

# Discussion Paper

Deutsche Bundesbank  
No 29/2012

## **Finding relevant variables in sparse Bayesian factor models: economic applications and simulation results**

Sylvia Kaufmann

(Study Center Gerzensee)

Christian Schumacher

(Deutsche Bundesbank)

**Editorial Board:**

Klaus Düllmann  
Heinz Herrmann  
Christoph Memmel

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,  
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Telex within Germany 41227, telex from abroad 414431

Please address all orders in writing to: Deutsche Bundesbank,  
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet <http://www.bundesbank.de>

Reproduction permitted only if source is stated.

ISBN 978-3-86558-857-9 (Printversion)

ISBN 978-3-86558-858-6 (Internetversion)

**Abstract:**

This paper considers factor estimation from heterogeneous data, where some of the variables are noisy and only weakly informative for the factors. To identify the irrelevant variables, we search for zero rows in the loadings matrix of the factor model. To sharply separate these irrelevant variables from the informative ones, we choose a Bayesian framework for factor estimation with sparse priors on the loadings matrix. The choice of a sparse prior is an extension to the existing macroeconomic literature, which predominantly uses normal priors on the loadings. Simulations show that the sparse factor model can well detect various degrees of sparsity in the data, and how irrelevant variables can be identified. Empirical applications to a large multi-country GDP dataset and disaggregated CPI inflation data for the US reveal that sparsity matters a lot, as the majority of the variables in both datasets are irrelevant for factor estimation.

**Keywords:**

Factor models, variable selection, sparse priors

**JEL-Classification:**

C38, C11

## Non-technical summary

Factor models are nowadays widely used econometric tools in central banks. Policy-relevant applications of these models include the estimation of composite business cycle indicators, forecasting and structural investigations of the transmission of shocks. The econometric methods to estimate factor models can consider large datasets, and thus allow for the simultaneous analysis of many economic activities.

A difficulty with respect to the use of large datasets is that the information content of some included time series might be low, for example, if they are noisy. The literature shows that keeping these variables in the data can distort factor estimation. In this case, variable preselection could be considered to separate the relevant variables from irrelevant ones and estimate factors from the reduced dataset. Another solution might be the modification of the estimation methods of the model such that the weight attached to irrelevant variables becomes small.

In this paper, we discuss methods for finding irrelevant variables in factor models. We also discuss a sparse specification of the factor model that allows to discriminate sharply between relevant and irrelevant variables. To estimate factors, we follow a Bayesian approach from the recent statistical literature, where sparsity is introduced in the prior distribution of the loadings matrix. Such sparse priors have rarely been used in the macroeconomic literature on factor models.

Simulations show that the factor model is a flexible tool in the presence of data with different degrees of sparsity, where the information content of the data and the number of irrelevant variables varies. The simulations also indicate that the posterior distributions of the loadings are useful to find irrelevant variables in the data.

As an empirical application, we estimate the sparse factor model using an international dataset. In the literature, this data is used to analyse global business cycles. The results show that only less than 50% of the countries are explained by the factors. In particular, Africa and Latin America seem to play only a minor role in international business cycles. An additional application concerns disaggregated CPI inflation in the US. In this data, we find even more sparsity, as only about 35% of the sectoral inflation rates are informative about the factors. Thus, sectoral price developments seem to be driven mainly by sector-specific rather than aggregate shocks.

## Nicht-technische Zusammenfassung

In Zentralbanken werden Faktormodelle zu verschiedensten Zwecken verwendet. Sie bilden die Grundlage der Berechnung von zusammengesetzten Konjunkturindikatoren, werden zur Analyse von strukturellen Schocks und bei der Prognose angewendet. Die ökonomischen Methoden zur Schätzung von Faktormodellen erlauben die Auswertung umfangreicher Datensätze und können daher eine Vielzahl ökonomischer Aktivitäten simultan berücksichtigen.

Eine Schwierigkeit bei der Verwendung großer Datensätze besteht darin, dass einige Variablen verrauscht sein können und ihr Informationsgehalt für die Schätzung gemeinsamer Faktoren daher gering ist. Empirischer Evidenz aus der Literatur zufolge kann dadurch die Faktorschätzung erheblich verzerrt werden. In diesem Fall ist in Erwägung zu ziehen, eine Vorauswahl der für die Schätzung relevanten Variablen zu treffen und die irrelevanten Variablen bei der Schätzung zu vernachlässigen. Eine andere Lösung besteht darin, das Schätzverfahren des Modells so zu modifizieren, dass den irrelevanten Variablen schon bei der Schätzung ein geringes Gewicht beigemessen wird.

Im vorliegenden Papier werden Methoden vorgestellt und diskutiert, mit denen irrelevante Variablen in Faktormodellen identifiziert werden können. Zudem wird eine sparsame Spezifikation des Faktormodells vorgeschlagen, mit der eine scharfe Trennung zwischen relevanten und irrelevanten Variablen ermöglicht wird. Als Analyserahmen dient ein bayesianischer Schätzansatz, wobei Sparsamkeit über die A-Priori-Verteilungen der Ladungskoeffizienten des Faktormodells implementiert wird. Ein solcher Modellrahmen wurde bisher selten für makroökonomische Fragestellungen verwendet.

In Simulationen zeigt sich, dass sich das Faktormodell flexibel an verschiedene Gegebenheiten in den Daten anpassen kann, wobei insbesondere der Informationsgehalt in den Daten und somit die Zahl der relevanten und irrelevanten Variablen variiert wird. Ferner gelingt es mit den Schätzergebnissen, irrelevante Variablen im Datensatz zu identifizieren. Hierbei können insbesondere die A-Posteriori-Verteilungen der Ladekoeffizienten verwendet werden.

Als eine empirische Anwendung wird das sparsame Faktormodell mit einem internationalen Datensatz geschätzt, der in der Literatur zur Analyse globaler Konjunkturzyklen verwendet wird. Es zeigt sich, dass weniger als 50% der Länder durch die Faktoren erklärt werden. Insbesondere Afrika und Lateinamerika scheinen bestenfalls eine geringe Rolle im internationalen Konjunktuzusammenhang zu spielen. In einer weiteren Anwendung wird das Modell mit disaggregierten Inflationsdaten der USA geschätzt. In diesen Daten zeigt sich eine noch größere Sparsamkeit, da weniger als 35% der sektoralen Inflations-

raten durch die Faktoren erklärt werden. Demnach scheinen die Preisentwicklungen in vielen Sektoren vornehmlich durch idiosynkratische Einflüsse bestimmt zu sein, während aggregierte, gesamtwirtschaftliche Einflüsse für diese Sektoren eine sehr untergeordnete Rolle spielen.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Sparse factor model and identification</b>	<b>3</b>
2.1	Model structure . . . . .	3
2.2	Relevant and irrelevant variables for factor analysis . . . . .	4
2.3	Identification . . . . .	5
<b>3</b>	<b>Bayesian setup and MCMC estimation</b>	<b>6</b>
3.1	Likelihood and prior specification . . . . .	6
3.2	Posterior inference . . . . .	10
<b>4</b>	<b>Simulations</b>	<b>12</b>
4.1	Factor DGP . . . . .	12
4.2	Models for estimation . . . . .	14
4.3	Statistics to evaluate the estimation performance . . . . .	14
4.4	Simulation results . . . . .	16
4.5	Checks for robustness . . . . .	20
<b>5</b>	<b>Empirical results</b>	<b>20</b>
5.1	Results for international GDP growth data . . . . .	21
5.2	Results for US sectoral inflation data . . . . .	27
<b>6</b>	<b>Conclusion</b>	<b>30</b>
<b>A</b>	<b>Posterior distributions</b>	<b>32</b>
A.1	Step (i): Simulate $f^T$ from $\pi(f^T X^T, \theta)$ . . . . .	32
A.2	Step (ii): Simulate the parameters from $\pi(\theta_{-\lambda} f^T, X^T, \lambda)$ . . . . .	34
<b>B</b>	<b>Country coverage in the international GDP growth dataset</b>	<b>34</b>
<b>C</b>	<b>Additional simulation results</b>	<b>37</b>

## List of Tables

1	Simulation results with 10 irrelevant variables (out of 50) following an AR(1) process, without setting zero loadings a priori for estimation . . . .	17
2	Simulation results with 10 relevant variables (out of 50), without setting zero loadings a priori for estimation . . . . .	19
3	Summary statistics on sparsity based on international growth data with alternative specifications . . . . .	25
4	Percent of relevant countries by world region under alternative specifications	26
5	Simulation results with 10 irrelevant variables (out of 50) following an AR(1) process, with Geweke and Zhou (1996) identification . . . . .	37
6	Simulation results with 10 relevant variables (out of 50), with Geweke and Zhou (1996) identification . . . . .	38
7	Simulation results with 10 irrelevant variables (out of 50) following an AR(1) process, without setting zero loadings a priori for estimation, with larger prior variance of loadings . . . . .	39
8	Simulation results with 10 relevant variables (out of 50), without setting zero loadings a priori for estimation, with larger prior variance of loadings .	40

## List of Figures

1	Prior distributions $\pi(\rho)$ , $\pi(\beta)$ and $\pi(\lambda)$ under $r_0 = 5$ , $s_0 \in \{0.1, 0.5, 0.9\}$ and $a = 1$ . . . . .	9
2	Black-White heat plot indicating significant loadings on international GDP growth data . . . . .	22
3	Heat plot for loadings based on international GDP growth data . . . . .	23
4	Black-White heat plot indicating significant loadings based on disaggregated US CPI inflation data . . . . .	28
5	Heat plot for loadings based on disaggregated US CPI inflation data . . . .	29



# Finding relevant variables in sparse Bayesian factor models: Economic applications and simulation results<sup>1</sup>

## 1 Introduction

Factor models based on large datasets have received increasing attention in the recent macroeconomic literature. Policy-relevant applications of these models include the estimation of composite business cycle indicators, forecasting and structural investigations of the transmission of shocks (Altissimo et al., 2010; Stock and Watson, 2002; Kose et al., 2003; Bernanke et al., 2005).

A rarely discussed issue in the literature is the composition of the data used to estimate factors. Economists can in principle make use of all the information contained in large macroeconomic datasets that are nowadays available, as proper factor estimation techniques like principal components (PC) analysis can be applied to these large datasets. The question is to what extent the whole datasets are informative for the above-mentioned purposes. One of the few papers that discusses the composition of the data used to estimate factors is Boivin and Ng (2006). In the context of factor estimation by PC, they show that it generally depends on the information content of data for the factors whether the inclusion of additional data is detrimental. The point is illustrated by Monte Carlo simulation results, where the data-generating process (DGP) includes some variables that depend strongly on the factors whereas others are noisy. Both types of variables together imply heterogeneous information in the data and a factor loadings matrix containing a considerable number of zeros.

In the present paper, we propose a factor estimation framework, which takes into account heterogeneous information in the data. The factors are estimated within a state-space model with Bayesian techniques. To account for heterogeneous information, we choose a sparse, parsimonious parameterization of the factor loadings matrix. In particular, the loadings are estimated under a sparse prior, which assumes that many zeros may be present in the loadings matrix. Such an approach has been recently proposed in the context of gene expression analysis (West, 2003; Carvalho, 2006; Lucas et al., 2006; Car-

---

<sup>1</sup>This paper represents the authors' personal opinions and does not necessarily reflect the views of the Swiss National Bank or the Deutsche Bundesbank. We thank Knut Are Aastveit, Sandra Eickmeier, Heinz Herrmann, Anne Sofie Jore, and Malte Knüppel for helpful comments and suggestions. Contact information: Sylvia Kaufmann: Study Center Gerzensee, Dorfstrasse 2, P.O. Box 21, CH-3115 Gerzensee, Email: [sylvia.kaufmann@szgerzensee.ch](mailto:sylvia.kaufmann@szgerzensee.ch). Christian Schumacher: Deutsche Bundesbank, Economic Research Centre, Wilhelm-Epstein-Str. 14, 60431 Frankfurt am Main, Germany. Email: [christian.schumacher@bundesbank.de](mailto:christian.schumacher@bundesbank.de).

valho et al., 2008). In an economic context, sparse priors have hardly been applied before. An exception is the paper by Frühwirth-Schnatter and Lopes (2010), who estimate an exchange rate factor model with sparse loadings. However, the majority of macroeconomic applications of Bayesian factor models typically employ normal priors on the loadings, for example, to investigate the role of structural shocks in factor models (Bernanke et al., 2005; Kose et al., 2003, 2010; Mackowiak et al., 2009). The sparse prior we use here is an extension to priors for variable selection in multiple regression (George and McCulloch, 1993, 1997; Geweke, 1996) and is more general than the normal prior. The sparse prior on a particular factor loading implies more mass at zero than the normal prior and therefore allows for a sharper distinction between relevant and irrelevant variables.

Based on the posterior distributions of the loadings, we can investigate the relevance of variables in the model. In particular, rows in the factor loadings containing only zero elements indicate irrelevant variables which do not contribute relevant information to factor inference. The results on the relevance of variables might have important implications for empirical applications. For example, Francis et al. (2012) follow Kose et al. (2003) and investigate international business cycles by estimating international factors from a large set of country-specific GDP series. If variables for a certain country are not explained by the international factors and vice versa do not help to estimate them, one could conjecture that these countries are not linked to international business cycles. Mackowiak et al. (2009) identify common and sector-specific shocks from a factor model estimated from disaggregated US CPI inflation data covering many different products and services. The identification of irrelevant variables might be helpful in this case, as shocks hitting the common factors do not have a significant impact on these variables. Thus, the results on relevant variables might provide additional insights regarding the relative importance of sector-specific and aggregate shocks.

