possible disaggregation level in sectoral analysis has become standard in econometric analysis, irrespectively of whether the focus lies on economic stance evaluation or forecasting (Forni et al. 2000, Stock and Watson 2002).

Specifying a sparse factor model for large economic datasets brings about valuable advantages. First, the inference on the factor loading matrix allows an explicit interpretation of the factors. Given that series might be affected by only less than all estimated factors, those with non-zero loadings are those definitively relevant for the interpretation of a factor. Second, the issue of selecting the variables containing most information on the common factors driving the variables is simultaneously addressed while estimating the model. The factor loadings of irrelevant variables are estimated to be zero, yielding rows of zeros in the factor loading matrix. Third, if the focus lies on forecasting, the estimate of the sparse factor model provides evidence on whether the panel contains relevant information for a variable of interest, and specifically which variables should be retained to compute the forecast.

We work within a Bayesian framework and pursue the parametric approach of Lucas et al. (2006). We extend the specification to a dynamic factor model with sparse factor loading matrix. Sparsity is induced by implementing a two-layer prior, each layer of which is a mixture of a Dirac delta function with infinite mass at zero and of, respectively, a normal and a Beta distribution for the factor loadings and the variable-specific probability of a non-zero loading. The posterior update then yields either variable and factor specific zero loadings or loadings shifted away from zero. Another avenue is proposed in Yoshida and West (2010) who estimate the loadings by solving an eigenvalue problem taking into account the sparse structure of the covariance matrix of the data.

There exist alternative approaches to implement sparsity or near-sparsity in the loading matrix. In the frequentist framework, one might use a sparse eigenvalue decomposition to induce sparsity (Zou et al. 2006). Eickmeier (2005) used the varimax approach of Kaiser (1958) to obtain an interpretation of the factors. After estimating the factors by principal components, the procedure looks for a rotation yielding a maximum correlation dispersion among the columns of the loading matrix. In the Bayesian framework, Bhattacharya and Dunson (2011) use a prior which increasingly shrinks the loadings towards zero as the number of estimated factors increases. While sampling, the number of factors is estimated by setting those loadings column-wise to zero which lie within a pre-defined threshold interval around zero. In the present paper, we estimate a model that explicitly includes sparsity. As already mentioned, the explicit sparse specification allows to select the relevant variables simultaneously while estimating the model. Thus, we also circumvent the usual two-step or ad hoc procedures which are followed traditionally to address the issue of variable selection (see e.g. Forni et al. 2001, Bai and Ng 2008). The usefulness of sparse priors to find relevant variables in factor models is the focus of a companion paper, see Kaufmann and Schumacher (2012). Here, we will primarily deal with identification and estimation issues.

In gene expression analysis, a high degree of sparsity is usually present and already delivers a unique factor identification. In economic data, the degree of sparsity is uncertain and might obviously be lower than in other disciplines. Therefore, we deal with the identification issue and propose an alternative to the traditional ones used in the literature. Geweke and Zhou (1996) and Aguilar and West (2000) work with restrictions on the upper diagonal of the factor loading matrix. Frühwirth-Schnatter and Lopes (2010) implement a generalized upper-diagonal zero loading matrix, in which the first non-zero loading of each factor lies on a higher row number than the ones for all previously ranked factors. Carvalho et al. (2008) pursue a heuristic approach after having estimated a "core" factor model for the variables of interest. Subsequently, variables are added which are perceived to highly correlate with the core model. Those are retained which correlate, and the "founder" of the potential additional factor is determined. Finally, Bernanke, Boivin, and Eliasz (2005) work with a restricted upper square loading matrix to identify a unique rotation of the factors. In all works but Frühwirth-Schnatter and Lopes (2010) and Carvalho et al. (2008), variable ordering is not perceived as an issue. Nevertheless, it is conceivable that the first ranked variables in the panel[1], the factor loadings of which are restricted and therefore are relevant for factor identification, might in fact be the relevant ones for only fewer than all assumed factors. It is conceivable that the first ranked variables might not contain any information at all about the factors. If N, the number of included variables is large, the introduced bias in the estimated k, $k << N$, factors might be small, if the remaining $N - k$ variables contain enough information about the common factors. Nevertheless, the bias is reduced if the factor founders are ranked first (Lucas et al. 2006, Frühwirth-Schnatter and Lopes 2010, Bai and Ng 2011).

We propose an identification scheme which is independent of variable ordering. The Markov chain Monte Carlo (MCMC) sampling schemes we design switches between two observationally equivalent factor representations. First, the $N \times k$ sparse factor loading matrix λ^* is estimated freely, under a $k \times k$ identity covariance matrix of the factor innovations as identifying restriction, $\Sigma_{\eta^*} = I$. Second, the factors are identified and sampled under the restriction of a semi-orthogonal factor loading matrix λ, $\lambda'\lambda = D$, where D is diagonal with elements ordered in descending magnitude. The factor loading matrix λ is obtained from the eigenvalue decomposition of $\lambda^{*\prime}\lambda^*$, using the eigenvectors H corresponding to the eigenvalues arranged in descending order of magnitude, $\lambda = \lambda^* H$. Given the sampled factors we switch back to the first step by transforming the system again to $\lambda^* = \lambda H'$, $f_t^* = H f_t$. For unrestricted static factor models, Bayesian estimation with order-independent identification has been derived in Chan et al. (2013).

The sampler formally identifies the factor model up to sign switching and up to trivial rotation, i.e. up to permutations of columns in the loading matrix together with corresponding factor permutation. There are two ways to obtain a unique ordering of the factors. If the columns of the loading matrix and the factors are permuted randomly while sampling, the highest (absolutely) correlated factor draws will define the "core" patterns according to which the draws are then re-ordered in a post-processing step. A second possibility is to order the factors according to the number of non-zero factor loadings while sampling. Post-processing of the draws then reduces to checking whether factors

[1]In panels with economic data, groups of data are usually deliberately ranked. For example, GDP and its components usually are ranked first, followed by a group of trade variables, then financial variables, and so on.

with an equal or similar number of non-zero loadings need additional permutations to uniquely identify the factors. Finally, overall sign identification is achieved by imposing the restriction of positive loadings for the majority of non-zero column-specific loading coefficients.

The paper is organized as follows. The next section presents briefly the model and the various observationally equivalent representations of it. Section 3 introduces sparsity and proposes an identification and estimation procedure which is independent of the variable ordering in the panel. Section 4 presents the Bayesian setup and the prior specification, in particular the two-layer sparse prior for the factor loading matrix. Section 5 describes the posterior inference. Exercises with simulated data in section 6 show the usefulness of the sampler and evaluate the performance of various prior specifications. In particular, we compare the estimation efficiency gains of using the two-layer prior for the loadings against using the one-layer and the normal prior. In section 7, we present an application to a large Swiss data set including 182 macroeconomic and price data series. Section 8 concludes. The appendix A contains details about the derivation of the posterior distributions.

2 Model specification

Assume X_t to be a $N \times 1$ vector of non-trending series observed in quarter $t = 1, \ldots, T$. Typically, when N is large, X_t may have a factor representation

$$\begin{aligned} X_t &= \lambda^* f_t^* + \xi_t \\ \Phi^*(L) f_t^* &= \eta_t^*, \quad \eta_t^* \sim N\left(0, \Sigma_{\eta^*}\right) \\ \Psi(L)\xi_t &= \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \Sigma_\varepsilon\right) \end{aligned} \tag{1}$$

where the factor loading matrix λ^* is $N \times k$. The processes $\Phi^*(L)$ and $\Psi(L)$ are of order p and q, respectively. The common dynamics in the series are captured by the k factors f_t^*, where $k << N$. The vector autoregressive (VAR) process $\Phi^*(L)$ is left unrestricted. However, the factor innovations are assumed to be orthogonal, Σ_{η^*} is diagonal. Given that the common dynamics are captured by the factors, the process of the idiosyncratic components ξ_t are independent of each other. Therefore, the process $\Psi(L)$ and the covariance matrix Σ_ε are diagonal, with element $\psi_i(L)$ and σ_i^2, respectively. Finally, the common and idiosyncratic components are uncorrelated, $E\left(\eta_t^* \varepsilon_t\right) = 0$.

The * indicates that the loading matrix is sparse, i.e. containing (some or many) zeros in the columns and even rows of zero loadings. The non-zero loadings in columns indicate the variables on which the factors load. These variables potentially define the factor, i.e. they potentially assign an interpretation to the factor. For instance, the factor which loads on many real variables may reflect the business cycle, while another factor which mainly loads on trade data may reflect external conditions. Rows of zero loadings indicate variables that are irrelevant for factor estimation. Thus, allowing for explicit zero factor loadings enables to perform variable selection simultaneously while estimating the factor model.

The estimation of the factor structure is based on the covariance structure of the data:

$$Cov(X) = \lambda^* Cov(f^*)\lambda^{*\prime} + Cov(\xi), \tag{2}$$

where $Cov(f^*)$ is non-diagonal given the $VAR(p)$ process and $Cov(\xi)$ is diagonal. The representation in (1) will still be satisfied for any non-singular matrix H (Lawley and Maxwell 1971, chapter 2):

$$\begin{aligned} X_t &= \lambda^* H H^{-1} f_t^* + \Psi(L)\xi_t = \lambda f_t + \Psi(L)\xi_t \\ \Phi(L) f_t &= \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta) \end{aligned} \tag{3}$$

in which $\Phi(L)$ is adjusted to $\Phi(L) = H^{-1}\Phi^*(L)H$ and $\eta_t = H^{-1}\eta_t^*$. Therefore, to identify the model, k^2 restrictions are necessary. By setting $\Sigma_{\eta^*} = I$ equal to the identity matrix, we implement $k(k+1)/2$ restrictions. A usual approach to implement the remaining $k(k-1)/2$ restrictions is to set to zero the upper-diagonal elements of the leading $k \times k$ submatrix in λ^*. This however needs a procedure to define the k leading relevant variables, in particular if the loading matrix is sparse (Carvalho et al. 2008, Frühwirth-Schnatter and Lopes 2010). Here, however, we propose to implement a procedure which identifies the factors independently of variable ordering. We achieve this by imposing the remaining restrictions on the vector product of the matrix λ, i.e. implementing $\lambda'\lambda = D$, D being diagonal with elements arranged in decreasing order of magnitude, while preserving the identity matrix of the factor innovation covariance matrix $\Sigma_\eta = I$. See below for further arguments and the exact implementation.

To obtain a solution to (2), we equate the number of left-hand side equations to the number of free parameters we want to estimate ($p \geq 2$, $q \geq 0$):

$$\begin{aligned} N(N+1)/2 &= Nk + N + Nq + pk^2 + k - k^2 \\ s &= (1-p)k^2 - k(N+1) + N(N-2q-1)/2 \end{aligned}$$

For large N, s will usually be positive. In this case, a factor representation usually is not trivial and the covariance structure of the data, $Cov(X)$ needs to feature some s constraints for a factor representation to exist, (Anderson and Rubin 1956). The solution for k under $s \leq 0$, in which case a unique factor representation (for $s = 0$) or a factor representation with potentially an infinite choice of λ and $Cov(\xi)$ (for $s < 0$) in general exists, yields

$$k \geq \frac{(N+1) - \sqrt{(N+1)^2 - 2N(1-p)(N-1-2q)}}{2(1-p)}$$

For large N, the minimum k turns out to be quite large. We reproduce some values for k for given N, assuming $p = q = 2$ (see also Anderson and Rubin 1956):

$N =$	5	10	50	150	300
$k =$	0	1.9	16.6	53.2	108.2

Given that in general $k << N$, in the following we proceed by assuming that the covariance structure $Cov(X)$ allows a factor representation as in (1).

Assuming λ^* to be sparse induces zero loadings. If the pattern of estimated zero loadings is such that it is destroyed by any non-singular transformation of the factors, excluding trivial column switching, then λ^* and $Cov(\xi)$ are uniquely determined (Lawley and Maxwell 1971). In this case, for given k, we interpret the sparse factor model as the most parsimonious factor representation of the data X.

Finally, note that the representation also covers the case where lagged factors enter the observation equation in (1) (see also Peña and Poncela 2006). Assume that:

$$\begin{aligned} X_t &= \lambda^* \tilde{\lambda}(L) \tilde{f}_t + \Psi(L)\xi_t \\ \tilde{\Phi}(L)\tilde{f}_t &= \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta). \end{aligned} \tag{4}$$

The specification in (1) can be recovered with $f_t^* = \tilde{\lambda}(L)\tilde{f}_t$ and $\Phi^*(L) = \tilde{\Phi}(L)\tilde{\lambda}(L)^{-1}$.

3 Sparsity and identification

As shown in the previous section, the different parametrization in (1) and (3) are observationally equivalent. To identify the model (up to sign identification), we will choose H in such a way that $\lambda'\lambda = D$ and $\Sigma_\eta = I$. Usual identification schemes would set the upper diagonal elements of the factor loading matrix λ^* equal to zero, $\lambda^*_{ij} = 0$, together with $\Sigma_{\eta^*} = I$. for $j > i$ (Aguilar and West 2000, Geweke and Zhou 1996).

Generally, if the factor loading matrix has full rank and k is properly chosen, variable ordering does not affect theoretical model identification, given that for any variable permutation BX_t, we obtain an observationally equivalent factor model with factor loading matrix $B\lambda^*$, which may not have the lower triangular form. With an appropriately chosen orthonormal matrix O, $B\lambda^*O$ will again be lower triangular and the factors Of_t^* will have the same probability distribution as the factors in the original factor model (Lopes and West 2004). Note first however, that the sparse structure in λ^* will usually get lost by rotation, hence λ^*O will thus not feature a sparse structure anymore. Second, in particular in case of a sparse loading matrix, variable ordering may be an issue for model estimation and for estimation of the number of factors. When working with the zero upper diagonal identification scheme, the first k leading variables obtain a considerable weight in determining the factors.

To illustrate this, assume that for a given ordering of N variables,[2] the corresponding sparse factor loading matrix λ^* in specification (1) would be λ^* in (5) while we would use

[2]Large panels of economic data are usually clustered, ranking first real variables like GDP and its components, then adding trade data, price data, financial variables etc.

the upper-zero diagonal identification in $\tilde{\lambda}^*$, with $\tilde{\lambda}^*_{jj} > 0$, to estimate it.

$$\lambda^* = \begin{bmatrix} \lambda^*_{11} & 0 & \dots & \dots & \lambda^*_{1k} \\ 0 & 0 & \dots & & 0 \\ 0 & \lambda^*_{32} & 0 & \dots & 0 \\ \vdots & \vdots & & & \vdots \\ \dots & \dots & 0 & \lambda^*_{k-1,k-1} & \lambda^*_{k-1,k} \\ 0 & \dots & \dots & \lambda^*_{k,k-1} & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & & 0 \\ \lambda^*_{N1} & \dots & \dots & & \lambda^*_{Nk} \end{bmatrix}, \; \tilde{\lambda}^* = \begin{bmatrix} \tilde{\lambda}^*_{11} & 0 & \dots & \dots & 0 \\ \tilde{\lambda}^*_{21} & \tilde{\lambda}^*_{22} & 0 & \dots & 0 \\ \vdots & \vdots & & & \vdots \\ \tilde{\lambda}^*_{k1} & \tilde{\lambda}^*_{k2} & \tilde{\lambda}^*_{k3} & \dots & \tilde{\lambda}^*_{kk} \\ \tilde{\lambda}^*_{k+1,1} & \tilde{\lambda}^*_{k+1,2} & \tilde{\lambda}^*_{k+1,3} & \dots & \tilde{\lambda}^*_{k+1,k} \\ \vdots & \vdots & & & \vdots \\ \tilde{\lambda}^*_{N1} & \dots & \dots & & \tilde{\lambda}^*_{Nk} \end{bmatrix} \tag{5}$$

In λ^*, the second-ranked variable does not provide any information on the factors and the second factor loads on the third, but not on the second variable. Working with $\tilde{\lambda}^*$ would be inappropriate without re-ordering the variables.

In the literature using the upper-zero diagonal identification scheme for sparse factor models, several approaches exist to determine the relevant k leading variables. Carvalho et al. (2008) resolve the issue by proposing a heuristic approach to determine the number of factors and the selection and ordering of the k leading variables. First, they estimate a "core" factor model for the variables of interest. Subsequently, they add variables which they perceive to highly correlate with the core model. Those which correlate are retained, and the "founder" of the potential new factor is determined. Frühwirth-Schnatter and Lopes (2010) propose a generalized upper-zero triangular identification scheme, in which $r_1 < \dots < r_k$, where r_j denotes the row of the top non-zero entry in $\tilde{\lambda}^*$, such that $\tilde{\lambda}^*_{l_i,j} \neq 0$, $l_i \geq r_j$ and $\tilde{\lambda}^*_{ij} = 0,\ i < l_i$. Their sampling scheme is designed in a way to explore at the same time appropriate orderings and the relevance of the variables, and various factor dimensions as well.

Here, we propose to identify and estimate the sparse factor model independently of the variable ordering. To this end, while estimating the model, we will switch between the specifications (1) and (3):

$$\lambda^* f^*_t = (\lambda^* H)\left(H^{-1} f_t\right) = \lambda f_t \tag{6}$$

The sparse matrix λ^* is estimated freely, i.e. independently of the variable ordering. Conditional on λ^*, the system is rotated into specification (3), in which the factors are estimated under the identification constraints $\lambda'\lambda = D$ and $\Sigma_\eta = I$. The rotation matrix H corresponds to the eigenvectors of $\lambda^{*\prime}\lambda^*$ arranged in descending order of magnitude of the corresponding eigenvalues.

The sampler we propose below thus iterates over the following steps:

(i) Simulate λ^* from $\pi\left(\lambda^*|f^{*T}, X^T, \Psi(L), \Sigma_\varepsilon\right)$ under a sparse prior.
Transform to $\lambda = \lambda^* H$, such that $\lambda'\lambda = D$, with D diagonal containing the elements arranged in descending order of magnitude. The matrices H and D are the solution to $\lambda^{*\prime}\lambda^* = HDH'$. They respectively correspond to the eigenvectors and the eigenvalues of $\lambda^{*\prime}\lambda^*$.
Transform the dynamics to $\Phi(L) = H'\Phi^*(L)H$. The factor innovation covariance is $\Sigma_\eta = H'H = I$.

(ii) Simulate f^T from $\pi\left(f^T|X^T, \theta\right)$ under the identification $\lambda'\lambda = D$.
Transform the system back again to $f_t^* = Hf_t$

(iii) Simulate the rest of the parameters, $\Phi^*(L)$, $\Psi(L)$, Σ_ε, from $\pi\left(\theta_{-\lambda^*}|f^T, X^T, \lambda^*\right)$.

The estimation provides a statistically identified factor model up to sign switching under the specification (3) used in step (ii) and up to sign switching and trivial rotation in specification (1) used in step (i). While sampling it may well be that factors are subject to these permutations randomly. However, if the degree of sparsity is high, factors might well be identified and the sampler might stabilize at one mode of the posterior. This is the outcome of the simulation study, which also shows that the number of factors is rightly inferred by the sampler, yielding mean zero columns in the loading matrix for the factors estimated redundantly.

In empirical work, to control for column switching in a first round, having sampled λ^* we might arrange the factors according to the number of non-zero column-specific loadings. This procedure also helps in determining the appropriate number of factors. In post-processing the simulations of the factors, the researcher might have to re-arrange some draws of the factors according to the identified factor patterns (see next paragraph). This may be the case if the number of non-zero loadings do not differ a lot across factors. In a final step, to ensure sign identification, we require the majority of non-zero loadings of each factor to be positive, $\left(\sum_{i=1}^N I_{\{\lambda_{ij}^*>0\}} / \sum_{i=1}^N I_{\{\lambda_{ij}^*\neq 0\}}\right) > 0.5\ \forall j$. If the condition is violated the specific loading and factor draws are multiplied by -1, and also the corresponding rows and columns of $\Phi^*(L)$.

Alternatively, one might randomly trivially rotate the factors f_t^*, the factor loadings λ^* and the factor-specific parameters, and randomly switch factor signs at the end of each sweep of the sampler. This ensures that the whole posterior is explored. In post-processing the MCMC output, the relevant factor patterns can be inferred by grouping the highly absolutely correlated factor draws and rearranging each factor draw according to the identified patterns. This procedure is similar to the identification procedure by $k-means$ clustering motivated in Früwirth-Schnatter (2011). In a final step, sign identification is again achieved by ensuring column-wise that the majority of non-zero factor loadings is positive.

4 Bayesian specification

4.1 Likelihood and prior specification

The complete-data likelihood for specification (1) takes the form

$$L\left(X^T|f^{*T},\theta\right)=\prod_{t=1}^{T}\pi\left(X_t|f^{*t},\theta\right), \tag{7}$$

where $X^t=(X_t,X_{t-1},\ldots,X_1)$ denotes all observations up to period t. The parameter θ includes all model parameters to be estimated of the specification in (1), $\theta=(\lambda^*,\Phi^*,\Psi,\Sigma_\varepsilon)$, given that $\Sigma_{\eta^*}=I$. The observation density in (7) is multivariate normal

$$\pi\left(X_t|f^{*t},\theta\right)=\frac{1}{(2\pi)^{N/2}|\Sigma_\varepsilon|^{1/2}}\exp\left\{-\frac{1}{2}\left(\Psi(L)(X_t-\lambda^*f_t^*)\right)'\Sigma_\varepsilon^{-1}\left(\Psi(L)(X_t-\lambda^*f_t^*)\right)\right\} \tag{8}$$

The prior density of the unobserved factors is

$$\pi\left(f^{*T}|\theta\right)=\prod_{t=p+1}^{T}\pi\left(f_t^*|f^{*t-1},\theta\right)\pi\left(f^{*p}|\theta\right) \tag{9}$$

where f^{*p} contains the initial states $f^{*p}=\left(f_p^*,\ldots,f_1^*,f_0^*\right)$.

For the parameters we assume independent priors,

$$\pi\left(\lambda^*,\Phi^*,\Psi,\Sigma_\varepsilon\right)=\pi\left(\lambda^*\right)\pi\left(\Phi^*\right)\pi\left(\Psi\right)\pi\left(\Sigma_\varepsilon\right), \tag{10}$$

where the hierarchical prior $\pi(\lambda^*)$ has a spike and slab specification:

$$\pi\left(\lambda_{ij}^*\right)=(1-\beta_{ij})\delta_0\left(\lambda_{ij}^*\right)+\beta_{ij}N\left(0,\tau_j\right) \tag{11}$$

$$\pi\left(\beta_{ij}\right)=(1-\rho_j)\delta_0\left(\beta_{ij}\right)+\rho_jB\left(a_jb_j,a_j(1-b_j)\right) \tag{12}$$

$$\pi\left(\rho_j\right)=B\left(r_{0j}s_{0j},r_{0j}(1-s_{0j})\right) \tag{13}$$

where the beta distribution $B((ab,a(1-b))$ has mean b and variance $b(1-b)/(1+a)$. The prior for ρ_j favours very small values such that $r>0$ is large and s is a small probability. The function $\delta_0(\cdot)$ is the Dirac delta function at zero.

The prior (11)-(13) implies a common probability $1-\rho_jb_j$ across variables of a zero loading on factor j. In addition, the layer (12) reflects the viewpoint that for many variables, the probability of association with anyone factor is zero, while for a few it will be high. The hierarchical prior circumvents the observed problem that uncertainty about the significance of the loadings increases with increasing N, a feature that arises when working with a prior assuming a common base rate ρ_j across series to load on factor j (Lucas et al. 2006).

Frühwirth-Schnatter and Lopes (2010) use a hierarchical prior on the factor loadings, in which the loading significance is governed by an indicator $\delta_{ij}=\{0,1\}$, with $P\left(\delta_{ij}=1|\rho_j\right)$.

Table 1: Kurtosis in $\pi(\lambda^*)$ under various parametrizations of $\pi(\tau) \sim IG(g, G)$ and $\pi(\rho) \sim B(r_0 s_0, r_0(1-s_0))$ for $\pi(\beta) \sim B(0.9, 0.1)$

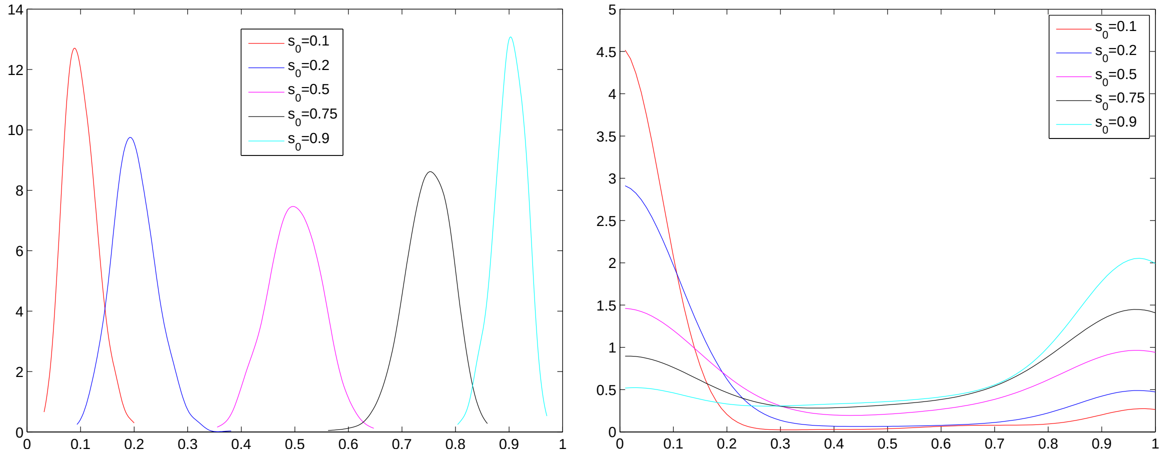
	$r_0 = 100$				
$g = 2$	$s_0 = 0.1$	$s_0 = 0.2$	$s_0 = 0.5$	$s_0 = 0.75$	$s_0 = 0.9$
$G = 0.18$	89.69	47.70	26.52	19.89	14.75
$G = 0.5$	85.53	47.88	27.15	18.57	17.23
$G = 1$	88.40	48.81	25.36	18.55	14.77

This corresponds to a one-layer sparse prior specification (see below section 6.3). The marginal two-layer sparse prior with expected inclusion probability of $\rho_j b_j$ is implementable by restricting ρ_j to lie in the interval $[0\ b_j]$. Still, for large N the two-layer specification alleviates the disadvantage of the one-layer specification manifesting itself in large standard deviations of factor loadings' marginal posterior distributions. In addition, the second layer (12) renders the variable- and factor-specific inclusion probability bimodal with mass separated either near zero or one, while the one-layer specification has a uni-modal distribution near the mean inclusion probability (see figure 1 below).