In the paper, we also discuss identification issues prevalent in factor models in the context of sparsity. A widely used identification scheme sets the upper diagonal elements of the factor loading matrix to zero (Geweke and Zhou, 1996; Aguilar and West, 2000). Obviously, when the number of variables is large and the factor loading matrix has a sparse structure, variable ordering becomes an important issue (Frühwirth-Schnatter and Lopes, 2010). Here, we propose to estimate the sparse factor model independently of the variable ordering without any zero restrictions imposed a priori as in Bhattacharya and Dunson (2011). We make use of the fact that the sparsity priors generally imply enough zeros in the loadings matrix a posteriori. In this regard, the sparse prior specification also contributes to the identification of factors and loadings.

To assess the quantitative properties of the sparse factor model, we provide Monte

Carlo simulations, where the DGP allows for different degrees of sparsity. Compared to a benchmark normal prior for the loadings, our prior well detects sparsity in the simulated data. This even holds, if there is only little sparsity in the DGP. We also compare different specifications of sparse priors. It turns out that the performance of the one-layer specification used in Frühwirth-Schnatter and Lopes (2010) is similar to the performance of the hierarchical two-layer prior in Carvalho et al. (2008). We also compare different ways to separate the irrelevant from the relevant variables in the data, and find in general stable results, independently of the degree of sparsity in the DGP.

To further illustrate the properties of the sparse factor model, we provide an empirical application to the multi-country GDP data used by Francis et al. (2012) and to the US sectoral inflation data by Mackowiak et al. (2009). Posterior inference on the factor loadings in both applications reveals that many time series have zero loading rows. In the international data, it turns out that African countries and most countries in Latin America are not driven by world factors. In the US inflation data, we find that more than 65% of the sectoral inflation rates are not explained by factors and thus mainly driven by sector-specific determinants.

The paper proceeds as follows. Section 2 provides an overview of the factor model, its identification, and the definition of relevant and irrelevant variables. In Section 3, we discuss the prior distributions, in particular of the loadings, and the posterior sampler. Section 4 contains the Monte Carlo simulation results, and Section 5 the empirical results. Finally, Section 6 summarizes and concludes.

## 2 Sparse factor model and identification

### 2.1 Model structure

Assume that  $X_t$  is a  $N \times 1$  vector of stationary series observed in time period  $t = 1, \dots, T$ . The number of time series  $N$  will typically be large, and we will work with the factor model

$$X_t = \lambda f_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon), \quad (1)$$

$$\Phi(L)f_t = \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta), \quad (2)$$

where the  $(k \times 1)$  factors  $f_t$  with  $k \ll N$  capture the common dynamics in the series. The factor loading matrix  $\lambda$  has dimensions  $N \times k$ . The vector autoregressive lag polynomial  $\Phi(L)$  is of order  $p$  with  $Lf_t = f_{t-1}$ . The idiosyncratic components  $\varepsilon_t$  are independent of each other,  $E(\varepsilon_{it}\varepsilon_{jt}) = 0$ ,  $i \neq j$ , i.e.  $\Sigma_\varepsilon$  is diagonal with elements  $\sigma_i^2$  for  $i = 1, \dots, N$ . For identification purposes, see Section 2.3 below, we assume that factor innovations

are independent of each other, rendering  $\Sigma_\eta$  diagonal. Finally, the factor innovations the idiosyncratic components are also assumed to be independent,  $E(\eta_t \varepsilon_t) = 0$ . Of course, if the common dynamics are not sufficient to capture all dynamics in the data, one might specify additional idiosyncratic dynamics, e.g. modify (1) to  $X_t = \lambda f_t + \xi_t$  with  $\Psi(L)\xi_t = \varepsilon_t$  and lag polynomial  $\Psi(L)$  of order  $q$ . However, to simplify notation we focus on the specification (1).

## 2.2 Relevant and irrelevant variables for factor analysis

As outlined in the introduction, we address the relevance of variables in the factor model by focussing on the loadings matrix in (1). In particular, we aim at finding those variables among the  $N$  series, which are not related to the factors and thus have only zero loadings elements in the respective row. For a particular series  $x_{it} \in X_t$ , the factor model is defined as  $x_{it} = \lambda_i f_t + \varepsilon_{it}$ , where  $\lambda_i$  is the  $i$ th ( $1 \times k$ ) row of  $\lambda$ . If all elements in the row vector  $\lambda_i$  are equal to zero, namely

$$\lambda_i = (\lambda_{i1} \lambda_{i2} \dots \lambda_{ik}) = (0 \ 0 \ \dots \ 0), \quad (3)$$

the common factors do not have an impact on variable  $x_{it}$ , and it is driven by idiosyncratic noise only,  $x_{it} = \varepsilon_{it}$ . We call these variables irrelevant for factor analysis. On the other hand, the relevant variables are the remaining ones, namely, those with at least one non-zero loading element.

In the simulations and empirical exercises below, we estimate the factor model and address the relevance of variables using Bayesian techniques. The search can be carried out based on the standard posterior output of Bayesian factor estimation. Therefore, searching for relevant variables might augment the inferential toolbox of factor analysis. Details how the relevance of the data is addressed will be provided in Section 4.3.

Note that zero rows in the loadings matrix have implications for the estimation of the unobserved factors, because the factor estimates are typically obtained by solving a signal extraction problem using the observed variables. If, for example, the Kalman filter is used to estimate the unobserved factors, a variable with corresponding zero loadings row does not contain relevant information for estimating the factors. Koopman and Harvey (2003) show how Kalman filtered and smoothed states can be expressed as functions of the observed data. These weight functions are equal to zero, if the loadings matrix, which is also the observation matrix in state space, has a zero row for the variable under consideration.

## 2.3 Identification

In model (1), loadings and factors are not separately identified. For any non-singular  $k \times k$  matrix  $H$ , there is an observationally equivalent model

$$X_t = \lambda H H^{-1} f_t + \varepsilon_t, \quad H^{-1} \Phi(L) H H^{-1} f_t = H^{-1} \eta_t. \quad (4)$$

To identify the model - up to sign identification -, we have to choose a non-singular matrix  $H$ . As it contains  $k^2$  elements, we need  $k^2$  identifying restrictions to pin down  $\lambda$  and  $f_t$ , see for example Bai and Ng (2010). A standard identification scheme sets the upper diagonal elements of the factor loading matrix  $\lambda$  equal to zero,  $\lambda_{ij} = 0$  for  $j > i$  (Geweke and Zhou, 1996; Aguilar and West, 2000). This provides us with  $k(k-1)/2$  restrictions. If we also impose the diagonality restriction on  $\Sigma_\eta$  and set each diagonal element to one,  $\Sigma_\eta = I$ , we additionally have  $k(k+1)/2$  restrictions. Thus we end up with  $k^2$  identifying restrictions.

An alternative approach that we pursue in the paper, is to estimate the factor loadings freely, without setting  $k(k-1)/2$  zero restrictions in the loadings matrix a priori, under a sparse prior specification, see also Bhattacharya and Dunson (2011) and Knowles and Ghahramani (2011). Details will be provided in Section 3.1. Under the sparse prior, the Gibbs sampler induces sparsity by setting loadings to zero, if the data support this decision. Thus, we obtain sparsity a posteriori, if the data indeed support enough zero restrictions. In our results we generally find a lot of support in favour of sparsity, and thus obtain well-defined posterior samples of the loadings and factors. This approach might also be justified, because the focus of our investigation is related somehow to variable selection (Bhattacharya and Dunson, 2011). Note that the hypothesis of variable irrelevance as defined in Section 2.2 is independent of the identification or rotation scheme chosen. Partition the true variables  $X_t$  into two  $(N_1 \times 1)$  and  $(N_2 \times 1)$  blocks,  $X_t = [X'_{1,t} \ X'_{2,t}]'$ , where the second block would have zero factor loadings,  $\lambda_2 = 0$ . The rotation with any invertible matrix  $Q$

$$\tilde{\lambda} = \lambda Q = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} Q = \begin{pmatrix} \lambda_1 Q \\ \lambda_2 Q \end{pmatrix} \quad (5)$$

yields  $\tilde{\lambda}_2 = \lambda_2 Q = 0Q = 0$ . Thus, sparsity in the factor model in terms of zero rows in the loadings matrix cannot be removed by rotation. Therefore, we expect that the investigation of sparsity on the basis of zero rows in the loadings matrix will not be heavily affected by the choice of the identification scheme.

In the literature on factor models, the issue of factor identification is crucially linked to finding an ordering of the variables such that the  $k$  leading variables be strongly associated

with the factors. In the lower triangular identification scheme of Geweke and Zhou (1996) and Aguilar and West (2000), the first variable is assumed to be affected by the first factor only, the second one by the second one and potentially by the first, etc. Large panels of macroeconomic data are usually clustered, for example, ranking first real output variables like industrial components and its components, then adding other categories of data as in the benchmark dataset of Stock and Watson (2002). Given this ordering of data, a zero upper diagonal identification scheme would establish mostly the first group of real variables as the founders of the different factors. This might be undesirable, as we might think of one factor mainly capturing the comovement between real variables, another one capturing the comovement between financial variables. To resolve the issue, Carvalho (2006) proposes an evolutionary approach to determine the number of factors and the selection and ordering of the  $k$  leading variables. Frühwirth-Schnatter and Lopes (2010) propose a generalized upper-zero triangular identification scheme, in which  $l_1 < \dots < l_k$ , where  $l_j$  denotes the row of the top non-zero entry in  $\tilde{\lambda}$  for  $j = 1, \dots, k$ , such that  $\tilde{\lambda}_{l_j, j} \neq 0$ , and  $\tilde{\lambda}_{ij} = 0 \forall i < l_j$ . As in the present paper, Frühwirth-Schnatter and Lopes (2010) estimate zero and non-zero loadings based on a latent indicator variable  $\gamma_{ij} \in \{0, 1\}$ , where  $\gamma_{ij} = 1$ , if the corresponding loading  $\lambda_{ij}$  is unequal to zero, and  $\gamma_{ij} = 0$  otherwise. At each step of the Gibbs sampler, the decision whether  $\lambda_{ij} \neq 0$  or  $\lambda_{ij} = 0$  is based on the posterior odds ratio  $P(\lambda_{ij} \neq 0 | \cdot) / P(\lambda_{ij} = 0 | \cdot)$ . Finally, Kaufmann and Schumacher (2012) propose to identify the factors independently of the variable ordering by combining a restriction-free sparse estimation of the factor loading matrix with the identifying restriction of semi-orthogonal loadings, where  $\lambda' \lambda$  is a diagonal matrix.

### 3 Bayesian setup and MCMC estimation

#### 3.1 Likelihood and prior specification

The complete-data likelihood takes the form

$$L(X^T | f^t, \theta) = \prod_{t=1}^T \pi(X_t | f^t, \theta), \quad (6)$$

where  $X^t = (X_t, X_{t-1}, \dots, X_1)$  denotes all observations up to period  $t$  and the parameter vector  $\theta$  includes all model parameters,  $\theta = (\lambda, \Phi, \Sigma_\varepsilon)$ . The observation density in (6) is multivariate normal

$$\pi(X_t | f^t, \theta) = \frac{1}{(2\pi)^{N/2} |\Sigma_\varepsilon|^{1/2}} \exp \left\{ -\frac{1}{2} (X_t - \lambda f_t)' \Sigma_\varepsilon^{-1} (X_t - \lambda f_t) \right\}. \quad (7)$$

The prior density of the unobserved factors is

$$\pi(f^T|\theta) = \prod_{t=p+1}^T \pi(f_t|f^{t-1}, \theta) \pi(f^p|\theta), \quad (8)$$

where  $f^p$  contains the initial states  $f^p = (f_{-p}, \dots, f_{-1}, f_0)$ . For the parameters we assume independent priors

$$\pi(\lambda, \Phi, \Sigma_\varepsilon) = \pi(\lambda) \pi(\Phi) \pi(\Sigma_\varepsilon) \quad (9)$$

where apart from the sparse prior on  $\lambda$ , all prior distributions are conjugate distributions. For the dynamic parameters we assume a multivariate normal prior truncated to the stationary region, according to

$$\pi(\Phi) = N(p_0, P_0) I_{\{Z(\Phi) > 1\}} \quad (10)$$

where  $I_{\{\cdot\}}$  is the indicator function and  $Z(\Phi) > 1$  means that the roots of the characteristic equation of the process  $\Phi(L)$  lie outside the unit circle. To specify  $p_0$  and  $P_0$  of the VAR process we implement a standard Minnesota prior. Given that  $\Sigma_\varepsilon$  is diagonal, we assume independent inverse Gamma prior distributions for the variances,  $\pi(\sigma_{\varepsilon,i}^2) = IG(u_0, U_0)$ , for  $i = 1, \dots, N$ .

The sparse prior specification for the loadings  $\pi(\lambda)$  follows the approach of West (2003), Lucas et al. (2006), and Carvalho et al. (2008), and extends the variable selection priors based on mixtures surveyed in George and McCulloch (1997):

$$\pi(\lambda_{ij}) = (1 - \beta_{ij})\delta_0(\lambda_{ij}) + \beta_{ij}N(0, \tau_j), \quad (11)$$

$$\pi(\beta_{ij}) = (1 - \rho_j)\delta_0(\beta_{ij}) + \rho_j B(a_j b_j, a_j(1 - b_j)), \quad (12)$$

$$\pi(\rho_j) = B(r_{0j} s_{0j}, r_{0j}(1 - s_{0j})). \quad (13)$$

The sparsity prior for  $\pi(\lambda_{ij})$  in (11) assigns each loading element  $\lambda_{ij}$  an individual probability  $\beta_{ij}$  of taking a non-zero value, see Carvalho et al. (2008). If the loading is non-zero, its values are drawn from a normal prior with variance  $\tau_j$ , which follows an inverse Gamma prior distribution,  $\pi(\tau_j) = IG(g_0, G_0)$ . Otherwise, the loading is zero, as implied by the Dirac delta function  $\delta_0(\cdot)$ . Layer (12) defines the element-specific probability  $\beta_{ij}$  of association between variable  $i$  and factor  $j$  and is specified such that inclusion probabilities become either very small or very large. The prior of  $\beta_{ij}$  is itself a mixture of a Dirac delta function and a Beta function  $B(ab, a(1 - b))$  with mean  $b$  and precision  $a$ . In layer (13), the Beta prior on  $\rho_j$  favours very small values if the precision  $r_{0j} > 0$  is large and the mean  $s_{0j}$ , which can be regarded as prior probability of nonzero values of  $\beta_{ij}$ , is small.



The prior (11)-(13) implies a common probability  $1 - \rho_j b_j$  across variables of a zero loading on factor  $j$ . The hierarchical prior circumvents the observed problem that for increasing  $N$ , the posterior of the factor loading  $\lambda_{ij}$  is spread out, resulting in a large uncertainty about the significance of the loading (Lucas et al., 2006). However, due to the rich structure of the hierarchical prior, this might also come at a cost of increased computation and perhaps reduced predictive performance, suggesting that the uncertainty removed from the posterior actually would be important (Knowles and Ghahramani, 2011).

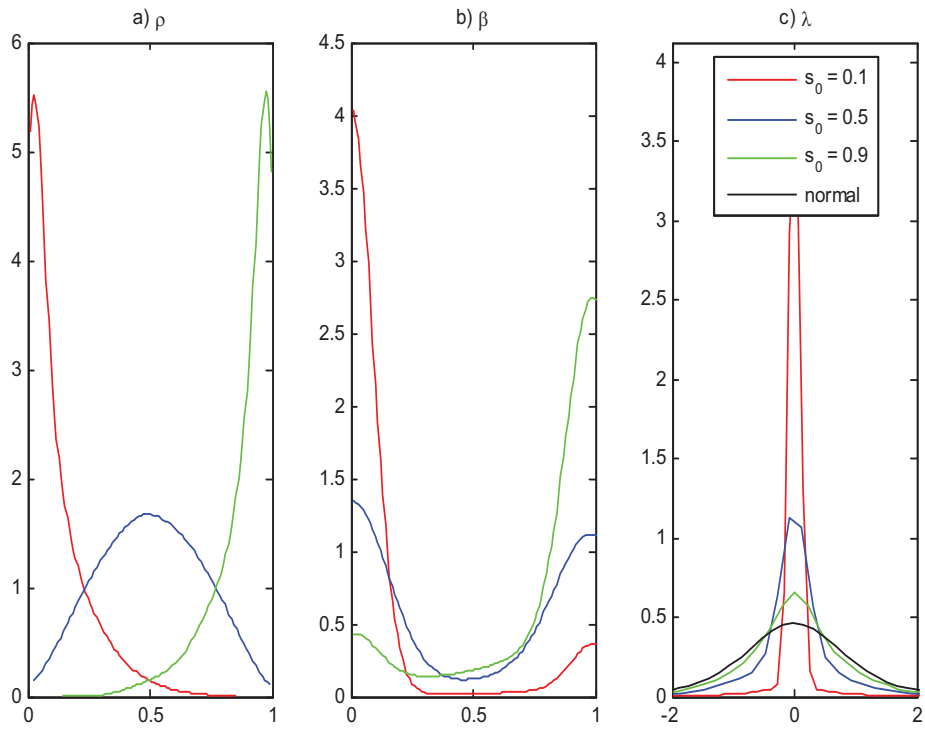
To obtain an intuition of the shape of the prior distribution of  $\lambda_{ij}$ , we provide a numerical example with 10000 draws from the prior. We first simulate out of the prior in layer (13), where we assume a precision of  $\rho_{0j} = 5$  with different means  $s_{0j} \in \{0.1, 0.5, 0.9\}$ . Figure 1, Panel a), plots the distribution for  $\rho$  under the various parameter settings. Obviously, with increasing mean, the distribution is shifted from left to right. In the second layer (12), the prior for  $\beta_{ij}$  has a bimodal shape in Panel b), and an increase in  $s_{0j}$  leads to a higher probability that  $\beta_{ij}$  will be non-zero, given the hyperparameter constellation of  $a = 1$  and  $b = 0.9$ . We obtain an expected base rate of a non-zero factor loading of 0.09. On average, 10% of the loadings will have an expected 90% probability of a non-zero value. By additionally setting  $g_0 = 2$  and  $G_0 = 1$ , which induces  $E(\tau_j) = 1$ , we generally obtain a fat-tailed prior distribution for  $\lambda_{ij}$  with a lot of mass at zero depicted in Panel c) of Figure 1. Increasing  $s_{0j}$  renders  $\pi(\lambda)$  less peaked and thus leads a priori to a higher inclusion probability and a smaller degree of sparsity in the loadings.

The normal prior can be obtained from the hierarchical prior by letting  $s_{0j} \rightarrow 1$ , see again how the distribution in Panel c) of Figure 1 converges to the normal distribution as  $s_{0j}$  increases. Thus, the normal prior is a special case of the hierarchical prior chosen here. Note that macroeconomic applications of Bayesian factor models typically employ the normal prior for the loadings (Bernanke et al., 2005; Kose et al., 2003, 2008; Mackowiak et al., 2009). By choosing  $s_{0j} < 1$ , we assume that there is a priori more sparsity than in the normal prior. Note that the hierarchical prior does not only imply a different shape compared to the normal prior. Due to the mixture specification in layer (11), we will effectively set some  $\lambda_{ij}$  exactly equal to zero, if sparsity matters in the data. The continuous normal prior does not have this property. Thus, by using the hierarchical prior, we hope to obtain a sharper distinction between relevant and irrelevant variables in the macroeconomic applications. Note that the zero loadings can also help to identify loadings and factors, as setting zero restrictions is a standard way of factor identification, see the discussion in Section 2.3 above.

Compared to the existing literature, the two-layer prior can also be regarded as an



Figure 1: Prior distributions  $\pi(\rho)$ ,  $\pi(\beta)$  and  $\pi(\lambda)$  under  $r_0 = 5$ ,  $s_0 \in \{0.1, 0.5, 0.9\}$  and  $a = 1$ .



extension to priors from the Bayesian literature on variable selection (George and McCulloch, 1993, 1997):

$$\pi(\lambda_{ij}) = (1 - \rho_j)\delta_0(\lambda_{ij}) + \rho_j N(0, \tau_j), \quad (14)$$

$$\pi(\rho_j) = B(r_{0j}s_{0j}, r_j(1 - s_{0j})). \quad (15)$$

This prior is equivalent to the two-layer prior (11)-(13), if we neglect the layer for  $\beta_{ij}$ . Thus, the two-layer prior (11)-(13) from Carvalho et al. (2008) actually also nests the one-layer prior from George and McCulloch (1997). Frühwirth-Schnatter and Lopes (2010) design a hierarchical prior, in which the significance of the factor loading is governed by an indicator  $\gamma_{ij} \in \{0, 1\}$  with  $P(\gamma_{ij} = 1 | \rho_j) = \rho_j$ , where  $\rho_j$  also follows a Beta distribution. The prior can be related to the two-layer prior by constraining  $\rho_j b_j$  a priori to lie in  $[0, b_j]$ . Thus, the prior used in Frühwirth-Schnatter and Lopes (2010) is actually closely related to the one-layer prior from George and McCulloch (1997), see also Geweke (1996).

A contribution of the present paper is to compare the performance of the different priors. We will do so in a simulation exercise, see Section 4, where we will compare the different priors with respect to their ability to detect different degrees of sparsity in the data.

### 3.2 Posterior inference

Updating the prior with data information yields the inference on the posterior distribution of  $\vartheta = (f^T, \theta)$ ,  $\pi(\vartheta | X^T) \propto L(X^T | f^T, \theta) \pi(f^T | \theta) \pi(\theta)$ . The sampler is based on the following steps

- (i) Simulate  $f^T$  from  $\pi(f^T | X^T, \theta)$ .
- (ii) Simulate  $\Phi, \Sigma_\varepsilon$  from  $\pi(\theta_{-\lambda} | f^T, X^T, \lambda)$ , where  $\theta_{-\lambda}$  excludes  $\lambda$  from  $\theta$ .
- (iii) Simulate  $\lambda$  from  $\pi(\lambda | f^T, X^T, \Sigma_\varepsilon)$  under a sparse prior.
- (iv) Update the hyperparameters of the sparse prior.

Step (i) and (ii) are standard by now in the Bayesian simulation setup. The path  $f^T$  is simulated using a multi-move sampler as proposed in Carter and Kohn (1994), Shephard (1994), Frühwirth-Schnatter (1994). The parameter simulation needs further blocking. Given the conjugate priors, the posterior distributions are multivariate normal and inverse Wishart. The distributions can be found in the appendix.

Step (iii) and (iv) update the sparse prior of the factor loading matrix and the hyperparameters, respectively, given the factors. As the derivation is less standard, see also Carvalho (2006) and Geweke (1996), we describe it in more detail. In particular, the

posterior  $\pi(\lambda_{ij}|f^T, X^T, \Sigma_\varepsilon)$  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  $E(\beta_{ij}) = \rho_j b_j$  across variables. The marginal prior becomes

$$\pi(\lambda_{ij}|\rho_j) \sim (1 - \rho_j b_j)\delta_0(\lambda_{ij}) + \rho_j b_j N(0, \tau_j) \quad (16)$$

For each factor  $j$ , transform the variables to

$$x_{it}^* = x_{it} - \sum_{l=1, l \neq j}^k \lambda_{il} f_{lt} = \lambda_{ij} f_{jt} + \varepsilon_{it}$$

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

$$\pi(\lambda_{ij}|\cdot) = \prod_{t=q+1}^T \pi(x_{it}^*|\cdot) \{(1 - \rho_j b_j)\delta_0(\lambda_{ij}) + \rho_j b_j N(0, \tau_j)\} \quad (17)$$

$$= P(\lambda_{ij} = 0|\cdot) \delta_0(\lambda_{ij}) + P(\lambda_{ij} \neq 0|\cdot) N(m_{ij}, M_{ij}) \quad (18)$$

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

$$M_{ij} = \left( \frac{1}{\sigma_{\varepsilon, i}^2} \sum_{t=1}^T f_{jt}^2 + \frac{1}{\tau_j} \right)^{-1}, \quad m_{ij} = M_{ij} \left( \frac{1}{\sigma_{\varepsilon, i}^2} \sum_{t=1}^T f_{jt} x_{it}^* \right) \quad (19)$$

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

$$PO_{ij} = \frac{P(\lambda_{ij} \neq 0|\cdot)}{P(\lambda_{ij} = 0|\cdot)} = \frac{\pi(\lambda_{ij})|_{\lambda_{ij} \neq 0}}{\pi(\lambda_{ij})|_{\lambda_{ij} = 0}} \frac{\rho_j b_j}{1 - \rho_j b_j} = \frac{N(0; 0, \tau_j)}{N(0; m_{ij}, M_{ij})} \frac{\rho_j b_j}{1 - \rho_j b_j} \quad (20)$$

We choose  $\lambda_{ij} \neq 0$  if  $U \leq PO_{ij} / (1 + PO_{ij})$ , where  $U$  is a draw from the uniform distribution over  $[0, 1]$ . To address the relevance of variables later we follow Frühwirth-Schnatter and Lopes (2010) and compute the binary indicators  $\gamma_{ij}$  defined as  $\gamma_{ij} = 1$  if  $\lambda_{ij} \neq 0$ , and  $\gamma_{ij} = 0$  if  $\lambda_{ij} = 0$ . Thus the indicators  $\gamma_{ij}$  show whether variable  $x_{it}$  is associated with factor  $f_{jt}$  or not.

If the binary decision leads to  $\lambda_{ij} \neq 0$  in this step,  $\lambda_{ij}$  is drawn from  $N(m_{ij}, M_{ij})$ , and otherwise zero. Conditional on  $\lambda_{ij}$  we update the variable specific probabilities  $\beta_{ij}$  and sample from  $\pi(\beta_{ij}|\lambda_{ij}, \cdot)$ . When  $\lambda_{ij} = 0$ ,

$$\begin{aligned} \pi(\beta_{ij}|\lambda_{ij} = 0, \cdot) &\propto (1 - \beta_{ij}) [(1 - \rho_j)\delta_0(\beta_{ij}) + \rho_j B(a_j b_j, a_j(1 - b_j))] \\ P(\beta_{ij} = 0|\lambda_{ij} = 0, \cdot) &\propto (1 - \rho_j), \quad P(\beta_{ij} \neq 0|\lambda_{ij} = 0, \cdot) \propto (1 - b_j)\rho_j \end{aligned} \quad (21)$$

That is, with posterior odds  $(1 - b_j)\rho_j/(1 - \rho_j)$  we sample from  $B(a_j b_j, a_j(1 - b_j) + 1)$  and set otherwise  $\beta_{ij}$  equal to zero. Conditional on  $\lambda_{ij} \neq 0$  we obtain

$$\begin{aligned} \pi(\beta_{ij}|\lambda_{ij} \neq 0, \cdot) &\propto \beta_{ij} N(\lambda_{ij}; 0, \tau_j) [(1 - \rho_j)\delta_0(\beta_{ij}) + \rho_j B(a_j b_j, a_j(1 - b_j))] \\ P(\beta_{ij} = 0|\lambda_{ij} \neq 0, \cdot) &= 0, \quad P(\beta_{ij} \neq 0|\lambda_{ij} \neq 0, \cdot) = 1 \end{aligned} \quad (22)$$

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

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

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

## 4 Simulations

With simulations, we want to investigate how the different priors in the model affect factor estimation under different degrees of sparsity in the data. The DGP will vary with respect to the information content of the data for the factors, which is implemented through different specifications of the loadings, see Section 4.1. Given the simulated data, we estimate the common components and compare them to the simulated values from the DGP following the simulations with heterogenous data in Boivin and Ng (2006). The specification of the estimated models is discussed in Section 4.2. We also simulate the search for irrelevant variables. In Section 4.3 we compare different procedures to find out whether irrelevant variables, which are defined to be idiosyncratic in the DGP, can indeed be identified as such based on the estimation results. We do the same with simulated relevant variables in the DGP. The simulation results are provided in Section 4.4.