4.2 The sparse prior on λ^*

To obtain an intuition of the influence of the hyperparameters ρ_j and β_{ij} in shaping the prior distribution of λ^*_{ij}, we simulate out of the prior assuming various prior means for ρ_j, $s_0 = (0.1, 0.2, 0.5, 0.75, 0.9)$, with precision $r_0 = 100$. Figure 1, panel (a), plots the distribution for ρ under the various parameter settings. We observe that the distribution becomes increasingly skewed as the mean is shifted away from 0.5. Decreasing the precision would reinforce the pattern. Figure 1, panel (b), shows that decreasing sparsity shifts the mass of the series-specific non-zero factor loading probability towards and smoothes the bimodal shape of the distribution. With the hyperparameter constellation of $b = 0.9$ and $a = 1$, we obtain expected base rates of a non-zero factor loading of $E(\rho_j b) = (0.09, 0.18, 0.45, 0.68, 0.81)$. In the most sparse setup, on average 10% of the $\lambda^*_{\cdot j}$ will have an expected 90% probability of a non-zero value. The resulting fat-tailed prior distribution for λ^* is depicted in figure 2, where, from left to right, the prior for τ, $IG(g, G)$ is parameterized by $g = 2$ and $G = (0.18, 0.5, 1)$, yielding an expected value $E(\tau) = (0.09, 0.25, 0.5)$, respectively. Conditional on the parameterization for $\pi(\tau)$, decreasing sparsity renders $\pi(\lambda^*))$ less peaked. This is documented in table 1, which reproduces the kurtosis in $\pi(\lambda^*)$ under the various prior parameterizations. Given the parameterization for $\pi(\rho)$, becoming more diffuse in $N(0, \tau)$ apparently renders $\pi(\lambda^*))$ less peaked. However, table 1 shows that the kurtosis generally is not very sensitive to changes in G. The differences are largest for the most sparse specifications.

Figure 1: Prior distributions $\pi(\rho)$ and $\pi(\beta)$ under $r_0 = 100$, $s_0 = (0.1, 0.2, 0.5, 0.75, 0.9)$, b=0.9 and a=1.



(a) $\pi(\rho)$ (b) $\pi(\beta)$

4.3 Prior specification for the remaining parameters

The remaining parameters have standard prior distributions. For the dynamic parameters we assume multivariate normal priors truncated to the stationary region:

$$\begin{aligned} \pi(\Phi^*) &= N(\mathrm{p}_0, \mathrm{P}_0)\, I_{\{Z(\Phi^*)>1\}} \\ \pi(\Psi) &= N(\mathrm{q}_0, \mathrm{Q}_0)\, I_{\{Z(\Psi)>1\}} \end{aligned}$$

where $I_{\{\cdot\}}$ is the indicator function and $Z(\varphi) > 1$ means that the roots of the characteristic equation of the process $\varphi(L)$ lie outside the unit circle.

Given that Σ_ε is diagonal, we assume independent inverse Gamma prior distributions for the variances:

$$\pi\left(\sigma_i^2\right) = IG(\mathrm{u}_0, \mathrm{U}_0), \quad i = 1, \ldots, N \tag{14}$$

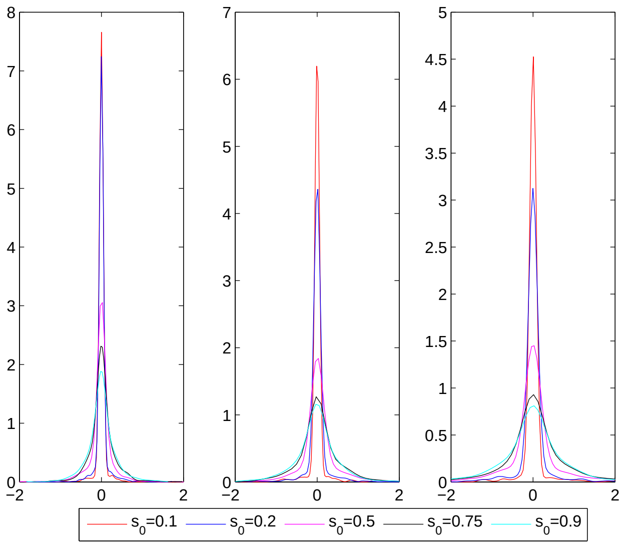
5 Posterior inference

5.1 Sampling design

Updating the prior with data information yields the inference on the posterior distribution of $\vartheta = \left(f^{*T}, \theta\right)$, $\pi\left(\vartheta | X^T\right) \propto L\left(X^T | f^{*T}, \theta\right) \pi\left(f^{*T} | \theta\right) \pi(\theta)$.

The sampler is based on the following steps

Figure 2: Prior distribution $\pi(\lambda^*)$, becoming (from left to right) more diffuse in $N(0,\tau)$ with the parametrization for $\pi(\tau) \sim IG(g, G)$ as $g = 2$ and $G = (0.18, 0.5, 1)$. Each panel plots the marginal distribution for various parametrizations in $\pi(\rho)$, $r_0 = 100$ and $s_0 = (0.1, 0.2, 0.5, 0.75, 0.9)$.



(i) Simulate λ^* from $\pi\left(\lambda^* | f^{*T}, X^T, \Psi(L), \Sigma_\varepsilon\right)$ under a sparse prior.
Transform to $\lambda = \lambda^* H$, such that $\lambda'\lambda = D$, with D diagonal containing the elements arranged in descending order of magnitude. The matrices H and D correspond to the eigenvectors and the eigenvalues of $\lambda^{*\prime}\lambda^*$, respectively.
Transform the dynamics to $\Phi(L) = H'\Phi^*(L)H$. The factor innovation covariance remains $\Sigma_\eta = H'H = I$.

(ii) Update the hyperparameters of the sparse prior.

(iii) Simulate f^T from $\pi\left(f^T | X^T, \theta\right)$ under the identification $\lambda'\lambda = D$.
Transform the system back again to $f_t^* = Hf_t$

(iv) Simulate the rest of the parameters, $\Phi^*(L)$, $\Psi(L)$, Σ_ε, from $\pi\left(\theta_{-\lambda^*} | f^{*T}, X^T, \lambda^*\right)$.

Step (i) and (ii) closely follow Carvalho (2006) and will described in more details in the next subsection. Step (iii) and (iv) are by now standard in the Bayesian simulation setup. The path f^T may be simulated using a multi-move sampler as proposed in Carter and Kohn (1994), Shephard (1994), Frühwirth-Schnatter (1994). Here however, we will implement the blocked sampling scheme proposed in Chan and Jeliazkov (2009), which takes advantage of the fact that the system matrices are banded when observables and latent variables are stacked. The interested reader finds details in appendix A. The parameter simulation needs further blocking. Given the conjugate priors, the posterior distributions are multivariate normal and inverse Wishart.

Depending on the way factor identification is implemented, the sampler is completed by either a permutation step ordering the factors f_t^* in descending order of the number of factor-specific non-zero loadings and identifying the sign of the factor such that the majority of non-zero loadings be positive; or a random column-switch and sign-switch permutation step to ensure that the full posterior distribution is being explored.

5.2 Sampling from the sparse posterior $\pi\left(\lambda_{ij}^{*}|\cdot\right)$

The posterior $\pi\left(\lambda_{ij}^{*}|f^{*T},X^{T},\Psi(L),\Sigma_{\varepsilon}\right)$ is obtained by first integrating out the variable specific prior probability of zero loading for each factor j. The prior in (11)-(13) implies a common base rate of a non-zero factor loading of $E\left(\beta_{ij}\right)=\rho_{j}b_{j}$ across variables. The marginal prior becomes

$$\pi\left(\lambda_{ij}^{*}|\rho_{j}\right)\sim(1-\rho_{j}b_{j})\delta_{0}(\lambda_{ij}^{*})+\rho_{j}b_{j}N\left(0,\tau_{j}\right) \tag{15}$$

For each factor j, transform the variables to

$$x_{it}^{*}=\psi_{i}(L)x_{it}-\sum_{l=1,l\neq j}^{k}\lambda_{il}^{*}\psi_{i}(L)f_{lt}^{*}=\lambda_{ij}^{*}\psi_{i}(L)f_{jt}^{*}+\varepsilon_{it} \tag{16}$$

which basically isolates the effect of factor j in variable i. Combine the marginal prior with data information to sample independently across i from

$$\begin{aligned}\pi\left(\lambda_{ij}^{*}|\cdot\right) &= \prod_{t=q+1}^{T}\pi(x_{it}^{*}|\cdot)\left\{(1-\rho_{j}b_{j})\delta_{0}(\lambda_{ij}^{*})+\rho_{j}b_{j}N\left(0,\tau_{j}\right)\right\} &(17)\\ &= P\left(\lambda_{ij}^{*}=0|\cdot\right)\delta_{0}(\lambda_{ij}^{*})+P\left(\lambda_{ij}^{*}\neq 0|\cdot\right)N\left(m_{ij},M_{ij}\right) &(18)\end{aligned}$$

with observation density $\pi(x_{it}^{*}|\cdot)=N\left(\lambda_{ij}^{*}\psi_{i}(L)f_{jt}^{*},\sigma_{i}^{2}\right)$ and where

$$\begin{aligned}M_{ij} &= \left(\frac{1}{\sigma_{i}^{2}}\sum_{t=q+1}^{T}\left(\psi_{i}(L)f_{jt}^{*}\right)^{2}+\frac{1}{\tau_{j}}\right)^{-1} &(19)\\ m_{ij} &= M_{ij}\left(\frac{1}{\sigma_{i}^{2}}\sum_{t=q+1}^{T}\left(\psi_{i}(L)f_{jt}^{*}\right)x_{it}^{*}\right) &(20)\end{aligned}$$

To obtain the posterior odds $P\left(\lambda_{ij}^{*}\neq 0|\cdot\right)/P\left(\lambda_{ij}^{*}=0|\cdot\right)$ we update the prior odds of non-zero factor loading:

$$\frac{P\left(\lambda_{ij}^{*}\neq 0|\cdot\right)}{P\left(\lambda_{ij}^{*}=0|\cdot\right)}=\frac{\pi\left(\lambda_{ij}^{*}\right)|_{\lambda_{ij}^{*}=0}}{\pi\left(\lambda_{ij}^{*}|\cdot\right)|_{\lambda_{ij}^{*}=0}}\frac{\rho_{j}b_{j}}{1-\rho_{j}b_{j}}=\frac{N\left(0;0,\tau_{j}\right)}{N\left(0;m_{ij},M_{ij}\right)}\frac{\rho_{j}b_{j}}{1-\rho_{j}b_{j}} \tag{21}$$

Conditional on λ_{ij}^{*} we update the variable specific probabilities β_{ij} and sample from $\pi(\beta_{ij}|\lambda_{ij}^{*},\cdot)$. When $\lambda_{ij}^{*}=0$

$$\begin{aligned}&\pi(\beta_{ij}|\lambda_{ij}^{*}=0,\cdot)\propto(1-\beta_{ij})\left[(1-\rho_{j})\delta_{0}(\beta_{ij})+\rho_{j}B\left(a_{j}b_{j},a_{j}(1-b_{j})\right)\right] &(22)\\ &P(\beta_{ij}=0|\lambda_{ij}^{*}=0,\cdot)\propto(1-\rho_{j}),\quad P(\beta_{ij}\neq 0|\lambda_{ij}^{*}=0,\cdot)\propto(1-b_{j})\rho_{j}\end{aligned}$$

That is, with posterior odds $(1-b_j)\rho_j/(1-\rho_j)$ we sample from $B\left(a_jb_j, a_j(1-b_j)+1\right)$ and set otherwise β_{ij} equal to zero. Conditional on $\lambda^*_{ij} \neq 0$ we obtain

$$\begin{aligned}
&\pi(\beta_{ij}|\lambda^*_{ij} \neq 0, \cdot) \propto \beta_{ij} N\left(\lambda^*_{ij}; 0, \tau_j\right)\left[(1-\rho_j)\delta_0(\beta_{ij}) + \rho_j B\left(a_jb_j, a_j(1-b_j)\right)\right] \\
&P(\beta_{ij}=0|\lambda^*_{ij} \neq 0, \cdot) = 0, \quad P(\beta_{ij} \neq 0|\lambda^*_{ij} \neq 0, \cdot) = 1
\end{aligned} \tag{23}$$

In this case we sample β_{ij} from $B\left(a_jb_j+1, a_j(1-b_j)\right)$.

The posterior update of the hyperparameters τ_j and ρ_j is sampled from an inverse Gamma, $\pi\left(\tau_j|\cdot\right) \sim IG\left(g_j, G_j\right)$, and a Beta distribution, $\pi\left(\rho_j|\cdot\right) \sim B\left(r_{1j}, r_{2j}\right)$, respectively, with

$$\begin{aligned}
&g_j = g_0 + \frac{1}{2}\sum_{i=1}^{N} I_{\{\lambda^*_{ij}\neq 0\}}, \quad G_j = G_0 + \frac{1}{2}\sum_{i=1}^{N} {\lambda^*_{ij}}^2 \\
&r_{1j} = r_{0j}s_{0j} + S_j, \quad r_{2j} = r_{0j}(1-s_{0j}) + N - S_j \\
&\text{where } S_j = \sum_{i=1}^{N} I_{\{\beta_{ij}\neq 0\}}
\end{aligned}$$

6 Simulations

In the following, we will address various issues that arise in factor model estimation. After describing the data generating process, we first assess the ability of the sampler in recovering the true number of factors and the true factor loading structure (subsection 6.2). Recovering the true factor loading structure goes hand in hand with excluding the right series, i.e. recovering the rows with zero factor loadings on all factors. In a second exercise we assess the estimation performance of the sampler. We compare the root mean squared estimation error for the common component across specifications using various numbers of factors and using various prior specifications on the loading matrix. In particular, besides using the two-layer sparse prior introduced in section 4, we also estimate a model using a one-layer prior and the usually used normal prior for the factor loadings. The posterior sampling distributions under the latter two specifications are described in subsection 6.3. Subsection 6.4 contains the results.

6.1 Simulation setup

We simulate $N = 100$ time series of length $T = 100$, driven by a $k = 3$ first-order autoregressive factor process. We assume $q = 1$ for the idiosyncratic processes. Table 2 depicts the two settings with different degrees of sparsity. With a high precision, $r_{0j} = 500$, we simulate data with a high degree of sparsity, $s_{0j} = (0.2, 0.2, 0.1)$, and data with a lower degree of sparsity, $s_{0j} = (0.9, 0.75, 0.5)$. Together with the hyperparameters specifying the series-specific probability of non-zero factor loading, $b_j = 0.8$ and $a_j = 0.01$, this yields respectively relatively low and high expectations of non-zero factor loadings, $E(\rho_jb_j) = (0.16, 0.16, 0.08)$ and $E(\rho_jb_j) = (0.72, 0.6, 0.4)$. On average, in case of a high degree of sparsity, 65% of the series will have zero factor loadings on all factors, and in case of a low degree of sparsity, this will be expected for only 7% of the series. The non-zero factor loadings are simulated out of normal distributions $N(m_j, M)$, see table 3. The variance $M = 0.01$ is chosen relatively tight in order to well separate the group of time series on which each factor loads. In this way, we intend to maximize the estimation performance of the sampler based on the normal prior for the factor loadings. The gains in estimation performance obtained by using the sparse prior specification might thus be interpreted as a lower bound of potential gains.

Each autoregressive coefficient of the idiosyncratic components, ψ_i, is simulated out of a normal distribution $N(0, 0.09)$, while the variance of the idiosyncratic error terms is fixed to $1 - 1/3\left(0.91^2 + 0.75^2 + 0.64^2\right)$ (see table 3). The hyperparameter and parameter constellation is chosen such that each factor, if solely loading on a time series and taking into account its autoregressive process, approximately accounts for a variance share lying between 65% and 74%, see table 4.

Assigning the same importance to each factor (on its own), provided that it loads on a time series, in our view renders the following simulation exercise less dependent on the importance of the common components, hence less dependent on differences between factors in signal to noise ratios. We want to evaluate the performance of the estimation when a lower and a larger number of factors than appropriate is assumed to drive the variables and whether the advantage of the two-layer sparsity prior persists also in the case of a decreasing degree of sparsity.

Table 2: Hyperparameter settings in simulations and estimations, with factor-specific expected probability of non-zero factor loadings, $E(\rho_j b_j)$. Expected share of zero factor loading rows is $r0$.

$b_j = 0.8,\ a_j = 0.01$	$s_{0j} =$	
$r_{0j} = 500$	(0.2,0.2,0.1)	(0.9,0.75,0.5)
$E(\rho_j b_j)$	(0.16,0.16,0.08)	(0.72,0.6,0.4)
$r0 = \prod_{j=1}^{3} E(P(\lambda_{ij} = 0))$	0.65	0.07
Estimation prior		
$r_{0j} = 50$	(0.2,0.2,0.1)	(0.9,0.75,0.5)
	(0.3,0.3,0.3)	(0.75,0.75,0.75)

To sum up, the simulated data generating process takes the form:

$$X_{it} = \lambda_i^* f_t^* + \xi_{it} \tag{24}$$

$$f_t^* = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{bmatrix} f_{t-1}^* + \eta_t^*,\ \eta_t^* \sim N(0, I) \tag{25}$$

$$\xi_{it} = \psi_i \xi_{i,t-1} + \varepsilon_{it},\ \varepsilon_{it} \sim N\left(0, \sigma_i^2\right) \tag{26}$$

where the non-zero factor loadings in λ_i^* and the coefficients ψ_i, $i = 1, \ldots, N$, are simulated out of the normal distributions depicted in table 3.

Table 3: Parameter settings of the simulation distributions

Parameters	Setting, distribution
$\lambda_{ij}^*,\ i=1,\dots N$	$N\left(m_j, M\right)$
	$M=0.01,\ m_1=0.91,\ m_2=0.75,\ m_3=0.64$
$\psi_i,\ i=1,\dots N$	$N\left(0, 0.09\right)$
σ_i^2	$1-1/3\left(0.91^2+0.75^2+0.64^2\right)$

Table 4: Autoregressive factor process and approximate variance share

$$\Phi^* = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{bmatrix},\ \Sigma_{\eta^*} = I$$

	Approximate signal variance $Var\left(\lambda_{\cdot j}^* f_{jt}^*\right)$	Approx. variance share of common component
Factor 1	0.91	0.69
Factor 2	0.75	0.65
Factor 3	1.14	0.74

For each of the sparsity degrees depicted in table 2 we produce 10 replications of data. The dynamic factor model is then estimated for $k=2,3,4$ factors. To evaluate the estimation performance we additionally estimate the factor models under the one-layer sparsity prior and the normal prior for the factor loadings. We draw 6,000 times from the posterior, discard the first 2,000 draws and retain every second one to evaluate the performance in recovering the true factor loading structure and to compute the estimation error statistics and the estimation precision of non-zero factor loadings. 6,000 draws might seem a low number. However, it turns out that the sampler converges quite quickly under both, the informative and the less informative, prior settings for the factor loadings (see table 2, bottom panel) and under the informative priors for the rest of the parameters (see table 5).

6.2 Series exclusion and factor structure

Table 6 depicts the average exclusion error and the average error in estimating the factor structure. The average absolute exclusion error (AEE) computes the excess number of (rightly excluded) series that have all zero factor loadings – working with the median of the posterior, $\bar{\lambda}^*$, while the relative (REE) exclusion error sets the absolute exclusion error

Table 5: Prior specification in the simulation exercise.

Distribution	Hyperparameter setting
Sparse prior (ρ_j, β_{ij})	$r_{0j} = 50$, s_{0j}: various specifications, see table 2
	$b_j = 0.8$, $a_j = 0.01$
Factor loading (τ_j)	$h_0 = 2$, $H_0 = 0.25$
Factor autoregression (Φ)	$p_0 = 0$
	P_0: Minnesota with prior diagonal variance 0.16
	shrink factor for off-diagonals 0.078
Idiosyncratic autoregression (ψ_i)	$q_0 = 0$, $Q_0 = 0.16$
Idiosyncratic error variance (σ_i^2)	$u_0 = 2$, $U_0 = 1$

in relation to the number of truely excluded series in the panel, $N_0^{(r)} = \sum_{i=1}^{N} I_{\{\lambda_{ik}^{*(r)}=0,\forall k\}}$:

$$\begin{aligned}
\text{AEE} &= \frac{1}{R}\sum_{r=1}^{R}\left(\sum_{i=1}^{N} I_{\{\bar{\lambda}_{ik}^{*(r)}=0,\forall k\}} I_{\{\lambda_{ik}^{*(r)}=0,\forall k\}}\right) - N_0^{(r)} \\
\text{REE} &= \frac{1}{R}\sum_{r=1}^{R}\frac{1}{N_0^{(r)}}\left[\left(\sum_{i=1}^{N} I_{\{\bar{\lambda}_{ik}^{*(r)}=0,\forall k\}} I_{\{\lambda_{ik}^{*(r)}=0,\forall k\}}\right) - N_0^{(r)}\right]
\end{aligned} \tag{27}$$

For the specification $k = 2$, the first line reports the overall exclusion error, i.e. AEE and REE without multiplying with $I_{\{\lambda_{ik}^{*(r)}=0,\forall k\}}$. We observe that assuming too few factors leads to a larger estimated number of excluded series. Nevertheless, usually the truely excluded series are rightly estimated to be excluded. For example, given a high degree of sparsity and a sparse prior, an excess of 6 series is excluded – which amounts on average to 10% of $N_0^{(r)}$ – while basically all of the truely excluded series are rightly excluded – see the second line corresponding to $k = 2$. For the factor specification $k = 3, 4$ we only report the AEE and REE, because the numbers are the same for the overall exclusion error. We observe that up to 1 series which should be is usually not excluded. With decreasing sparsity, 1 series accounts for a larger share of excluded series, e.g. to 13% of truely excluded series.

Table 6 also reports the average factor structure error, again in absolute numbers of loadings and relatively to N:

$$\frac{1}{NR}\sum_{r=1}^{R}\left(\mathbf{I}'_{\{\bar{\lambda}_{ik}^{*(r)}\neq 0\}} \cdot \mathbf{I}_{\{\lambda_{ij}^{*(r)}\neq 0\}}\right) \tag{28}$$

where $\mathbf{I}_{\{\lambda_{ij}^{*(r)}\neq 0\}}$ is a $N \times k$ 0-1 matrix with 1's indicating the non-zero factor loadings. Numbers on the diagonal in table 6, m_{jj}, indicate the number of non-zero factor loadings for factor j, off-diagonal numbers, m_{jl} indicate how many non-zero loadings the factors j and l have in common. Given that the matrices are symmetric, we only report the lower diagonal part.

Overall, the true number of factors is detected. The specification with $k = 4$ on average yields a column of zero factor loadings or a column with less than three non-zero factor loadings. In case of a high degree of sparsity, the loading structure is recovered with less than 5% error. It pays off to reduce the information degree of the prior, see the panels with $r_{0j} = 50$ and, respectively $s_{0j} = (0.3, 0.3, 0.3)$ and $s_{0j} = (0.75, 0.75, 0.75)$. The estimation error is reduced, particularly when the degree of sparsity is low and according to the simulations when $k = 4$.

Table 6: Exclusion error and loading structure error for data with a high and a low degree of sparsity.