### 4.1 Factor DGP

For the simulation study, we simulate  $N = 50$  time series of length  $T = 100$ , driven by a  $k = 2$  dimensional factor process. The dataset will be simulated in two parts: In the first part, for  $i = 1, \dots, 40$ , we simulate data with different degrees of sparsity. To address the relevance of variables, we simulate another 10 variables in the second part, for  $i = 41, \dots, 50$ , which are either relevant or irrelevant by definition:

1. The DGP for the first block of variables for  $i = 1, \dots, 40$ , is a 2-factor DGP with a factor VAR with  $p = 1$  lags and takes the form

$$x_{it} = \lambda_i f_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma_{\varepsilon,i}^2), \quad (23)$$

$$f_t = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.8 \end{bmatrix} f_{t-1} + \eta_t, \quad \eta_t \sim N(0, I), \quad (24)$$

To sample the loadings  $\lambda_i$ , we rely on the two-layer prior distributions in equations (11)-(13) from Section 3.1. The degree of sparsity in  $\lambda_i$  is successively relaxed by increasing the hyperparameter  $s_{0j}$ , namely  $s_{0j} \in \{0.1, 0.5, 0.9\}$  for  $j = 1, 2$ , see equation (13) above. Increasing  $s_{0j}$  from 0.1 to 0.9 implies that the fraction of non-zero factor loadings increases from 10% to 90% based  $b_j = 0.8$ . The marginal expected factor-specific probability of a non-zero factor loading thus increases from 8% to 72%. As regards precision, we fix  $r_{0j} = 50$  and  $a_j = 0.01$  for  $j = 1, 2$ . The non-zero factor loadings are simulated out of the normal distributions  $N(m_j, M)$  with  $M = 0.01$  and  $m_1 = 0.60$ ,  $m_2 = 0.40$ . The variance  $M = 0.01$  is chosen relatively tight in order to sharply define non-zero loadings. Variances of the idiosyncratic components are fixed to  $\sigma_{\varepsilon,i}^2 = 0.74$  for  $i = 1, \dots, 40$ .

2. The DGP for the second block of variables, for  $i = 41, \dots, 50$ , contains either relevant or irrelevant variables:

- (a) Irrelevant variables: We simulate 10 completely idiosyncratic variables with DGP  $x_{it} = 0.5 \times x_{it-1} + \varepsilon_t$  with  $\varepsilon_t \sim N(0, \sigma_{\varepsilon,i}^2)$  and  $\sigma_{\varepsilon,i}^2 = 0.74$ .
- (b) Relevant variables: The DGP for relevant variables is defined by  $x_{it} = \lambda_i f_t + \varepsilon_{it}$ , where the loadings are sampled according to  $\lambda_{ij} \sim N(m_j, M)$  with  $M = 0.01$  and  $\varepsilon_t \sim N(0, \sigma_{\varepsilon,i}^2)$  and  $\sigma_{\varepsilon,i}^2 = 0.37$ . Concerning the loadings means, we assume that one factor explains a relevant variable, but which one is randomly determined. With equal probability, we choose for each sample whether the first or the second factor explains the relevant variable. If the first factor is chosen, we sample  $\lambda_{i1} \sim N(m_1, M)$  with  $m_1 = 0.60$ , and if the second factor is chosen, we set  $m_2 = 0.40$  to sample  $\lambda_{i2} \sim N(m_2, M)$ , corresponding to the DGP of the first block of variables above.

Overall, we obtain six different specifications of DGPs depending on the three sparsity degrees  $s_{0j} \in \{0.1, 0.5, 0.9\}$  for the first 40 variables and depending on whether the final 10 variables are either relevant or irrelevant. For each DGP specification, we produce  $R = 50$  replications of the multivariate data as well as true loadings  $\lambda_i^{(r)}$  and factors  $f_t^{(r)}$  for  $r = 1, \dots, R$ .

## 4.2 Models for estimation

The sparse factor model is estimated for  $k = 2$  factors under different sparse prior distributions for the loadings. We compare the results of the sparse factor model estimated with the two-layer prior to the factor model estimated with the normal prior. Furthermore, we compare the estimates obtained with the two-layer prior (11)-(13) against those obtained with the one-layer prior (14)-(15). To estimate the factor model under the sparse priors, we assume that the true degree of sparsity in the DGP is unknown and choose an intermediate degree of sparsity  $s_{0j} = 0.5$  for  $j = 1, 2$ . Thus, we can investigate how sparsity is detected with the different priors, when the true degree of sparsity in the DGP is not known a priori. We additionally fix  $r_{0j} = 3$  and  $a_j = 0.5$  for  $j = 1, 2$ , implying low precision a priori when estimating the model compared to the DGP. For  $b_j$ , we assume  $b_j = 0.8$ . Furthermore, the prior on  $\tau_j$  in  $N(0, \tau_j)$  is parameterized by  $IG(2, 0.5)$ . Concerning the variances of the idiosyncratic components, we define  $\pi(\sigma_{\varepsilon,i}^2) = IG(u_0, U_0)$  for  $i = 1, \dots, N$  with  $u_0 = 2.0$  and  $U_0 = 1.0$ . To specify hyperparameters of the factor VAR lag polynomial in (10), we implement a Minnesota prior with variance equal to 0.16 on the first autoregressive lag and a shrinkage factor of 0.0278 for off-diagonal parameters. As regards identification, we will estimate the factor model without any identifying restrictions following the discussion in Section 2.3. As a check for robustness, we will later also consider the identification scheme by Geweke and Zhou (1996). To estimate the sparse model for a given simulated dataset indexed by  $r$  for  $r = 1, \dots, R$ , we draw 5000 times from the posterior, discard the first 1000 draws, keep every second one, and evaluate the posterior with  $G^{\text{eff}} = (5000 - 1000)/2 = 2000$  retained iterations. Thus, we obtain loadings and factor estimates  $\lambda_i^{(r,g)}$  and  $f_t^{(r,g)}$ , respectively, for  $g = 1, \dots, G^{\text{eff}}$ .

## 4.3 Statistics to evaluate the estimation performance

The results in the Tables below will be based on three different statistics. As first statistic of estimation accuracy, we follow the simulations by Boivin and Ng (2006) and assess the accuracy of the estimates of the common components. For that purpose, we compute the root mean-squared error  $\text{RMSE}_r$  of the common components by

$$\text{RMSE}_r = \frac{1}{N} \frac{1}{G^{\text{eff}}} \sum_{i=1}^N \sum_{g=1}^{G^{\text{eff}}} \sqrt{\frac{1}{T} \sum_{t=1}^T \left( \lambda_i^{(r,g)} f_t^{(r,g)} - \lambda_i^{(r)} f_t^{(r)} \right)^2}, \quad (25)$$

for all the  $r = 1, \dots, R$  simulated datasets. Finally, we average over the replications to obtain the  $\text{RMSE} = \frac{1}{R} \sum_{r=1}^R \text{RMSE}_r$ . We obtain RMSEs for data simulated with different

sparsity degrees  $s_{0j} \in \{0.1, 0.5, 0.9\}$ , and for estimating the factor model under the one-layer, two-layer, and the normal prior.

In addition, we address the relevance or irrelevance of variables from the posterior draws of the model parameters. From the DGP, we know where the irrelevant variables are located, namely at the end of the dataset,  $i = 41, \dots, 50$ . To assess the relevance of a variable, we follow two different approaches.

1. First, we look at the posterior draws of each loading  $\lambda_{ij}^{(g)}$ ,  $g = 1, \dots, G^{\text{eff}}$  and  $j = 1, 2$ , and search for those variables that have a loading posterior distribution which is significantly different from zero. For this purpose, we compute 95% HPD regions for each  $\lambda_{ij}$  from the posterior samples and check whether zero is included or not. If zero is not included, we say that variable  $x_{it}$  is associated with factor  $f_{jt}$ . To make statements about the relevance of variables in the factor model, we have to assess the rows of the loadings matrix, see Section 2. We term a variable  $x_{it}$  to be relevant if zero is outside the 95% HPD region for at least one loading element in row  $i$  of the loadings matrix.
2. Second, we can use posterior association probabilities to assess the relevance of variables. In the sampler, step (20) above, we compute the posterior odds  $P(\lambda_{ij} \neq 0 | \cdot) / P(\lambda_{ij} = 0 | \cdot)$  conditional on the factors, amongst others. Based on this ratio, we decide whether the loading is different from zero,  $\lambda_{ij} \neq 0$ , or not,  $\lambda_{ij} = 0$ . Frühwirth-Schnatter and Lopes (2010) as well as Carvalho et al. (2008) define the indicator matrix  $\gamma_{ij} = 1$  if  $\lambda_{ij} \neq 0$  and  $\gamma_{ij} = 0$  if  $\lambda_{ij} = 0$  and use the posterior distributions of the  $\gamma_{ij}^{(g)}$  for  $g = 1, \dots, G^{\text{eff}}$  to obtain the posterior association probability

$$P(\gamma_{ij} = 1 | X^T) = \frac{1}{G^{\text{eff}}} \sum_{r=1}^{G^{\text{eff}}} \gamma_{ij}^{(r)} \quad (26)$$

of variable  $i$  with factor  $j$ . Again a variable is called relevant for the model, if there is at least one loading element in row  $i$  with  $P(\gamma_{ij} = 1 | X^T) > 0.95$ .

In the simulations, we will consider 10 variables which are relevant or irrelevant in the DGP. To provide a summary statistic across all these variables, we compute the average inclusion frequency. In particular, we count how many of the 10 irrelevant variables are actually regarded as relevant in the estimation of the model given a simulated dataset based on the HPD regions of the loadings and the association probabilities. We then average this number over the  $R$  simulation samples and the irrelevant variables. Assume we simulate from the DGP with irrelevant data. If the average inclusion frequency is close

to zero, we can confirm that the irrelevant variables are detected correctly. An average inclusion frequency close to one indicate that all the irrelevant variables are regarded as relevant in the estimation.

Note that the literature mostly employs posterior association probabilities (26) to assess the relevance of variables rather than the posterior draws of the loadings, see Frühwirth-Schnatter and Lopes (2010). In the exercise below, we also use the HPD regions based on  $\lambda_{ij}^{(g)}$  for  $g = 1, \dots, G^{\text{eff}}$ , as this additionally allows us to address the performance of the normal prior, which does not necessitate the computation of the posterior odds  $P(\lambda_{ij} \neq 0 | \cdot) / P(\lambda_{ij} = 0 | \cdot)$ . In the simulations below, we can check how well the HPD regions and the association probabilities work in separating relevant from irrelevant variables.

#### 4.4 Simulation results

Table 1 contains simulation results, where the last 10 variables are irrelevant in the DGP, and Table 2 contains results with relevant variables in the DGP, see Section 4.1. The standard output in the Tables below consists of five panels. In Panel A, we provide the RMSE as defined above in (25), and in Panel B the RMSE of the model based on the normal and one-layer prior relative to the RMSE of the model estimated using the two-layer sparse prior. Thus, a relative RMSE greater than one indicates a superior performance of the two-layer prior. To evaluate the relative performance of the three priors further, Panel C contains the average differences between two  $\text{RMSE}_r$  based on different priors. The differences are computed as the  $\text{RMSE}_r$  of the one-layer prior minus the  $\text{RMSE}_r$  of the two-layer prior, and the same using the normal prior. If the mean of the sequence of differences is greater than zero, the two-layer prior is superior. We also provide the 95% highest probability density (HPD) region to assess the significance of the differences between priors. Panels D and E of Table 1 contain the average inclusion frequencies of the irrelevant variables. Panel D contains results based on HPD regions of the loadings posterior draws, whereas Panel E contains results based on the posterior association probabilities (26). Panels A and B of Table 1 show that the differences between the RMSEs based on the two-layer prior and one-layer prior are quite small, whereas both clearly improve over the normal prior. The HPD regions in Panel C indicate no systematic differences between the two-layer prior and one-layer prior, whereas the two-layer prior yields a significantly lower RMSE than the normal prior for  $s_0 = 0.1$  and  $s_0 = 0.5$ . However, for a low degree of sparsity in the DGP ( $s_0 = 0.9$ ), the RMSE of the two-layer prior is not significantly lower than the normal prior in this case. These results from



Table 1: Simulation results with 10 irrelevant variables (out of 50) following an AR(1) process, without setting zero loadings a priori for estimation

	$s_0 =$		
	0.1	0.5	0.9
A. RMSE, mean over MC replications			
two-layer prior	0.0918	0.2516	0.3463
one-layer prior	0.0913	0.2514	0.3472
normal prior	0.2100	0.2941	0.3515
B. RMSE, normalized to one for two-layer prior			
two-layer prior	1.0000	1.0000	1.0000
one-layer prior	0.9941	0.9991	1.0025
normal prior	2.2868	1.1690	1.0148
C. RMSE difference, mean over MC replications (95% HPD region below)			
RMSE(one-layer)-RMSE(two-layer)	-0.0006 (-0.0071;0.0070)	-0.0002 (-0.0011;0.0006)	0.0006 (-0.0013;0.0063)
RMSE(normal)-RMSE(two-layer)	0.1173 (0.0952;0.1394)	0.0429 (0.0308;0.0567)	0.0049 (-0.0016;0.0131)
D. Average inclusion frequency based on 95% HPD of $\lambda_{ij}$ with respect to irrelevant variables			
two-layer prior	0.000	0.006	0.004
one-layer prior	0.000	0.006	0.002
normal prior	0.004	0.048	0.026
E. Average inclusion frequency based on $\gamma_{ij}$ with respect to irrelevant variables			
two-layer prior	0.000	0.006	0.004
one-layer prior	0.000	0.006	0.006

Panels A, B, and C are confirmed in Table 2.