Specification	Exclusion error	Factor structure error	
	AEE/REE	absolute error	relative to $N = 100$
Simulation: $r_{0j} = 500,\ s_{0j} = (0.2, 0.2, 0.1)$			
– Estimation prior: $r_{0j} = 50,\ s_{0j} = (0.2, 0.2, 0.1)$			
$k = 2$	6.2/0.10		
	-0.2/-0.00		
$k = 3$	-0.6/-0.01	$\begin{bmatrix} 2.1 & & \\ 4.4 & 2.8 & \\ 0.1 & 0.2 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0.02 & & \\ 0.04 & 0.03 & \\ 0.0 & 0.0 & 0.0 \end{bmatrix}$
$k = 4$	-0.4/-0.01	$\begin{bmatrix} 0.2 & & & \\ 0 & 0.1 & & \\ 0.2 & 0.1 & 0.3 & \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.0 & & & \\ 0 & 0.0 & & \\ 0.0 & 0.0 & 0.0 & \\ 0 & 0 & 0 & 0 \end{bmatrix}$
– Estimation prior: $r_{0j} = 50,\ s_{0j} = (0.3, 0.3, 0.3)$			
$k = 2$	5.3/0.08		
	-0.2/-0.00		
$k = 3$	-0.5/-0.01	$\begin{bmatrix} 0.2 & & \\ 0.3 & 0.3 & \\ 0 & 0.2 & 0.4 \end{bmatrix}$	$\begin{bmatrix} 0.0 & & \\ 0.0 & 0.0 & \\ 0 & 0.0 & 0.0 \end{bmatrix}$
$k = 4$	-0.4/-0.01	$\begin{bmatrix} 0.2 & & & \\ 0.2 & 0.2 & & \\ 0 & 0.1 & 0.3 & \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.0 & & & \\ 0.0 & 0.0 & & \\ 0 & 0.0 & 0.0 & \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Simulation: $r_{0j} = 500,\ s_{0j} = (0.9, 0.75, 0.5)$			
– Estimation prior: $r_{0j} = 50,\ s_{0j} = (0.9, 0.75, 0.5)$			
$k = 2$	0.13/0.10		
	-0.8/-0.15		

A diagonal figure f_{jj} indicates the number of non-zero loadings for factor j, the off-diagonal figure f_{jl} indicates the number of common non-zero loadings for factor j and l.

Table 6: Exclusion error and loading structure error, continued.

Specification	Exclusion error AEE/REE	Factor structure error absolute error	relative to $N = 100$
$k = 3$	-0.63/-0.13	$\begin{bmatrix} 8.4 & & \\ 18.3 & 13.3 & \\ 10.0 & 15.9 & 5.8 \end{bmatrix}$	$\begin{bmatrix} 0.08 & & \\ 0.18 & 0.13 & \\ 0.10 & 0.16 & 0.06 \end{bmatrix}$
$k = 4$	-0.63/-0.13	$\begin{bmatrix} 9.1 & & & \\ 15.1 & 9.6 & & \\ 5.0 & 9.3 & 1.9 & \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.09 & & & \\ 0.15 & 0.10 & & \\ 0.05 & 0.09 & 0.02 & \\ 0 & 0 & 0 & 0 \end{bmatrix}$
– Estimation prior: $r_{0j} = 50,\ s_{0j} = (0.75, 0.75, 0.75)$			
$k = 2$	0.4/0.07 -0.4/-0.05		
$k = 3$	-1.4/-0.27	$\begin{bmatrix} 9.0 & & \\ 16.4 & 8.5 & \\ 13.3 & 15.5 & 12.4 \end{bmatrix}$	$\begin{bmatrix} 0.09 & & \\ 0.16 & 0.09 & \\ 0.13 & 0.16 & 0.12 \end{bmatrix}$
$k = 4$	-1.1/-0.24	$\begin{bmatrix} 7.9 & & & \\ 12.6 & 5.4 & & \\ 8.0 & 6.8 & 6.6 & \\ 2.3 & 1.5 & 1.5 & 2.9 \end{bmatrix}$	$\begin{bmatrix} 0.08 & & & \\ 0.13 & 0.05 & & \\ 0.08 & 0.07 & 0.07 & \\ 0.02 & 0.02 & 0.02 & 0.03 \end{bmatrix}$

A diagonal figure f_{jj} indicates the number of non-zero loadings for factor j, the off-diagonal figure f_{jl} indicates the number of common non-zero loadings for factor j and l.

6.3 Relaxing the sparse prior

In the next section we will evaluate the performance of the sparse factor model using a prior loading specification with two layers against the factor model estimated under a one-layer prior loading specification and against the factor model estimated under the widely used normal prior (Bernanke, Boivin and Eliasz, 2005, inter alia). Therefore, in this section we derive the posterior distributions to sample from when the two-layer sparse prior is relaxed to a one-layer sparse prior and to a normal prior.

The one-layer sparse prior assumes that there is a common probability across units of zero loading on factor j:

$$\pi\left(\lambda_{ij}^{*}\right) = (1-\rho_j)\delta_0\left(\lambda_{ij}^{*}\right) + \rho_j N\left(0, \tau_j\right) \tag{29}$$

$$\pi\left(\rho_j\right) = B\left(r_{0j}s_{0j}, r_j(1-s_{0j})\right) \tag{30}$$

where, as in (13), s_{0j} is a small expected probability of non-zero factor loading and r_{0j} is large. The posterior for λ_{ij}^{*} is basically governed by the same moments as derived in

(18)–(20), with adjusted posterior odds of zero factor loading, however:

$$\frac{P\left(\lambda_{ij}^{*}\neq 0|\cdot\right)}{P\left(\lambda_{ij}^{*}=0|\cdot\right)}=\frac{\pi\left(\lambda_{ij}^{*}\right)|_{\lambda_{ij}^{*}=0}}{\pi\left(\lambda_{ij}^{*}|\cdot\right)|_{\lambda_{ij}^{*}=0}}\frac{\rho_j}{1-\rho_j}=\frac{N\left(0;0,\tau_j\right)}{N\left(0;m_{ij},M_{ij}\right)}\frac{\rho_j}{1-\rho_j} \tag{31}$$

The posterior of hyperparameter ρ_j is in this case $\pi\left(\rho_j|\cdot\right)\sim B\left(r_{1j},r_{2j}\right)$ with

$$r_{1j}=r_{0j}s_{0j}+S_j,\quad r_{2j}=r_{0j}(1-s_{0j})+N-S_j$$
$$\text{where } S_j=\sum_{i=1}^{N}I_{\{\lambda_{ij}^{*}\neq 0\}}$$

Under a normal prior for $\lambda_i^{*}=(\lambda_{i1}^{*},\ldots,\lambda_{iK}^{*})'$,

$$\pi\left(\lambda_i\right) \sim N\left(0,\tau\right) \quad \tau=\begin{bmatrix}\tau_1 & & 0\\ & \ddots & \\ 0 & & \tau_k\end{bmatrix} \tag{32}$$

$$\pi\left(\tau_j\right) \sim IG\left(g_0,G_0\right), \tag{33}$$

we may sample independently over i using the transformation

$$x_{it}^{*} = \psi_i(L)x_{it}=\sum_{j=1}^{k}\lambda_{ij}^{*}\psi_i(L)f_{kt}^{*}+\varepsilon_{it},\ \varepsilon_{it}\text{ i.i.d } N\left(0,\sigma_i^2\right) \tag{34}$$

from the posterior

$$\pi\left(\lambda_i|\cdot\right) = N\left(m_i,M_i\right) \tag{35}$$

$$M_i = \left(\frac{1}{\sigma_i^2}F_i^{*\prime}F_i^{*}+\tau^{-1}\right)^{-1} \tag{36}$$

$$m_i = M_i\left(\frac{1}{\sigma_i^2}F_i^{*\prime}X_i^{*}\right) \tag{37}$$

where F_i^{*} and X_i^{*} are, respectively, the predictor matrix and the vector of transformed variables in equation (34):

$$X_i^{*}=\begin{bmatrix}x_{i,q+1}^{*}\\ \vdots\\ x_{iT}^{*}\end{bmatrix},\qquad F_i^{*}=\begin{bmatrix}\psi_i(L)f_{1,q+1}^{*} & \cdots & \psi_i(L)f_{k,q+1}^{*}\\ \vdots & & \vdots\\ \psi_i(L)f_{1,T}^{*} & \cdots & \psi_i(L)f_{k,T}^{*}\end{bmatrix} \tag{38}$$

The posterior of the hyperparameter τ_j is $\pi\left(\tau_j|\cdot\right)=IG\left(g_j,G_j\right)$, with $g_j=g_0+0.5N$ and $G_j=G_0+0.5\sum_{i=1}^{N}{\lambda_{ij}^{*}}^2$.

6.4 Estimation efficiency

The performance is evaluated based on the root mean squared estimation error of the common component:

$$RMSEE^{(s)}=\frac{1}{R}\frac{1}{N}\frac{1}{M}\sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{m=1}^{M}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\lambda}_i^{*(m,r,s)}\hat{f}_t^{*(m,r,s)}-\lambda_i^{*(r,s)}f_t^{*(r,s)}\right)^2} \tag{39}$$

Table 7: Relative root mean squared estimation error of the common component under different prior specifications for the factor loadings, averaged over series, and replications. The numbers are expressed relative to the model estimated for $k = 3$ using the two-layer sparse prior specification.

	Scenarios: $r_{0j} = 500$, $s_{0j} =$			
	(0.2,0.2,0.1)		(0.9,0.75,0.5)	
	Estimation prior: $r_{0j} = 50$, $s_{0j} =$			
Estimated model	(0.2,0.2,0.1)	(0.3,0.3,0.3)	(0.9,0.75,0.5)	(0.75,0.75,0.75)
Two-layer sparse prior				
$k = 2$	1.65	1.55	2.39	2.34
$k = 4$	1.03	1.10	1.02	1.05
One-layer sparse prior				
$k = 2$	1.65	1.49	2.43	2.29
$k = 3$	1.00	1.01	1.00	1.00
$k = 4$	1.03	1.12	1.04	1.10
Normal prior				
$k = 2$	2.38	2.18	2.55	2.30
$k = 3$	2.07	2.05	1.12	1.08
$k = 4$	2.25	2.24	1.19	1.15

where for sparsity scenario s, $R = 10$, $M = 2000$, $N = 100$, $T = 100$ refer to the number of replications, the number of draws from the estimated posterior distribution, the number of simulated series and their length, respectively.

In table 7 we observe that the model estimated with the right number of factor always performs best. The efficiency losses to incur when using a one-layer instead of a two-layer prior appear minor when the true number of factors is assumed. In case of misspecification, the loss increases as the degree of sparsity decreases. Moreover, with a decreasing degree of sparsity, the efficiency gain of using the two-layer prior against the normal prior decreases. Finally, the loss in efficiency is largest for models estimated with a lower than the true number of factors.

7 Application: A large Swiss data set

7.1 Data and prior specification

To illustrate the method, we estimate a sparse factor model for a large Swiss panel data set. The dataset has been assembled and used by Kaufmann and Lein (2012), who investigate the existence of a price puzzle within a factor augmented VAR (FAVAR) approach. The data include 137 macroeconomic time series and 145 price series, measured at the quarterly frequency and covering the period of the first quarter of 1978 to the third quarter of 2008. The macroeconomic data series characterize the main aspects of the Swiss econ-

Table 8: Determining the number of factors. Distribution of variance share explained by the common component (CC).

	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
Bai&Ng IC_{p3}	-0.20	-0.19	-0.16	-0.14	-0.13
Share of excluded series	0.15	0.17	0.15	0.15	0.14
Share of series with a variance share of CC					
$\leq 20\%$	0.39	0.34	0.35	0.29	0.29
$\geq 60\%$	0.07	0.10	0.09	0.11	0.14
$\geq 80\%$	0.01	0.02	0.02	0.02	0.03
Variance share explained by CC					
Mean	0.27	0.29	0.29	0.32	0.33
Median	0.22	0.24	0.25	0.27	0.29
Increase in the variance share of CC					
Mean		0.01	0.01	0.02	0.01

omy and include, besides GDP and its components, series on real activity and the labor market, housing, financial markets and the GDP of Switzerland's major trading partners. Leading information is provided by consumer confidence surveys, production and price expectations in the manufacturing, wholesale and retail sale sectors, respectively. The price series cover about 90% of the Swiss CPI, including services and rents, non-durable, semi-durable and durable goods. For a detailed description of the data compilation see Kaufmann and Lein (2012).

We orientate ourselves towards Kaufmann and Lein (2012) to select in a first round the number of factors to estimate. They include three factors and the 3-month Libor as an additional one. Therefore, we estimate sparse specifications with three to seven factors to assess the appropriate number. In all specifications, we allow for two autoregressive lags in the idiosyncratic component and two lags for the factor dynamics. The sparse prior assumes a mean non-zero factor loading of $s_0 = 0.2$ with a relatively high precision of $r_0 = 100$, corresponding to 35% of the observed 282 time series. The series specific prior assumes a non-zero expected probability of $b_0 = 0.8$, with $a_0 = 0.01$. The prior variance of the normal prior for the non-zero factor loadings is specified by $\pi(\tau_j) \sim IG(2, 0.25)$.