Panels D and E of Table 1 contain the results with respect to the irrelevant variables in the DGP. The one- and two layer priors yield an average inclusion frequency of at most 0.6% ( $s_0 = 0.5$ ) and thus help to identify the irrelevant variables in more than 99% of the cases. The results for both sparsity priors do not depend much on the degree of sparsity  $s_0$  used to simulate the first block of data. The normal prior obtains an average inclusion frequency between 0.4% and 4.8%, and thus identifies irrelevant variables correctly in more than 95% of the cases. Thus we can confirm that the performance of all the priors is overall very good with respect to identify irrelevant variables; however, the sparse priors help to separate good from bad indicators slightly better than the normal prior. If we compare the different ways to compute average inclusion frequencies of the irrelevant variables, we see no major differences between Panel D, which contains results based on HPD regions of the loadings posterior draws, and Panel E, which contains results based on the posterior association probabilities (26). Thus, both methods seem to be reliable to find irrelevant variables.

Table 2 contains the results based on relevant variables in the DGP. The one- and two-layer sparse priors obtain similar average inclusion frequencies, which range from 95% ( $s_0 = 0.1$ ) to 99% ( $s_0 = 0.9$ ). The normal prior performs less reliable with very sparse data, where the average inclusion frequency is 77% ( $s_0 = 0.1$ ), but performs only slightly worse than the sparse priors for  $s_0 = 0.5$  and  $s_0 = 0.9$ . Thus, with respect to relevant variables in the DGP, the sparse priors show clearer advantages over the normal prior as with respect to irrelevant data.

To summarize, we find that sparsity priors generally improve the estimation accuracy of the common components in most of the cases over the normal prior. With a decreasing degree of sparsity in the DGP, the efficiency gain obtained using the sparse priors decreases over the normal prior. However, we are in general more flexible to tackle different degrees of heterogeneity in the simulated data by using the hierarchical sparsity priors compared to the normal prior. The two-layer prior cannot systematically improve over the one-layer prior, the differences in RMSEs between the two are insignificant in all cases. Therefore, the richer two-layer prior does not seem to contribute much to the simpler one-layer prior in terms of RMSE.

With respect to finding irrelevant variables, the performance of sparsity priors is only slightly better compared to the normal prior. However, there are clearer advantages of the sparse priors with respect to finding relevant variables. Overall, it seems to pay off having a more general prior specification than the normal prior. Between the one-layer and two-layer prior, we find no systematic significant differences. Furthermore, we find

Table 2: Simulation results with 10 relevant variables (out of 50), without setting zero loadings a priori for estimation

	$s_0 =$		
	0.1	0.5	0.9
A. RMSE, mean over MC replications			
two-layer prior	0.1843	0.2667	0.3040
one-layer prior	0.1838	0.2670	0.3055
normal prior	0.2605	0.2970	0.3104
B. RMSE, normalized to one for two-layer prior			
two-layer prior	1.0000	1.0000	1.0000
one-layer prior	0.9974	1.0012	1.0050
normal prior	1.4139	1.1137	1.0212
C. RMSE difference, mean over MC replications (95% HPD region below)			
RMSE(one-layer)-RMSE(two-layer)	-0.0003 (-0.0018;0.0009)	0.0002 (-0.0009;0.0014)	0.0015 (-0.0006;0.0033)
RMSE(normal)-RMSE(two-layer)	0.0758 (0.0601;0.0943)	0.0299 (0.0203;0.0425)	0.0066 (-0.0025;0.0123)
D. Average inclusion frequency based on 95% HPD of $\lambda_{ij}$ with respect to relevant variables			
two-layer prior	0.952	0.978	0.990
one-layer prior	0.946	0.978	0.992
normal prior	0.774	0.944	0.956
E. Average inclusion frequency based on $\gamma_{ij}$ with respect to relevant variables			
two-layer prior	0.956	0.982	0.992
one-layer prior	0.954	0.982	0.996

that both the HPD regions computed from the posterior samples of individual loadings  $\lambda_{ij}$  as well as the inclusion probabilities obtained from the posterior distribution of the indicator matrix  $\gamma_{ij}$  help to separate relevant from irrelevant variables well. Thus, both ways of searching for irrelevant variables might be helpful in practice.

## 4.5 Checks for robustness

To check for the robustness of the results, we carried out a number of additional simulations. For example, we carried out the same exercise with 25 idiosyncratic variables in the DGP rather than 10. We also ran the simulations with the identification scheme by Geweke and Zhou (1996). Additionally, we changed the persistence in the DGP of the irrelevant variables and the mean of the loadings in the prior distributions, as well as the variance of the loadings. Finally, we carried out simulations with a different DGP for the relevant variables, where both factors explain the relevant variables by definition. In all the cases, we obtained very similar results to those reported above. Some of the results, in particular those based on the identification scheme of Geweke and Zhou (1996), are provided in Appendix C.

## 5 Empirical results

The methods to identify relevant variables in sparse factor models can be regarded as quite general tools of model assessment and might be useful for many different applications. To illustrate the approaches, we will provide empirical examples based on two different economic datasets:

1. We use a multi-country macroeconomic dataset as in Francis et al. (2012), who investigate the role of global business cycles for many countries in a large factor model.
2. We investigate disaggregated US inflation data from Mackowiak et al. (2009) in a sparse factor model setting. Mackowiak et al. (2009) use a factor model as an empirical benchmark to compare alternative theoretical models of price-setting behavior.

The two datasets differ quite a lot in terms of data coverage, and we can expect to find different degrees of heterogeneity and sparsity. We estimate the sparse factor model for each of these datasets to assess the information content in the variables for estimating the factors. To this aim, we identify relevant and irrelevant variables by

searching for zero loadings rows. Since we have found in Section 4 above that HPD regions of loadings and association probabilities perform similarly, we report results for the association probabilities in (26) only. Furthermore, given that the two-layer and the one-layer prior perform very similarly and both tend to outperform the normal prior, we provide empirical results below for the two-layer prior only.

## 5.1 Results for international GDP growth data

The empirical application follows the literature on international business cycles estimated in large factor models such as Kose et al. (2003, 2010) and Del Negro and Otrok (2008). Recently, Francis et al. (2012) estimate global and regional business cycles from a large set of country-specific GDP growth series. In the application below, we follow this literature and investigate the role of global business cycles for each country by estimating the factor model with sparse loadings priors and searching for relevant variables. If variables for a certain country, region or continent are all irrelevant according to our definitions, one could conjecture that these countries are not linked to international business cycles.

We follow Francis et al. (2012) and choose the Penn World Tables (PWT) to construct the dataset. We use the version 7.0, see Heston et al. (2011), and take all those GDP series, which are available from 1960 onwards to match the sample period in Francis et al. (2012). We end up with  $N = 57$  countries, and compute GDP growth rates by taking first differences of the logarithm of each GDP series. We estimate a sparse factor model using this data and assess the relevance of each country in the model.

To estimate the factor model, we choose a specification with  $k = 3$  factors,  $p = 4$  lags in the factor VAR and  $q = 2$  lags in the idiosyncratic components. Checks for robustness are provided below. Regarding the prior distributions for the loadings, we choose the two-layer prior with hyperparameters  $s_0 = 0.5$  and precision  $r_0 = 3$ . The other hyperparameters are specified as before in the Monte Carlo exercise. We sample 15000 posterior draws, and discard the first 1500 replications as burn-in draws. From the remaining replications, we pick every third one.

In Figures 2 and 3, we provide results on the posterior distributions of the loadings  $\lambda_{ij}$  and the association probabilities in (26). We provide two different heat plots: In the first one in Figure 2, black indicates a significant loading defined by  $P(\gamma_{ij} = 1 | X^T) > 0.95$ . The entry is white otherwise. The heat plot in Figure 3 shows the mean of the posterior distribution of  $\lambda_{ij}$ , where from dark red to dark blue color the entries vary from positive to negative values. In both Figures, there are a lot of white regions in the heat plots, indicating that variables are often related only to some of the three factors. On average,

Figure 2: Black-White heat plot indicating significant loadings on international GDP growth data

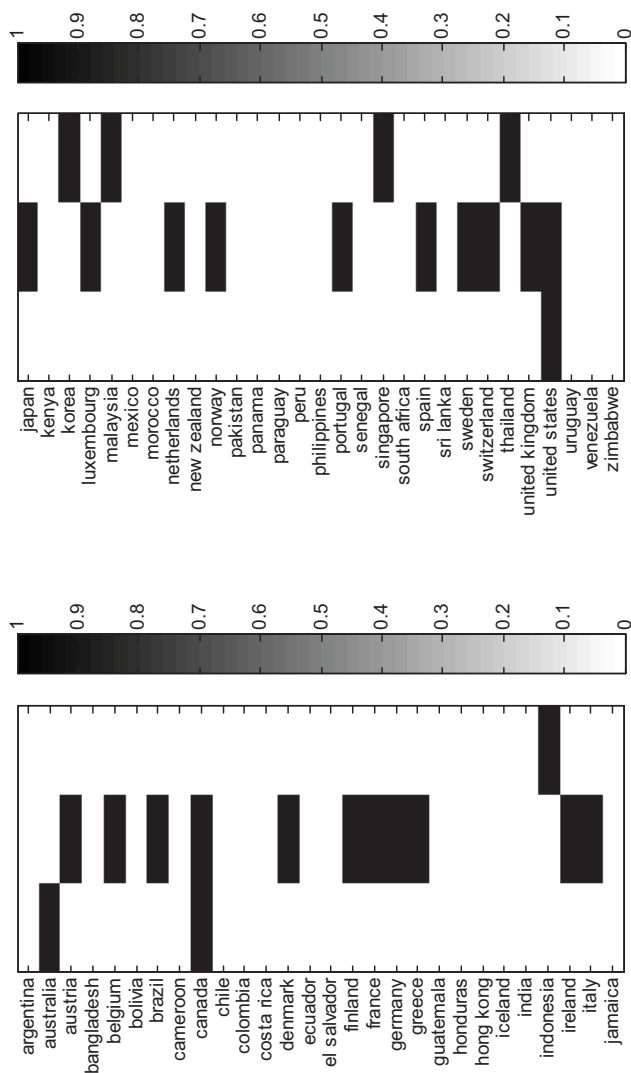
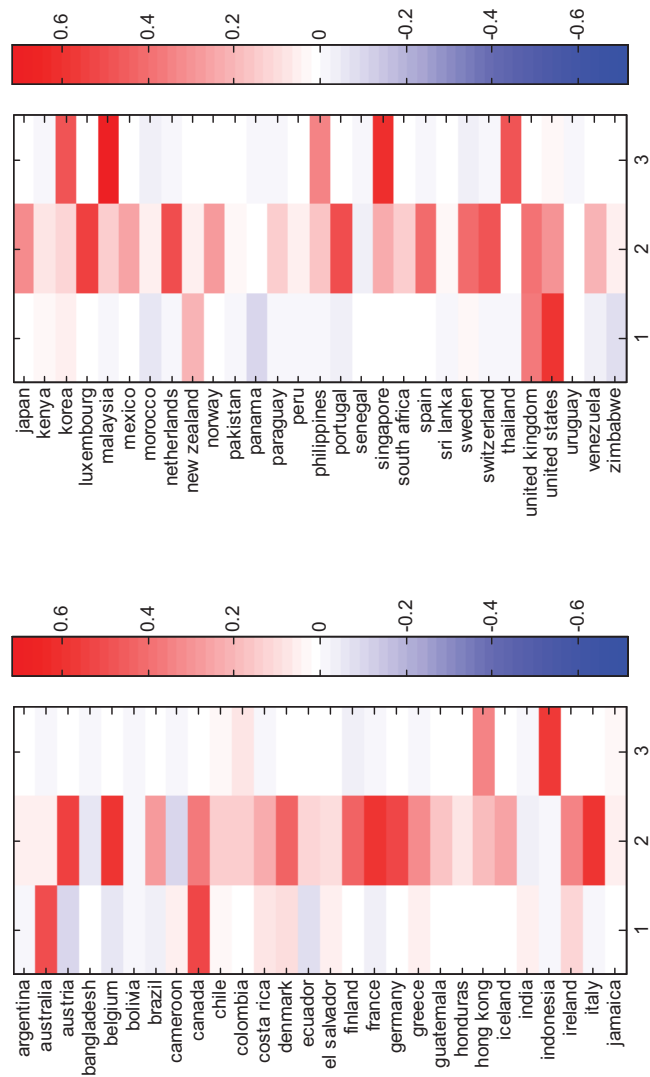


Figure 3: Heat plot for loadings based on international GDP growth data



only 17% of the elements in the loadings matrix are significantly different from zero. We find that 27 out of 57 variables (47.4%) have at least one loading different from zero, defined by a posterior association probabilities being greater than 0.95 for at least one loading element in a row.

If we separate the countries into geographic regions following Kose et al. (2003) and Francis et al. (2012), we can assess, which regions are well explained by the international factors. Kose et al. (2003) divide the world into six regions: Africa, Asia 1 (less developed), Asia 2 (more developed), Europe, Latin America, North America and Oceania. The detailed country classification can be found in Appendix B. In each group, the proportion of variables with at least one loading different from zero is (absolute numbers in brackets): Africa 0% (0 of 6 countries), Asia I 16.7% (1 of 6), Asia II 83.3% (5 of 6), Europe 94.4% (17 of 18), Latin America 6.3% (1 of 16), North America 66.7% (2 of 3), Oceania 50% (1 of 2). According to these results, African countries are not at all explained by international factors. In Latin America, there is only one country related to the international factors. Overall, sparsity seems to matter quite a lot in the international GDP growth data.