7.2 Estimation and identification

Table 8 depicts several measures which are useful to assess the number of factors. We also choose $k = 4$ based on the various measures presented. On the first line, we report the information criterion IC_{p3} of Bai and Ng (2002). We adjust it and include only series with non-zero factor loadings, i.e. N is substituted by $\sum_{i=1}^{N} I_{\{\lambda_{ij}^{*} \neq 0, \forall j\}}$, and we average over all MCMC simulations. There is a minimal increase in the criterion from $k = 3$ to $k = 4$. Moreover, we do not obtain a significant increase in the variance share explained by the common component by increasing the number of factors above four, see the last line in table 8. At $k = 4$, we exclude a maximum share of series (47 series or 17%) without loosing much of the share of the common component in explaining the variance of the series. At $k = 4$, we explain up to 20% of the variance for 34% of the series and more than 60% and 80% of the variance for 10% and 2% of the series, respectively. Overall the variance share explained averages 29% in the panel. Finally, in the model $k = 4$, the first three factors load each exclusively on a subset of series, while the fourth factor loads on series that are also influenced by at least one of the other factors. So, in view of the simulation results, with which we document that the loss in efficiency is smaller when a larger rather than a lower than the true number of factors is assumed, we work with $k = 4$.

All the measures but the IC_{p3} criterion presented in table 8 are based on identified factor models for different k and are computed as the mean over all iterations using the significant non-zero factor loadings. A loading is defined as significant, if the median of all draws is different from zero, but we obtain the same inference when using the median of the simulated inclusion probabilities (the β_{ij}s). Identification is obtained while estimating the model: In the present case, while iterating over the sampler, we re-order the sampled values at the end of each sweep according to the column-specific number of non-zero factor loadings. Given the evidence presented in Kaufmann and Lein (2012), we opted for this procedure. This yields a preliminary ordering of the factors (without sign-identification) according to their importance in terms of significant factor loadings.

To ensure that this preliminary ordering yields a unique factor ordering, we post-process the MCMC iterations and first group the factor-specific draws according to their highest correlation (higher than 0.9) in order to obtain the factor-specific defining pattern. The rest of the iterations is then re-ordered again according to highest absolute correlation with the defining factor-specific patterns. Accordingly, the corresponding iterations of the other factor-specific parameters, λ^*, β and $\Phi(L)^*$, and factor-specific prior hyperparameters, τ and ρ, are also re-ordered. Finally, we perform a sign switch on the iterations negatively correlated with the factor-specific pattern and achieve overall sign identification by ensuring that the majority of the non-zero factor-specific loadings are positive.

7.3 Results

Figure 3 depicts the mean of the identified factors along with a 95% confidence interval. We are able to identify two factors which are comparable to the ones estimated by Kaufmann and Lein (2012) and two which differ from their estimates. Thinking in terms of inflation, factor 1 can be related to the periods of high and low inflation. Indeed, in figure 4 we observe that most price series, the loadings of which are at the right of the yellow bar, are mainly related to factor 1. In table 9 we report the labels of the series which are loaded uniquely by one factor (except for factor 4). The inference is obtained based on the median of the MCMC draws and a * indicates a negative loading. Again, we observe that mostly price series, the ones with alphanumerical labels, are loaded by factor one (see table 10 for the corresponding price series). Factor 2, with dips at the beginning of the 1980s, 1990s and 2000s, obviously relates to economic activity. According to table 9, GDP and some components, some labor market series and survey series are solely affected by factor 2. Factor 3 and 4 differ from the ones estimated in Kaufmann and Lein (2012).

According to the series solely affected by factor 3, in particular the exchange rates of Switzerland's main trading partners, the oil price and energy prices (D090 is heating oil and G105 fuel), we can relate this one to international monetary conditions. The remaining factor 4 captures special features of a subgroup of price series. All series loaded by factor 4 are also affected by one of the previous factors. The observed spike in this factor (see figure 3) captures special price increases, mainly on domestic goods and services, registered in the course of the substitution of the sales tax by the value-added tax at the beginning of 1995. The labels in table 9 reveal that prices on electricity, public transport, cinema and alcoholic beverages in restaurants were registered to be exceptionally affected.

The evidence is completed by noting that, again based on the median of the draws for the factor loadings, 47 series are excluded from the factor model, which corresponds to roughly 17% of the data. In the top panel of table 9, we depict the estimated loading structure. The first two factors load on nearly all series with non-zero factor loadings, 227 series out of 235. Figure 5 renders yet another picture of the loadings. The scatter plots depict in blue dots the lower bound of the 95% highest posterior density interval against the upper bound and the red stars depict the lower bound of the interval against the median of the draws. From the zero line and the 45 degree line, we observe that some of the loadings with median different from zero still have one of the bounds including zero. However, for these factor loadings, the serie-specific median and mean probabilities of non-zero factor loadings are usually very close to one or higher than 50%, respectively.

Finally, figure 6 plots the marginal posterior distribution for ρ_j. The prior is considerably updated and shifted away from the mean of 0.2 for the factors 1,2, and 4, each in accordance to the number of non-zero factor-specific loadings.

Figure 3: Posterior mean of factors.

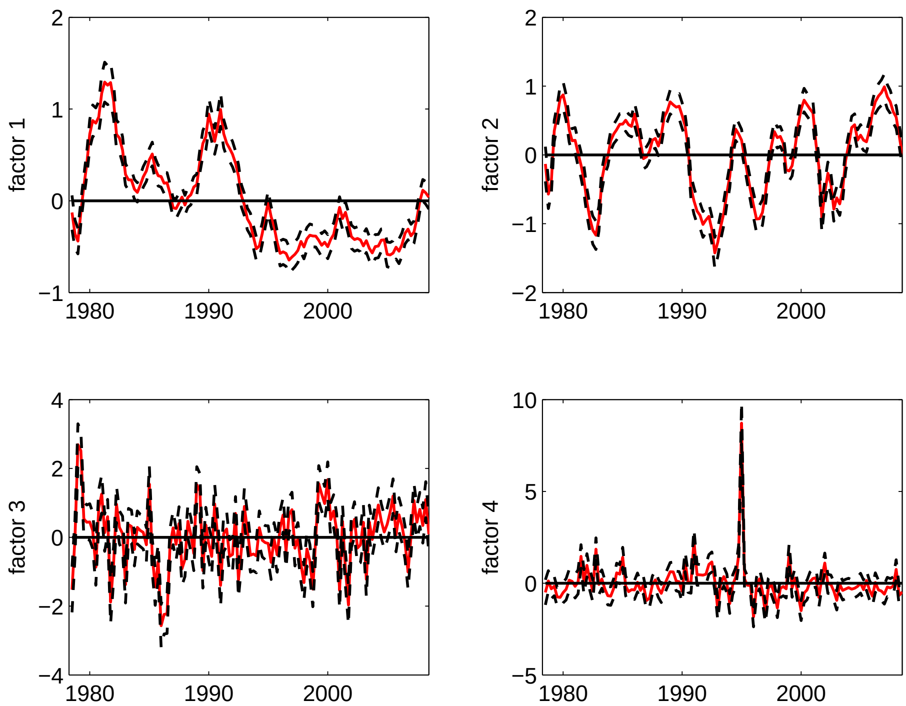
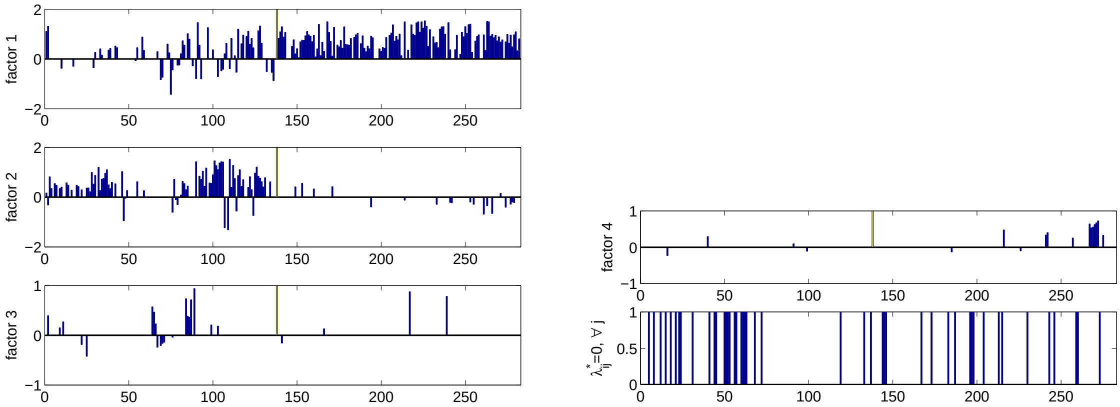


Table 9: Loading structure and series classification. For factor 1 to factor 3 we display the series uniquely affected by the respective factor. The series displayed for factor 4 are all also determined by (some of) the other factors. A * indicates series with negative loadings.

Loading structure	$\begin{bmatrix} 183 & & & \\ 57 & 101 & & \\ 11 & 9 & 24 & \\ 13 & 6 & 1 & 17 \end{bmatrix}$
Factor 1	*IPICHEM, EMPINS, *RESBUILD, HPRAPP, ZSEIDG10, MRATE, *SPRSNB, *NOTENUMLAUF, *SMPICRB, CONSPRI, KOF19, KOFRSLS, *USGDP, *WORLDTRADE, *PMIUSA, A003, A008, A014, A058, A065, A107, A115, A145, A170, A179, A200, A207, A212, A236, A246, A265, A285, A293, A297, A308, A347, A423, A449, A455, A481, A532, A539, A545, A552, A732, A862, B002, B010, B019, B046, B064, B082, C004, C015, C020, C027, C033, C050, C086, C093, C099, C126, C134, C175, C190, C198, C212, C220, C228, C237, D001, E002, E040, E050, E060, E071, E090, E100, E120, E150, E180, E221, F031, F036, G003, G062, G071, G082, G096, G113, H016, I003, I077, I085, I120, I211, I230, I300, I352, I420, I450, I501, I555, I570, L023, L100, L120, L130
Factor 2	GDP, PRICONS, INVEST, EQINV, INVENT, INDPROD, IPIMET, IPIENG, OECDLEAD, CEMENT, MANPOW, EMP2, EMPCHEM, EMPMET, EMPIND, HOURS, *URATE, OVERTIME, M2, *M3, EXPTOT, CONSFIN, CONSECO, CONSSAVE, KOF03, KOF05, KOF09, *KOF21, KOF27, KOFINDBS, KOFWSDEL, KOFWSEXPD, NOISEC2, UOISEC2, EMUGDP, *I400
Factor 3	EXPSER, CHFUSD, CHFEUR, CHFJPY, *REERUSD, OIL, D090, G105
Factor 4	*IPIWOOD, EMPREST, EXPPRIC, *CONSPURCH, *B058, D070, *E141, G210, G220, I436, K003, K052, K070, K075, K091, K103, L003
Excluded series	GOVCONS, CSTRINV, IMPSER, IPIFOOD, IPIMIN, IPIENWA, RETSALCF, RETSALFOOD, EMP1, EMPCOMM, EMPEDUC2, EMPADM, REDHRS, PARTRATE, HAPPR, HFINISH, CIVENG, HPAPP, HPINDU, HPSFH, UBS100, MSCI, TOTMAR, CHFJPYVOL, REERJPY, KOFRSJS, JPGDP, MSCIWLD, A076, A088, A097, A417, A519, B031, B075, C061, C067, C079, C168, D010, D050, F002, H001, I029, I465, I475, K170

The macroeconomic series are tabulated in Kaufmann and Lein (2011, table 4), the price series are found in appendix B (table 10).

Figure 4: Posterior median of factor loadings. The yellow bar indicates the sample split between macroeconomic and price variables. The bottom right panel plots the series with zero factor loading rows, these are 51 out of 282 series.



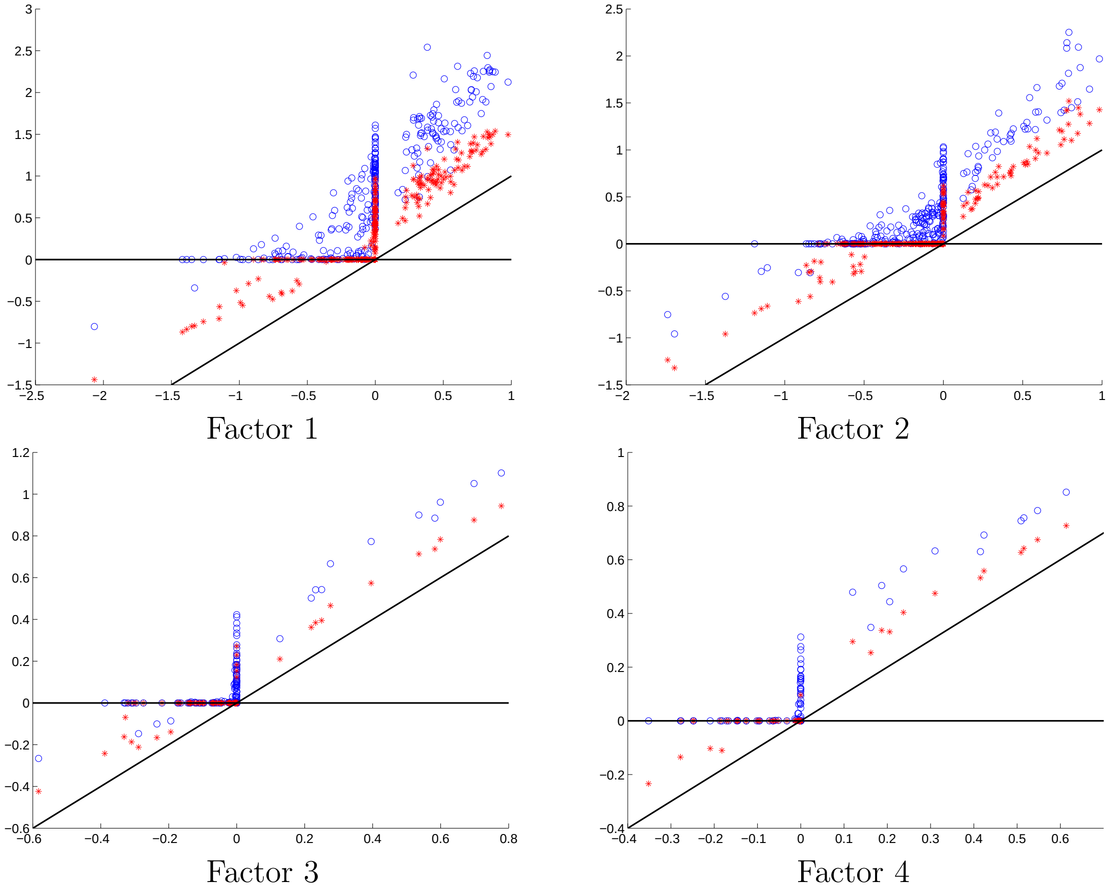
8 Conclusion

In the present paper we estimate a sparse dynamic factor model with Bayesian methods. The sparsity is useful to obtain an additional meaningful interpretation of the factors and to identify the variables irrelevant for factor estimation simultaneously while estimating the model. Variables with non-zero loadings in the same column are indicative of the interpretation of the factors. Variables with zero loadings on all factors are those which are irrelevant for the factor model. Moreover, we propose an identification procedure which is independent of variable ordering based on the semi-orthogonal representation of the sparse loading matrix. Sparsity is induced by designing a two-layer sparse prior on the factor loadings. A base rate governs the factor-specific mean probability of a non-zero loading, while, conditional upon the base probability, a unit-specific probability governs the probability of a non-zero loading for each series specifically.

The model is estimated with a Gibbs sampler. The estimation speed is greatly improved by applying a precision-based sampler. Simulations document that the true factor loading structure is nearly exactly recovered in settings with a high degree of sparsity. In settings with a low degree of sparsity, the loading structure is in general recovered with an error of less than 10%. With decreasing sparsity, the gains in estimation efficiency of using the two-layer sparse prior relatively to the normal prior specification on loadings diminishes but remain positive. Generally, compared with estimating a factor model with a fewer number of factors than the true number, the loss in estimation efficiency is lower when estimating factor models with a larger number of factors than the true number.

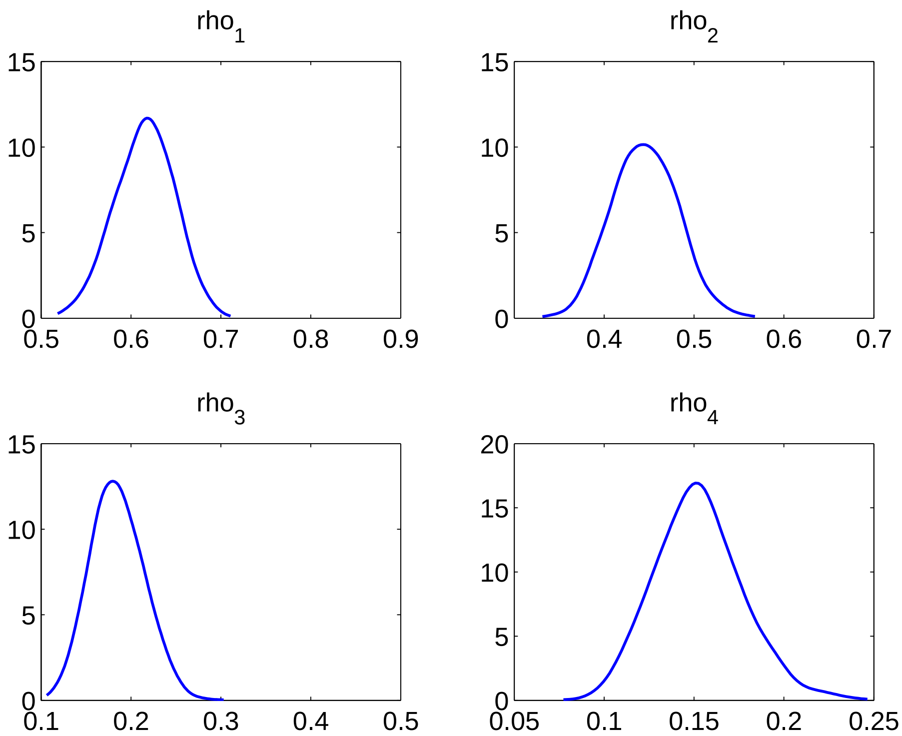
The estimation of a sparse factor model for a large dataset of Swiss macroeconomic (137) and price (145) series illustrates the method. We find four factors to be appropriate to capture the covariance structure in the data. The two factors loading on most series can be interpeted as an inflation and a business cycle factor, respectively. The third reflects international monetary conditions. The fourth factor, finally, affects a subset of the price

Figure 5: 95% highest posterior interval (blue, lower against higher bound) and median (red, lower bound against median) of factor loadings.



series and reflects price increases registered at the start of 1995 due to the substitution of the sales tax by the value-added tax.

Figure 6: Marginal posterior of ρ_j.



References

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics 18*, 338–357.

Anderson, T. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Prodeedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Economectrics, Industrial Research, and Psychometry*, pp. 111–150. Berkeley, California: University of California Press.

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

Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*, 304–317.

Bai, J. and S. Ng (2011). Principal components estimation and identification of the factors. Department of Economics Discussion Papers 1112-02, Columbia University.

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

Bhattacharya, A. and D. B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika 98*, 291–306.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 84*, 541–553.

Carvalho, C. M. (2006). *Structure and Sparsity in High-Dimensional Multivariate Analysis*. Ph. D. thesis, Institute of Statistics and Decision Sciences, Duke University.

Carvalho, C. M., J. Chang, J. E. Lucas, J. R. Nevins, W. Quanli, and M. West (2008). High-dimensional sparse factor modeling: Applications in gene expression genomics. *Journal of the American Statistical Association 103*, 1438–1456.

Chan, J., R. Leon-Gonzalez, and R. W. Strachan (2013). Invariant inference and efficient computation in the static factor model. mimeo.

Chan, J. C. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*, 101–120.

Eickmeier, S. (2005). Common stationary and non-stationary factors in teh euro area analyzed in a large-scale factor model. Discussion Paper 02/2005, Deutsche Bundesbank.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics & Statistics 82*, 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001). Coincident and leading indicators for the euro area. *The Economic Journal 111*, C62–C85.

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

Frühwirth-Schnatter, S. and H. F. Lopes (2010). Parsimonious Bayesian factor analysis when the number of factors is unknown. Technical report, The University of Chicago Booth School of Business.

Früwirth-Schnatter, S. (2011). Dealing with label switching under model uncertainty. In C. P. R. Kerrie L. Mengersen and D. M. Titterington (Eds.), *Mixtures: Estimation and Applications*, pp. 213–240. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom: John Wiley & Sons Ltd.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies 9*, 557–587.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrica 23*, 187–200.

Kaufmann, D. and S. Lein (2011). Sectoral inflation dynamics, idiosyncratic shocks and monetary policy. Working Paper 2011-7, Swiss National Bank.

Kaufmann, D. and S. Lein (2012). Is there a Swiss price puzzle? *Swiss Journal of Economics and Statistics 148*, 57–76.

Kaufmann, S. and C. Schumacher (2012). Finding relevant variables in sparse bayesian factor models: Economic applications and simulation results. Discussion Paper 29/2012, Deutsche Bundesbank.

Lawley, D. and A. Maxwell (1971). *Factor Analysis as a Statistical Method* (Second edition ed.). London: Butterworth & Co (Publishers) Ltd.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica 14*, 41–67.

Lucas, J., C. Carvalho, Q. Wang, A. Bild, J. Nevins, and M. West (2006). Sparse statistical modelling in gene expression genomics. In P. M. K.A. Do and M. Vannucci (Eds.), *Bayesian Inference for Gene Expression and Proteomics*. Cambridge University Press.

Peña, D. and P. Poncela (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference 136*, 1237–1257.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–131.

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

West, M. (2003). Bayesian factor regression models in the "large p, small n“ paradigm. *Bayesian Statistics 7*, 723–732.

Yoshida, R. and M. West (2010). Bayesian learning in sparse graphical factor models via variational mean-field annealing. *Journal of Machine Learning Research 11*, 1771–1798.

Zou, H., T. Hastie, and R. Tibshirani (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics 15*, 262–286.

A Posterior distributions

To derive the sampler, we first condense the variables in model (3) to:

$$\begin{aligned} \Psi(L)X_t = x_t^* &= \lambda f_t - \lambda \odot (\psi_{\cdot 1} \otimes \mathbf{1}_{1\times k}) f_{t-1} - \cdots - \lambda \odot (\psi_{\cdot q} \otimes \mathbf{1}_{1\times k}) f_{t-q} + \varepsilon_t & (40)\\ & \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon), \ \Sigma_\varepsilon \text{ diagonal} & \\ f_t &= \Phi_1 f_{t-1} + \cdots + \Phi_p f_{t-p} + \eta_t, \quad \eta_t \sim N(0, I_k) & (41) \end{aligned}$$

where $\odot$ and $\otimes$ represent the Hadamar and the Kronecker product, respectively. The row vector $\mathbf{1}_{1\times k}$ contains as elements k ones. We stack the observations to obtain the matrix representation:

$$\begin{aligned} \mathbf{X} &= \mathbf{\Lambda F} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, I_{T-q} \otimes \Sigma_\varepsilon) & (42)\\ \mathbf{\Phi F} &= \boldsymbol{\eta} \quad \boldsymbol{\eta} \sim N(0, \mathbf{S}) & (43) \end{aligned}$$

where $\mathbf{X} = (x_1^{*\prime}, \ldots, x_T^{*\prime})'$ contains all observed data, $\mathbf{F} = \left(f'_{q+1-\max(p,q)}, \ldots, f'_{q+1}, \ldots, f'_T\right)'$ stacks all unobserved factors, including initial states. The matrices $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ are respectively of dimension $(T-q)N \times (T+d)k$ and square $(T+d)k$, with $d = (p-q)I_{\{p>q\}}$. Typically, these matrices are sparse and banded around the main diagonal:

$$\mathbf{\Lambda} = \left[\begin{array}{c|cccccc} & -\lambda \odot (\psi_{\cdot q} \otimes \mathbf{1}_{1\times k}) & \ldots & \lambda & 0\ldots & & 0 \\ \mathbf{0}_{(T-q)N\times dk} & & \ddots & \ddots & & \ddots & \vdots \\ & 0\ldots & & 0 & -\lambda \odot (\psi_{\cdot q} \otimes \mathbf{1}_{1\times k}) & \ldots & \lambda \end{array}\right]$$

$$\mathbf{\Phi} = \left[\begin{array}{cccccccc} & I_p \otimes I_k & & & 0 & \ldots & & \\ \hline -\Phi_p & \ldots & -\Phi_1 & I_k & 0 & \ldots & & \\ & & & & \ddots & & & \\ & \ldots & 0 & -\Phi_p & \ldots & -\Phi_1 & I_k & \end{array}\right], \quad \mathbf{S} = \left[\begin{array}{ccc} I_p \otimes \Sigma_\eta^0 & 0 & \ldots \\ 0 & & \\ \vdots & & I_{T+d-p} \otimes I_k \end{array}\right]$$

where Σ_η^0 represents the variance of the initial states (see below).

The sampler proposed in section 5 consists in iterating over the following steps:

(i) Simulate f^T from $\pi\left(f^T | X^T, \theta\right)$ under the identification $\lambda'\lambda =$ diagonal.