To check the robustness of the results, we tried different settings of the hyperparameter  $s_0$ , which is key for determining prior sparsity in the loadings, the number of factors  $k$ , and different identification schemes. In particular, we also use  $s_{0j} \in \{0.25, 0.75\}$ . The results are summarized in Table 3. In all the alternative applications, we can confirm our main finding of sparsity. In particular, the number of variables affected by at least one factor is around 50%. The number of significant loadings varies between 13% and 25% across all factors and variables. This stability of the results with respect to changes in the priors indicate that the priors do not dominate the likelihood in this application. How the different regions of the world are represented by relevant countries can be verified in Table 4. Again, we find that all African countries are not explained by the factors in the different specifications. Latin America is only represented with at most one country. Following the logic of the empirical literature on international business cycles such as Kose et al. (2003, 2010), one could conjecture that these countries are not linked to international business cycles. On the other hand, most or almost all of the variables in the regions Europe, North America, and Asia II, are relevant in the model. These findings are robust across the various estimated specifications.



Table 3: Summary statistics on sparsity based on international growth data with alternative specifications

	variant 1	variant 2	variant 3	variant 4
a-priori zero restrictions	-	-	-	Geweke, Zhou (1996)
$s_0$	0.5	0.5	0.5	0.5
$k$	2	3	4	3
percent of variables with at least one significant loading	49.1	47.4	49.1	47.4
average number of relevant factors for each variable	0.49	0.51	0.53	0.51
percent of significant loadings in $\lambda$	24.6	16.9	13.1	16.9

	variant 5	variant 6
a-priori zero restrictions	-	-
$s_0$	0.25	0.75
$k$	3	3
percent of variables with at least one significant loading	45.6	47.4
average number of relevant factors for each variable	0.49	0.49
percent of significant loadings in $\lambda$	16.4	16.4

Note: We identify a loading as being significantly different from zero, if the inclusion probability is greater than 95%, as measured using the draws from the posterior of  $\gamma_{ij}$ . More details can be found in the main text.

Table 4: Percent of relevant countries by world region under alternative specifications

	variant 1	variant 2	variant 3	variant 4
a-priori zero restrictions	-	-	-	Geweke, Zhou (1996)
$s_0$	0.5	0.5	0.5	0.5
$k$	2	3	4	3
Africa	0.0 (0/6)	0.0 (0/6)	0.0 (0/6)	0.0 (0/6)
Asia I	16.7 (1/6)	16.7 (1/6)	16.7 (1/6)	16.7 (1/6)
Asia II	83.3(5/6)	83.3(5/6)	83.3(5/6)	83.3(5/6)
Europe	100.0 (18/18)	94.4 (17/18)	88.9 (16/18)	94.4 (17/18)
Latin America	6.3 (1/16)	6.3 (1/16)	18.8 (3/16)	6.3 (1/16)
North America	100.0 (2/3)	66.7 (2/3)	66.7 (2/3)	66.7 (2/3)
Oceania	0.0 (0/2)	50.0 (1/2)	50.0 (1/2)	50.0 (1/2)

	variant 5	variant 6
a-priori zero restrictions	-	-
$s_0$	0.25	0.75
$k$	3	3
Africa	0.0 (0/6)	0.0 (0/6)
Asia I	16.7 (1/6)	16.7 (1/6)
Asia II	83.3(5/6)	83.3(5/6)
Europe	88.9 (16/18)	94.4 (17/18)
Latin America	6.3 (1/16)	6.3 (1/16)
North America	66.7 (2/3)	100.0 (2/3)
Oceania	50.0 (1/2)	0.0 (0/2)

Note: We identify a loading as being significantly different from zero, if the inclusion probability is greater than 95%, as measured using the draws from the posterior of  $\gamma_{ij}$ . More details can be found in the main text.

## 5.2 Results for US sectoral inflation data

Mackowiak et al. (2009) estimate a factor model as an empirical benchmark to identify sector-specific and aggregate shocks. Identification is obtained through the assumption that sector-specific shocks are entirely idiosyncratic, whereas aggregate shocks are shocks to the common factors. The estimated impulse-response functions for the subindices of the US consumer price index (CPI) help to differentiate between different theoretical models of price-setting behavior. According to their empirical results focussed on the median sector, Mackowiak et al. (2009) find that price-setters in firms pay much more attention to sector-specific conditions than to aggregate (common) conditions.

The sparse factor model proposed here allows us to provide more results on the relative importance of aggregate and sector-specific determinants. In particular, we can investigate, which sectoral price inflation rates contain relevant information on the factors. Finding the relevant sectoral price inflation rates in the model might help to determine, which sectoral inflation rates comove with others and might be driven by common shocks, and which sectoral inflation rates are not explained by common factors, and thus are irrelevant according to the model.

The disaggregated US inflation data of Mackowiak et al. (2009) contain  $N = 79$  subindices of the US consumer CPI. We follow Mackowiak et al. (2009) and use  $k = 1$  factor. In our model, we also specify  $p = 12$  lags in the factor VAR and  $q = 6$  lags in the idiosyncratic components. The sparsity prior is specified with  $s_0 = 0.5$  and precision  $r_0 = 3$  as before.

According to the estimated inclusion probabilities, we find that 28 (35.4%) of the 79 variables have a non-zero factor loading. Thus, about 65% of the variables are not driven by the common inflation factor. Nakamura and Steinsson (2008) collect subindices of sectoral CPI in groups. If we summarize the results within their categories of data, we find that almost all groups of variables contain at least one series that obtains a non-zero factor loading. In particular, the proportions of variables with at least one non-zero loading are (absolute numbers in brackets): apparel 36.4% (4 of 11 variables), education and communication 25.0% (1 of 4), food and beverages 24.1% (7 of 29), housing 27.3% (3 of 11), medical care 100% (5 of 5), other goods and services 0% (0 of 2), recreation 71.4% (5 of 7), transportation 30.0% (3 of 10). Thus, other goods and services, food and beverages, and education and communication are, respectively, not at all or only weakly associated with the common factor. The other variable groups have two or more sectoral inflation rates that are relevant in the model. More details on the inclusion for each variable in the data are provided in the Figures 4 and 5. Overall, we find that sparsity

Figure 4: Black-White heat plot indicating significant loadings based on disaggregated US CPI inflation data

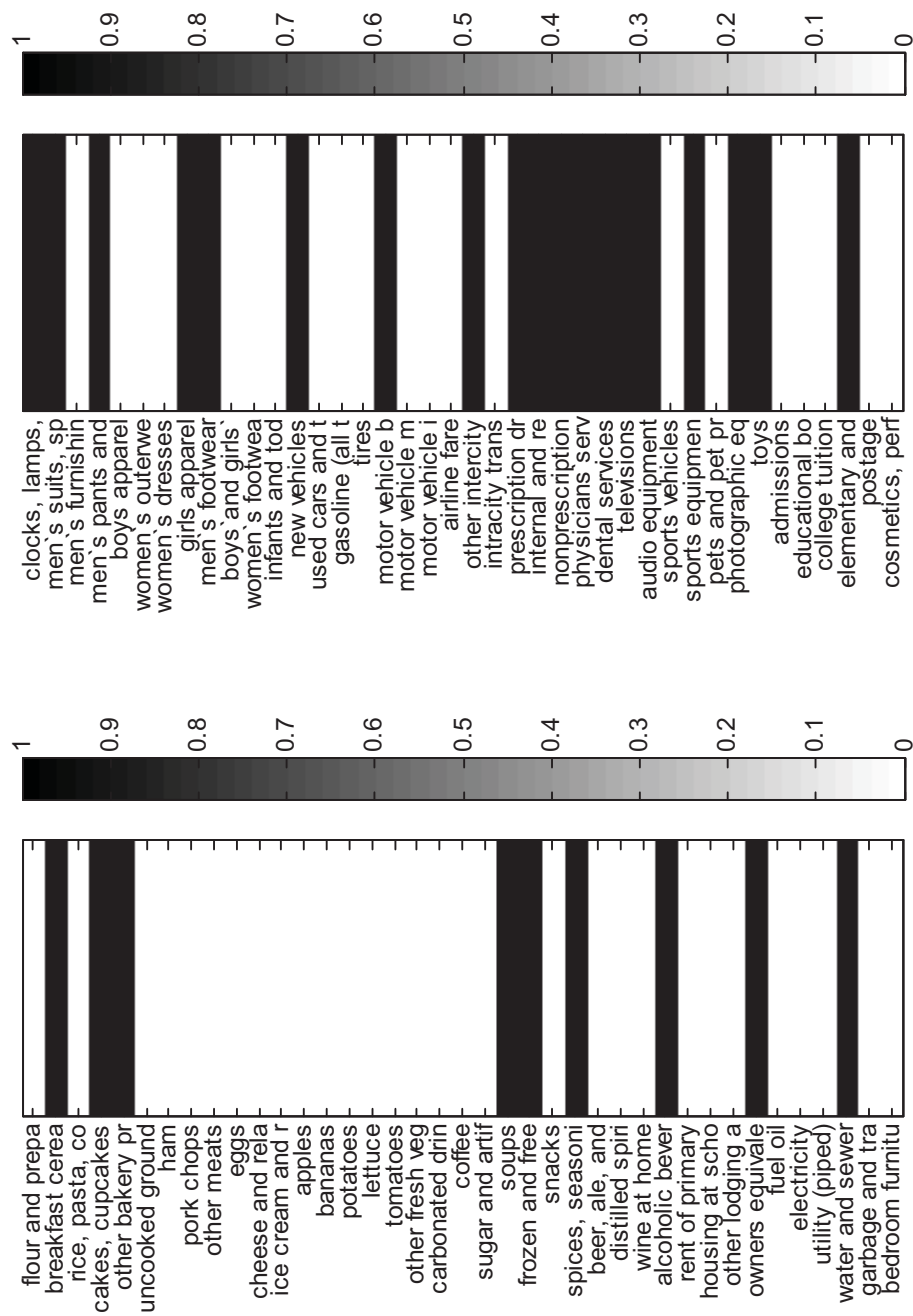
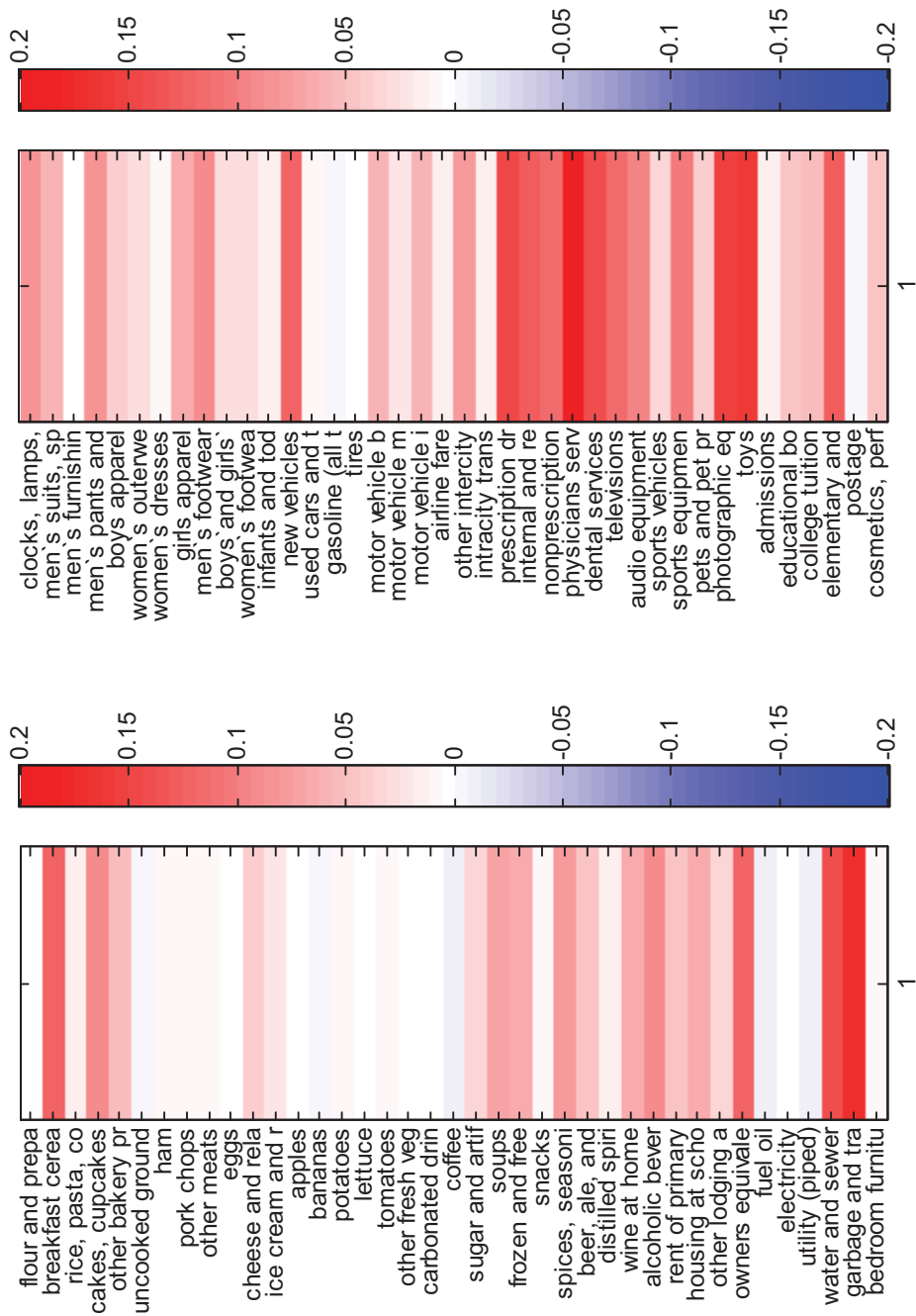


Figure 5: Heat plot for loadings based on disaggregated US CPI inflation data



matters a lot in the disaggregated CPI inflation data. The majority of loadings is not significantly different from zero, and only a subset of the indicators are influenced by the common factor. The results are robust to changes in the model specification, in particular, when the number of factors is increased to  $k = 2$ .