(ii) Simulate the parameters from $\pi\left(\theta_{-\lambda} | f^T, X^T, \lambda\right)$

(iii) Transform the system to $f_t^* = H^{-1} f_t$, with $H = \Sigma_\eta^{1/2}$ and $\Phi^*(L) = H^{-1}\Phi(L)H$, $\Sigma_{\eta^*} = I$.
Simulate λ^* from $\pi\left(\lambda^* | f^{*T}, X^T, \Psi(L), \Sigma_\varepsilon\right)$ under a sparse prior.
Transform to $\lambda = \lambda^* H^{-1}$ with $H = (\lambda^{*\prime}\lambda^*)^{1/2}$, $\Sigma_\eta = HH'$, $\Phi(L) = H\Phi^*(L)H^{-1}$.

Step (iii) has been described in the main part of the paper. Here we derive the relevant posterior distributions for step (i) and (ii).

(i) *Simulate* f^T *from* $\pi\left(f^T | X^T, \theta\right)$

We adapt the sampler proposed in Chan and Jeliazkov (2009) to the present setup, which allows to sample f^T in one sweep. Compared with multi-move sampling, we obtain an improvement in computing time by a factor of up to seven. Given the representation in (42)-(43), the complete data likelihood has a normal density

$$f(\mathbf{X}|\mathbf{F}, \theta) \sim N(\mathbf{\Lambda F}, I_{T-q} \otimes \Sigma_\varepsilon) \tag{44}$$

For the unobserved states, from (43) we obtain the following prior:

$$\begin{aligned} \mathbf{F}|\theta \;\; &\sim \;\; N\left(0, F_0^{-1}\right) \\ &F_0 = \mathbf{\Phi}'\mathbf{S}^{-1}\mathbf{\Phi} \end{aligned} \tag{45}$$

If in $\mathbf{S}$, the variance of the initial states, Σ_η^0, is not chosen to be diffuse, we may use as initial conditions the stationary variance arising from the condensed VAR(1) representation of (41), $\mathbf{F}_t = \tilde{\mathbf{\Phi}}\mathbf{F}_{t-1} + \boldsymbol{\eta}_t$, $\boldsymbol{\eta}_t \sim N\left(0, \left[\begin{array}{cc} I_k & 0_{k\times(p-1)k} \\ \multicolumn{2}{c}{0_{(p-1)k\times pk}} \end{array}\right]\right)$, with

$$\begin{aligned} \tilde{\mathbf{\Phi}} &= \left[\begin{array}{c} \tilde{\mathbf{\Phi}}_{\mathbf{1}} \\ \tilde{\mathbf{\Phi}}_{\mathbf{2}} \end{array}\right] \\ \tilde{\mathbf{\Phi}}_{\mathbf{1}} &= [\Phi_1 \; \ldots \; \Phi_p] \\ \tilde{\mathbf{\Phi}}_{\mathbf{2}} &= \left[I_{(p-1)k} \; \mathbf{0}_{(p-1)k\times k}\right] \end{aligned} \tag{46}$$

We then have $E(\mathbf{F}_t\mathbf{F}_t') = \tilde{\mathbf{\Phi}}E(\mathbf{F}_{t-1}\mathbf{F}_{t-1}')\tilde{\mathbf{\Phi}}' + \Sigma_{\boldsymbol{\eta}}$ and $\Sigma_{\mathbf{F}} = \tilde{\mathbf{\Phi}}\Sigma_{\mathbf{F}}\tilde{\mathbf{\Phi}}' + \Sigma_{\boldsymbol{\eta}}$. The vec operator yields

$$\operatorname{vec}(\Sigma_{\mathbf{F}}) = \left[\mathbf{I}_{(pk)^2} - (\tilde{\mathbf{\Phi}} \otimes \tilde{\mathbf{\Phi}})\right]^{-1} \times \operatorname{vec}\left(\Sigma_{\boldsymbol{\eta}}\right) \tag{47}$$

from which we can retrieve the corresponding values for Σ_η^0.

Combining the prior with the likelihood, the posterior is:

$$\begin{aligned} \mathbf{F}|\mathbf{X}, \theta &\sim N\left(\mathrm{f}, F^{-1}\right) \\ F &= F_0 + \mathbf{\Lambda}'\left(I_{T-q} \otimes \Sigma_\varepsilon^{-1}\right)\mathbf{\Lambda} \\ \mathrm{f} &= F^{-1}\mathbf{\Lambda}'\left(I_{T-q} \otimes \Sigma_\varepsilon^{-1}\right)\mathbf{X} \end{aligned} \tag{48}$$

To avoid the full inversion of F we take the Cholesky decomposition, $F = L'L$, then $F^{-1} = L^{-1}L^{-1\prime}$. We obtain a draw $\mathbf{F}$ by setting $\mathbf{F} = \mathrm{f} + L^{-1}\boldsymbol{\nu}$, where $\boldsymbol{\nu}$ is a $(T+d)k$ vector of independent draws from the standard normal distribution.

(ii) *Simulate the parameters from* $\pi\left(\theta_{-\lambda}|f^T, X^T, \lambda\right)$
We block the posterior simulation of the parameters. The dynamics of the idiosyncratic components $\psi_i = (\psi_{i1}, \ldots, \psi_{iq})'$, $i = 1, \ldots, N$ can be sampled individually.

$$\pi\left(\psi_i|X_i, f^T, \theta_{-\Psi}\right) = N\left(\mathrm{q}_i, \mathrm{Q}_i\right) I_{\{Z(\Psi)>1\}}, \quad i = 1, \ldots, N$$

where

$$\begin{aligned} \mathrm{Q}_i &= \left(\sigma_i^{-2}\tilde{X}_i^{-\prime}\tilde{X}_i^{-} + \mathrm{Q}_0^{-1}\right)^{-1} \\ \mathrm{q}_i &= \mathrm{Q}_i\left(\sigma_i^{-2}\tilde{X}_i^{-\prime}\tilde{X}_i + \mathrm{Q}_0^{-1}\mathrm{q}_0\right) \end{aligned}$$

where $\tilde{X}_i$ and $\tilde{X}_i^-$ are the vector of the transformed variable i and the predictor matrix of the transformed system

$$\tilde{X}_i = \left[\begin{array}{c} X_{i,q+1} - \lambda_i f_{q+1} \\ \vdots \\ X_{iT} - \lambda_i f_T \end{array}\right] \qquad \tilde{X}_i^- = \left[\begin{array}{ccc} X_{iq} - \lambda_i f_q & \cdots & X_{i1} - \lambda_i f_1 \\ \vdots & & \vdots \\ X_{i,T-1} - \lambda_i f_{T-1} & \cdots & X_{i,T-q} - \lambda_i f_{T-q} \end{array}\right]$$

The dynamics of the common factors $\Phi^* = \left[\Phi_1^* \ \ldots \ \Phi_p^*\right]'$ are jointly sampled from

$$\pi\left(vec\left(\Phi^*\right)|f^{*T}\right) = N\left(\mathrm{p}, \mathrm{P}\right) I_{\{Z(\Phi^*)>1\}}$$

where

$$\begin{aligned} \mathrm{P} &= \left(\left[I_k \otimes f^{*-}\right]' \left[I_k \otimes f^{*-}\right] + \mathrm{P}_0^{-1}\right)^{-1} \\ \mathrm{p} &= \mathrm{P}\left(\left[I_k \otimes f^{*-}\right]' vec\left(f^*\right) + \mathrm{P}_0^{-1}\mathrm{p}_0\right) \end{aligned}$$

where $f^* = \left[f_{p+1}^* \ \ldots \ f_T^*\right]'$ and

$$f^{*-} = \begin{bmatrix} f_p^{*\prime} & \cdots & f_1^{*\prime} \\ \vdots & & \vdots \\ f_{T-1}^{*\prime} & \cdots & f_{T-p}^{*\prime} \end{bmatrix}$$

The posterior distribution of Σ_η is inverse Wishart $IW = (\mathrm{e}, \mathrm{E})$ with $\mathrm{e} = \mathrm{e}_0 + 0.5(T-p)$ and $\mathrm{E} = \mathrm{E}_0 + 0.5\left(f - f^-\Phi\right)'\left(f - f^-\Phi\right)$.

We simulate σ_i^2 from independent $IG = (\mathrm{u}_i, \mathrm{U}_i)$ distributions, $i = 1, \ldots, N$, with $\mathrm{u}_i = \mathrm{u}_0 + 0.5(T-q)$ and $\mathrm{U}_i = \mathrm{U}_0 + 0.5\left(\tilde{X}_i - \tilde{X}_i^-\psi_i\right)'\left(\tilde{X}_i - \tilde{X}_i^-\psi_i\right)$.

B Price series labels

Table 10: Price series

Label	Description	Label	Description
A003	Rice	A008	Flour
A014	Bread	A033	Pastries
A058	Pasta	A065	Other cereal products
A076	Beef	A088	Veal
A097	Pork	A107	Lamb
A115	Poultry	A133	Other meat
A145	Sausages	A170	Other processed meat
A179	Frozen fish	A180	Fresh fish
A200	Whole milk	A207	Other type of milk
A212	Hard cheese	A236	Tinned fish and smoked fish
A246	Other dairy products	A265	Cream
A278	Eggs	A285	Butter
A293	Margarine, fats, edible oils	A297	Other cheese
A308	Fresh fruits	A347	Dried, frozen and tinned fruit
A361	Fresh vegetables	A417	Potatoes
A423	Dried, frozen, tinned vegetables	A449	Jam, honey, sweets
A455	Chocolate	A475	Sugar
A481	Soups, spices, sauces	A519	Coffee
A532	Tea	A539	Cocoa and nutritional beverages
A545	Natural mineral water	A552	Soft drinks
A732	Ready-made foods	A862	Fruit or vegetable juices
B002	Spirits/brandies	B010	Liqueurs and aperitifs
B019	Swiss red wine	B031	Foreign red wine
B046	Swiss white wine	B058	Foreign white and sparkling wine
B064	Beer	B075	Cigarettes
B082	Other tobacco products	C004	Men: coats, jackets
C015	Men: suits	C020	Men: trousers
C027	Men: shirts	C033	Men: sweaters
C041	Men: underwear	C050	Sportswear
C061	Women: coats, jackets	C067	Women: costumes, trouser suits, dresses
C079	Women: trousers	C086	Women: jackets
C093	Women: blouses	C099	Women: other clothing
C126	Children: coats and jackets	C134	Children: other clothing
C168	Garment fabrics	C175	Other clothing accessories
C190	Garment alterations	C198	Upkeep of textiles
C212	Women: footwear	C220	Men: footwear
C228	Children: footwear	C237	Shoe repairs
D001	Rent	D010	Products for housing maintenance and repair
D020	Services for housing maintenance and repair	D050	Natural gas
D070	Electricity	D090	Heating oil
E002	Furniture: livingroom and bedroom	E040	Furniture: kitchen and garden
E050	Furnishings	E060	Floor coverings and carpets

Table 10: Price series label, continued.

Label	Description	Label	Description
E071	Bed linen and household linen	E090	Curtains and curtain accessories
E100	Major household appliances	E120	Smaller electric household appliances
E141	Kitchen utensils	E150	Tableware and cutlery
E180	Tools for DIY and garden	E221	Goods for routine household maintenance
F002	Medicines	F031	Medical services
F036	Dental services	F059	Hospital services
G003	New cars	G062	Motorcycles
G071	Bicycles	G082	Spare parts
G096	Tyres and accessories	G105	Fuels
G113	Repair services and work	G210	Public transport: direct service
G220	Pubic transport: combined services	H001	Postal services
H016	Telecommunication services	I003	Television sets and audiovisual appliances
I029	Photographic, cinematographic equipment and optical instruments	I077	PC hardware
I085	Recording media	I120	Repair and installation
I211	Games, toys and hobbies	I230	Equipment for sport, camping and open-air recreation
I300	Plants and flowers	I320	Pets and related products
I352	Sporting events	I400	Sports and leisure activities
I420	Mountain railways, ski lifts.	I436	Cinema
I450	Theatre and concerts	I465	Radio and television licences
I475	Photographic services	I490	Leisure-time courses
I501	Books and brochures	I525	Daily newspapers and periodicals
I555	Writing and drawing materials	I570	Package holidays
J050	Life-long learning	K003	Meals taken in restaurants and cafs
K052	Wine taken in restaurants	K070	Beer taken in restaurants
K075	Spirits, other alcoholic drinks taken in restaurants	K091	Coffee and tea taken in restaurants
K103	Other beverages taken in restaurants	K170	Alternative accommodation facilities
K171	Hotels	L003	Hairdressing establishments
L023	Soaps and foam baths	L040	Hair-care products
L055	Dental-care products	L070	Beauty products and cosmetics
L100	Paper articles for personal hygiene	L120	First aid material
L130	Personal care appliances, electric		