The identification of irrelevant sectoral inflation rates here expands quantitatively on the results by Mackowiak et al. (2009), who find that price-setters in firms mainly pay attention to sector-specific conditions rather than to aggregate conditions. According to our results, we can conclude that about 65% of the sectoral inflation rates are entirely determined by sector-specific developments. Related to shocks, this implies that these sectoral inflation rates are driven by sector-specific shocks only, as irrelevant variables have zero rows in the loadings and thus cannot be influenced by shocks hitting the common factors.

## 6 Conclusion

In this paper, we discuss factor estimation with Bayesian methods, when some of the variables in the data are noisy and perhaps not informative to estimate the factors. In this sense, we define irrelevant variables as being entirely driven by idiosyncratic noise and not the common factors. For various economic questions, it might be helpful to know which variables in the data are actually irrelevant. Based on posterior inference, we provide different ways to separate the irrelevant variables from the relevant ones. We also propose to use a sparse prior distribution of the loadings matrix to sharply discriminate between the two types of variables. This prior is more general than the normal prior, which has often been used in the economic factor model literature so far.

In Monte Carlo simulations we show that the proposed sparse prior specification of the loadings generally improves the estimation accuracy compared to the standard normal prior used in the literature. In particular, estimates with the sparse prior are robust to different degrees of sparsity in the DGP. The higher the degree of sparsity in the data, the better is the estimation performance of the sparse prior over the normal prior. Even in the case of little sparsity in the DGP, working with the sparse prior does generally not lead to a worse estimation performance in comparison to the normal prior. To find relevant variables, we generally obtain good results by either using HPD regions based on the posterior samples of the loadings or posterior association probabilities. Searching for relevant variables is in general successful under different degrees of sparsity in the data.

In the empirical exercise, we search for relevant variables in two datasets from the literature. In the first empirical application, we estimate the sparse factor model using

an international dataset, which is used to analyse global business cycles in the literature by Francis et al. (2012) and Kose et al. (2003). The results show that only less than 50% of the countries are explained by the factors. In particular, Africa and Latin America seem to play only a minor role in international business cycles. An additional application concerns disaggregated CPI inflation in the US following Mackowiak et al. (2009). In this data, we find even more sparsity, as only about 35% of the sectoral inflation rates are informative about the factors. Thus, sectoral price developments seem to be mostly driven by sector-specific rather than aggregate determinants. Overall, we find a high degree of sparsity in both datasets. In general, however, we can expect that results will highly depend on the data and empirical application chosen. Overall, the methods discussed here seem to be flexible tools to estimate factors and determine relevant variables in the presence of heterogeneous information in economic datasets.

# Appendix

## A Posterior distributions

To derive the sampler, we cast model (1) into its state space form:

$$\begin{aligned} X_t &= \mathbf{\Lambda} \mathbf{F}_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \Sigma_\varepsilon) \\ \mathbf{F}_t &= \mathbf{\Phi} \mathbf{F}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(0, \Sigma_\eta) \end{aligned} \quad (27)$$

where the state vector contains  $f_t$  augmented by the appropriate number of lagged values,  $\mathbf{F}_t = (f'_t, \dots, f'_{t-p})'$ . The system matrices  $\mathbf{\Lambda}$  and  $\mathbf{\Phi}$  are of dimension  $N \times k(p+1)$  and square  $k(p+1)$ , respectively:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda & \mathbf{0}_{N \times kp} \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \end{bmatrix}, \quad \text{with} \quad \begin{aligned} \mathbf{\Phi}_1 &= [\Phi_1 \ \dots \ \Phi_p] \\ \mathbf{\Phi}_2 &= [\mathbf{I}_{kp} \ \mathbf{0}_{kp \times k}] \end{aligned}$$

The matrix  $\Sigma_\eta$  contains the identity matrix in the upper left  $k \times k$  sub-matrix.

In short form, the sampler proposed in Section 3.2 consists in iterating over the following steps:

- (i) Simulate  $f^T$  from  $\pi(f^T | X^T, \theta)$ .
- (ii) Simulate the parameters from  $\pi(\theta_{-\lambda} | f^T, X^T, \lambda)$ .
- (iii) Simulate  $\lambda$  from  $\pi(\lambda | f^T, X^T, \Sigma_\varepsilon)$  and update hyperparameters.

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

### A.1 Step (i): Simulate $f^T$ from $\pi(f^T | X^T, \theta)$

Given the prior density in (8) we can factorize the posterior distribution

$$\pi(f^T | X^T, \theta) = \pi(f_T | X^T, \theta) \prod_{t=1}^{T-1} \pi(f_t | f_{t+1}, X^t, \theta) \quad (28)$$

$$= \pi(f_T | X^T, \theta) \prod_{t=1}^{T-1} \pi(f_t | X^t, \theta) \pi(f_t | f_{t+1}, \theta) \quad (29)$$

The factorization has as typical element  $\pi(f_t | X^t, \theta)$  which we call the filter distribution of the unobservable factor  $f_t$  given information up to time  $t$ . Given Gaussian errors, the distribution is normal with mean  $f_{t|t}$  and variance  $V_{t|t}$ . We derive the filter distribution



for  $\mathbf{F}_t$  using the state space form in (27). The relevant moments of  $\pi(f_t|X^t, \theta)$  can be recovered using the selection matrix  $\iota = [I_k \quad \mathbf{0}_{k \times kp}]$  given that:

$$f_t = \iota \mathbf{F}_t$$

Since the state equation is stable and does not contain a constant term, the initial state  $\mathbf{F}_{0|0}$  is set to  $E(\mathbf{F}_t) = \mathbf{0}$ , see Hamilton (1995), p. 378. The initial covariance  $\mathbf{V}_{0|0}$  follows from  $\mathbf{F}_t = \Phi \mathbf{F}_{t-1} + \boldsymbol{\eta}_t$ ,  $E(\mathbf{F}_t \mathbf{F}_t') = \Phi E(\mathbf{F}_{t-1} \mathbf{F}_{t-1}') \Phi' + \Sigma_\eta$  and  $\Sigma_{\mathbf{F}} = \Phi \Sigma_{\mathbf{F}} \Phi' + \Sigma_\eta$ . The vec operator recovers

$$\text{vec}(\Sigma_{\mathbf{F}}) = [\mathbf{I}_{\text{size}(\mathbf{F},1)} - (\Phi \otimes \Phi)]^{-1} \times \text{vec}(\Sigma_\eta) \quad (30)$$

the coefficients  $\Sigma_{\mathbf{F}}$ . Then,

$$\text{vec}(\mathbf{V}_{0|0}) = \text{vec}(\Sigma_{\mathbf{F}}) \quad (31)$$

Given an initial state  $\mathbf{F}_{0|0}$  we iterate through the sample for  $t = 1, \dots, T$  to obtain  $\pi(\mathbf{F}_t|X^t, \theta)$  with mean  $\mathbf{F}_{t|t}$  and covariance  $\mathbf{V}_{t|t}$  given by

$$\mathbf{F}_{t|t} = \mathbf{F}_{t|t-1} + \mathbf{V}'_{t|t-1} \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \mathbf{V}_{t|t-1} \boldsymbol{\Lambda}' + \Sigma_\varepsilon)^{-1} (X_t - \boldsymbol{\Lambda} \mathbf{F}_{t|t-1}) \quad (32)$$

$$\mathbf{V}_{t|t} = \mathbf{V}_{t|t-1} - \mathbf{V}'_{t|t-1} \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \mathbf{V}_{t|t-1} \boldsymbol{\Lambda}' + \Sigma_\varepsilon)^{-1} \boldsymbol{\Lambda} \mathbf{V}_{t|t-1} \quad (33)$$

and  $\mathbf{F}_{t|t-1}$  and  $\mathbf{V}_{t|t-1}$  are obtained by the prediction equations:

$$\mathbf{F}_{t|t-1} = \Phi \mathbf{F}_{t-1|t-1} \quad (34)$$

$$\mathbf{V}_{t|t-1} = \Phi \mathbf{V}'_{t-1|t-1} \Phi' + \Sigma_\eta \quad (35)$$

Then, beginning in  $T$  we first sample  $f_T|X^T, \theta$  from  $N(f_{T|T}, v_{T|T})$  where

$$f_{T|T} = \iota \mathbf{F}_{T|T} \quad (36)$$

$$v_{T|T} = \iota \mathbf{V}_{T|T} \iota' \quad (37)$$

Then, for  $t = T - 1, \dots, 1$ , we derive

$$\pi(f_t|f_{t+1}, X^t, \theta) = N(f_{t|T}, v_{t|T})$$

treating the simulated  $f_{t+1}$  as an additional observation. We update the filter densities

$$\mathbf{F}_{t|T} = \mathbf{F}_{t|t} + \mathbf{V}_{t|t} (\iota \Phi)' (\iota \mathbf{V}_{t+1|T} \iota')^{-1} (f_{t+1} - \iota \mathbf{F}_{t+1|T})$$

$$\mathbf{V}_{t|T} = \mathbf{V}_{t|t} - \mathbf{V}_{t|t} (\iota \Phi)' (\iota \mathbf{V}_{t+1|T} \iota')^{-1} (\iota \Phi) \mathbf{V}_{t|t}$$

Again, the relevant moments  $f_{t|T}$  and  $v_{t|T}$  are obtained using (36) and (37), respectively.

## A.2 Step (ii): Simulate the parameters from $\pi(\theta_{-\lambda}|f^T, X^T, \lambda)$

We block the posterior simulation of the parameters.

The dynamics of the common factors  $\Phi = [\Phi_1 \dots \Phi_p]'$  are jointly sampled from

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

where

$$\begin{aligned} P &= \left( [I_k \otimes f^-]' [I_k \otimes f^-] + P_0^{-1} \right)^{-1} \\ p &= P \times \left( [I_k \otimes f^-]' \text{vec}(f) + P_0^{-1} p_0 \right) \end{aligned}$$

where  $f = [f_{p+1} \dots f_T]'$  and

$$f^- = \begin{bmatrix} f'_p & \cdots & f'_1 \\ \vdots & & \vdots \\ f'_{T-1} & \cdots & f'_{T-p} \end{bmatrix}$$

We simulate  $\sigma_{\varepsilon,i}^2$  from independent  $IG(u_i, U_i)$  distributions,  $i = 1, \dots, N$ , with  $u_i = u_0 + 0.5T$  and  $U_i = U_0 + 0.5 \sum_{t=1}^T (x_{it} - \lambda_i f_t)^2$ .

## B Country coverage in the international GDP growth dataset

Below, the countries covered in the international dataset used in Section above are shown. Each country belongs to a geographical region Africa, Asia 1 (less developed), Asia 2 (more developed), Europe, Latin America, North America and Oceania. The definition of geographical regions follows Kose et al. (2003).

1. Africa:

- (a) Cameroon
- (b) Kenya
- (c) Morocco
- (d) Senegal
- (e) South Africa
- (f) Zimbabwe

2. Asia 1:

- (a) Bangladesh
- (b) India
- (c) Indonesia
- (d) Pakistan
- (e) Philippines
- (f) Singapore
- (g) Sri Lanka

3. Asia 2:

- (a) Hong Kong
- (b) Japan
- (c) Korea
- (d) Malaysia
- (e) Thailand

4. Europe

- (a) Austria
- (b) Belgium
- (c) Denmark
- (d) Finland
- (e) France
- (f) Germany
- (g) Greece
- (h) Iceland
- (i) Ireland
- (j) Italy
- (k) Luxembourg
- (l) Netherlands

- (m) Norway
- (n) Portugal
- (o) Spain
- (p) Sweden
- (q) Switzerland
- (r) United Kingdom

## 5. Latin America

- (a) Argentina
- (b) Bolivia
- (c) Brazil
- (d) Chile
- (e) Colombia
- (f) Costa Rica
- (g) Ecuador
- (h) El Salvador
- (i) Guatemala
- (j) Honduras
- (k) Jamaica
- (l) Panama
- (m) Paraguay
- (n) Peru
- (o) Uruguay
- (p) Venezuela

## 6. North America

- (a) Canada
- (b) Mexico
- (c) USA

7. Oceania

- (a) Australia
- (b) New Zealand

## C Additional simulation results

The two tables in this section contain results under the identifying restrictions following Geweke and Zhou (1996). Table 5 contains simulation results, where the last 10 variables are irrelevant in the DGP, and Table 6 contains results with relevant variables in the DGP, see Section 4.1.

Table 5: Simulation results with 10 irrelevant variables (out of 50) following an AR(1) process, with Geweke and Zhou (1996) identification

	$s_0 =$		
	0.1	0.5	0.9
A. RMSE, mean over MC replications			
two-layer prior	0.0893	0.2516	0.3471
one-layer prior	0.0882	0.2509	0.3473
normal prior	0.2039	0.2902	0.3518
B. RMSE, normalized to one for two-layer prior			
two-layer prior	1.0000	1.0000	1.0000
one-layer prior	0.9879	0.9970	1.0005
normal prior	2.2830	1.1534	1.0134
C. RMSE difference, mean over MC replications (95% HPD region below)			
RMSE(one-layer)-RMSE(two-layer)	-0.0009 (-0.0073;0.0033)	-0.0003 (-0.0143;0.0073)	0.0008 (-0.0113;0.0085)
RMSE(normal)-RMSE(two-layer)	0.1163 (0.0877;0.1390)	0.0392 (0.0224;0.0610)	0.0049 (-0.0086;0.0150)
D. Average inclusion frequency based on 95% HPD of $\lambda_{ij}$ with respect to irrelevant variables			
two-layer prior	0.000	0.006	0.000
one-layer prior	0.000	0.006	0.000
normal prior	0.020	0.076	0.058
E. Average inclusion frequency based on $\gamma_{ij}$ with respect to irrelevant variables			
two-layer prior	0.000	0.006	0.002
one-layer prior	0.000	0.006	0.008

As another robustness check, we also respecified the hyperparameter of the variance of the loadings  $\tau_j$  when estimating the model, see equation (11). When estimating the factor model, the prior on  $\tau_j$  in  $N(0, \tau_j)$  is parameterized by  $IG(2, 2.5)$ , compared to  $IG(2, 0.5)$

Table 6: Simulation results with 10 relevant variables (out of 50), with Geweke and Zhou (1996) identification

	$s_0 =$		
	0.1	0.5	0.9
A. RMSE, mean over MC replications			
two-layer prior	0.1874	0.2686	0.3071
one-layer prior	0.1864	0.2691	0.3074
normal prior	0.2628	0.2973	0.3115
B. RMSE, normalized to one for two-layer prior			
two-layer prior	1.0000	1.0000	1.0000
one-layer prior	0.9949	1.0019	1.0008
normal prior	1.4023	1.1071	1.0142
C. RMSE difference, mean over MC replications (95% HPD region below)			
RMSE(one-layer)-RMSE(two-layer)	-0.0005 (-0.0105;0.0015)	0.0005 (-0.0120;0.0068)	0.0016 (-0.0081;0.0037)
RMSE(normal)-RMSE(two-layer)	0.0729 (0.0569;0.1054)	0.0287 (0.0177;0.0417)	0.0041 (-0.0062;0.0126)
D. Average inclusion frequency based on 95% HPD of $\lambda_{ij}$ with respect to relevant variables			
two-layer prior	0.976	0.976	0.974
one-layer prior	0.978	0.964	0.938
normal prior	0.894	0.992	0.988
E. Average inclusion frequency based on $\gamma_{ij}$ with respect to relevant variables			
two-layer prior	0.982	0.984	0.982
one-layer prior	0.982	0.976	0.944

in the baseline simulations in the main text. Table 7 contains simulation results, where the last 10 variables are irrelevant in the DGP, and Table 8 contains results with relevant variables in the DGP.

Table 7: Simulation results with 10 irrelevant variables (out of 50) following an AR(1) process, without setting zero loadings a priori for estimation, with larger prior variance of loadings

		$s_0 =$		
		0.1	0.5	0.9
A. RMSE, mean over MC replications				
	two-layer prior	0.0935	0.2523	0.3418
	one-layer prior	0.0927	0.2521	0.3426
	normal prior	0.2215	0.2988	0.3508
B. RMSE, normalized to one for two-layer prior				
	two-layer prior	1.0000	1.0000	1.0000
	one-layer prior	0.9907	0.9993	1.0023
	normal prior	2.3682	1.1841	1.0262
C. RMSE difference, mean over MC replications (95% HPD region below)				
	RMSE(one-layer)-RMSE(two-layer)	-0.0006 (-0.0178;0.0117)	-0.0002 (-0.0011;0.0006)	0.0008 (-0.0071;0.0028)
	RMSE(normal)-RMSE(two-layer)	0.1286 (0.1096;0.1562)	0.0475 (0.0272;0.0612)	0.0079 (-0.0030;0.0186)
D. Average inclusion frequency based on 95% HPD of $\lambda_{ij}$ with respect to irrelevant variables				
	two-layer prior	0.000	0.002	0.004
	one-layer prior	0.000	0.002	0.010
	normal prior	0.010	0.038	0.016
E. Average inclusion frequency based on $\gamma_{ij}$ with respect to irrelevant variables				
	two-layer prior	0.000	0.002	0.006
	one-layer prior	0.000	0.002	0.010

Table 8: Simulation results with 10 relevant variables (out of 50), without setting zero loadings a priori for estimation, with larger prior variance of loadings

	$s_0 =$		
	0.1	0.5	0.9
A. RMSE, mean over MC replications			
two-layer prior	0.1856	0.2667	0.2997
one-layer prior	0.1841	0.2668	0.3013
normal prior	0.2764	0.3023	0.3089
B. RMSE, normalized to one for two-layer prior			
two-layer prior	1.0000	1.0000	1.0000
one-layer prior	0.9919	1.0004	1.0054
normal prior	1.4891	1.1336	1.0307
C. RMSE difference, mean over MC replications (95% HPD region below)			
RMSE(one-layer)-RMSE(two-layer)	-0.0005 (-0.0049;0.0091)	0.0000 (-0.0007;0.0014)	0.0014 (-0.0018;0.0049)
RMSE(normal)-RMSE(two-layer)	0.0895 (0.0667;0.1220)	0.0355 (0.0250;0.0467)	0.0093 (0.0046;0.0157)
D. Average inclusion frequency based on 95% HPD of $\lambda_{ij}$ with respect to relevant variables			
two-layer prior	0.966	0.984	0.986
one-layer prior	0.964	0.982	0.992
normal prior	0.540	0.918	0.996
E. Average inclusion frequency based on $\gamma_{ij}$ with respect to relevant variables			
two-layer prior	0.968	0.986	0.994
one-layer prior	0.968	0.986	0.994



## References

- [1] Aguilar, O. and M. West (2000), Bayesian dynamic factor models and portfolio allocation, *Journal of Business & Economic Statistics* 18, 338–357.
- [2] Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, G. Veronese (2010), New Eurocoin: Tracking Economic Growth in Real Time, *Review of Economics and Statistics* 92, 1024-1034.
- [3] Bai, J. and S. Ng (2010), Principal Components Estimation and Identification of the Factors, Columbia University, mimeo.
- [4] Bernanke, B., J. Boivin and P. Elias (2005), Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, *Quarterly Journal of Economics* 120, 387-422.
- [5] Bhattacharya, A., D. Dunson (2011), Sparse Bayesian infinite factor models, *Biometrika* 98, 291-306.
- [6] Boivin, J. and S. Ng (2006), Are More Data Always Better for Factor Analysis, *Journal of Econometrics* 132, 169-194.
- [7] Carter, C. K. and R. Kohn (1994), On Gibbs sampling for state space models, *Biometrika* 84, 541–553.
- [8] Carvalho, C. M. (2006), Structure and Sparsity in High-Dimensional Multivariate Analysis, Ph. D. thesis, Institute of Statistics and Decision Sciences, Duke University.
- [9] Carvalho, C. M., J. Chang, J. E. Lucas, J. R. Nevins, Q. Wang and M. West (2008), High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics, *Journal of the American Statistical Association* 103, 1438-1456.
- [10] Del Negro, M., C. Otrok (2008), Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles, Fed New York, mimeo.
- [11] Francis, N., M. T. Owyang, and Ö. Savascin (2012), An Endogenously Clustered Factor Approach to International Business Cycles, Federal Reserve Bank of St. Louis Working Paper 2012-014A.
- [12] Frühwirth-Schnatter, S. (1994), Data augmentation and dynamic linear models, *Journal of Time Series Analysis* 15, 183–202.

- [13] Frühwirth-Schnatter, S. and H. F. Lopes (2010), Parsimonious Bayesian factor analysis when the number of factors is unknown, Technical report, University of Chicago Booth School of Business.
- [14] George, E. I. and McCulloch, R. E. (1993), Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88, 881-889.
- [15] George, E. I. and McCulloch, R. E. (1997), Approaches for Bayesian Variable Selection, *Statistica Sinica* 7, 339-373.
- [16] Geweke, J. (1996), Variable selection and model comparison in regression. *Bayesian Statistics* 5, 609-620.
- [17] Geweke, J. and G. Zhou (1996), Measuring the pricing error of the arbitrage pricing theory, *The Review of Financial Studies* 9, 557–587.
- [18] Heston, A., R. Summers, and B. Aten (2011), Penn World Table Version 7.0, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, June.
- [19] Kaufmann, S. and C. Schumacher (2012), Bayesian estimation of sparse dynamic factor models with order-independent identification, mimeo.
- [20] Knowles, D. and Z. Ghahramani (2011), Nonparametric Bayesian sparse factor models with application to gene expression modelling, *Annals of Applied Statistics* 5(2B), 1534-1552.
- [21] Koopman, S. J. and A. Harvey (2003), Computing observation weights for signal extraction and Filtering, *Journal of Economic Dynamics & Control* 27, 1317–1333.
- [22] Kose, M. A., C. Otrok and C. H. Whiteman (2003), International Business Cycles: World, Region and Country Specific Factors, *American Economic Review* 93, 1216-1239.
- [23] Kose, M. A., C. Otrok and C. H. Whiteman (2010), Understanding the evolution of world business cycles, *Journal of International Economics* 75, 110-130.
- [24] Lopes, H. F. and M. West (2004), Bayesian model assessment in factor analysis, *Statistica Sinica* 14, 41–67.

- [25] 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. Vanucci (Eds.), *Bayesian Inference for Gene Expression and Proteomics*, Cambridge University Press.
- [26] Mackowiak, B., E. Mönch and M. Wiederholt (2009), Sectoral price data and models of price setting, *Journal of Monetary Economics* 56, S78-S99.
- [27] Nakamura, E. and J. Steinsson (2008), Five facts about prices: a reevaluation of menu cost models, *Quarterly Journal of Economics* 123, 1415–1464.
- [28] Shephard, N. (1994), Partial non-Gaussian state space, *Biometrika* 81, 115–131.
- [29] Stock, J. and M. Watson (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics* 20, 147–162.
- [30] West, M. (2003), Bayesian factor regression models in the "large p, small n" paradigm, In: Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., and West, M. (eds.), *Bayesian Statistics 7*, Oxford University Press, 723–732.

## **The following Discussion Papers have been published since 2012:**

01	2012	A user cost approach to capital measurement in aggregate production functions	Thomas A. Knetsch
02	2012	Assessing macro-financial linkages: a model comparison exercise	Gerke, Jonsson, Kliem Kolasa, Lafourcade, Locarno Makarski, McAdam
03	2012	Executive board composition and bank risk taking	A. N. Berger T. Kick, K. Schaeck
04	2012	Stress testing German banks against a global cost-of-capital shock	Klaus Duellmann Thomas Kick
05	2012	Regulation, credit risk transfer with CDS, and bank lending	Thilo Pausch Peter Welzel
06	2012	Maturity shortening and market failure	Felix Thierfelder
07	2012	Towards an explanation of cross-country asymmetries in monetary transmission	Georgios Georgiadis
08	2012	Does Wagner's law ruin the sustainability of German public finances?	Christoph Priesmeier Gerrit B. Koester
09	2012	Bank regulation and stability: an examination of the Basel market risk framework	Gordon J. Alexander Alexandre M. Baptista Shu Yan
10	2012	Capital regulation, liquidity requirements and taxation in a dynamic model of banking	Gianni De Nicolò Andrea Gamba Marcella Lucchetta
11	2012	Credit portfolio modelling and its effect on capital requirements	Dilek Bülbül Claudia Lambert

12	2012	Trend growth expectations and U.S. house prices before and after the crisis	Mathias Hoffmann Michael U. Krause Thomas Laubach
13	2012	The PHF: a comprehensive panel survey on household finances and wealth in Germany	Ulf von Kalckreuth Martin Eisele, Julia Le Blanc Tobias Schmidt, Junyi Zhu
14	2012	The effectiveness of monetary policy in steering money market rates during the financial crisis	Puriya Abbassi Tobias Linzert
15	2012	Cyclical adjustment in fiscal rules: some evidence on real-time bias for EU-15 countries	Gerhard Kempkes
16	2012	Credit risk connectivity in the financial industry and stabilization effects of government bailouts	Jakob Bosma Micheal Koetter Michael Wedow
17	2012	Determinants of bank interest margins: impact of maturity transformation	O. Entrop, C. Memmel B. Ruprecht, M. Wilkens
18	2012	Tax incentives and capital structure choice: evidence from Germany	Thomas Hartmann-Wendels Ingrid Stein, Alwin Stöter
19	2012	Competition for internal funds within multinational banks: foreign affiliate lending in the crisis	Cornelia Düwel Rainer Frey
20	2012	Fiscal deficits, financial fragility, and the effectiveness of government policies	Markus Kirchner Sweder van Wijnbergen
21	2012	Saving and learning: theory and evidence from saving for child's college	Junyi Zhu

22	2012	Relationship lending in the interbank market and the price of liquidity	Falk Bräuning Falko Fecht
23	2012	Estimating dynamic tax revenue elasticities for Germany	Gerrit B. Koester Christoph Priesmeier
24	2012	Identifying time variability in stock and interest rate dependence	Michael Stein, Mevlud Islami Jens Lindemann
25	2012	An affine multifactor model with macro factors for the German term structure: changing results during the recent crises	Arne Halberstadt Jelena Stapf
26	2012	Determinants of the interest rate pass-through of banks – evidence from German loan products	Tobias Schlüter, Ramona Busch Thomas Hartmann-Wendels Sönke Sievers
27	2012	Early warning indicators for the German banking system: a macroprudential analysis	Nadya Jahn Thomas Kick
28	2012	Diversification and determinants of international credit portfolios: evidence from German banks	Benjamin Böninghausen Matthias Köhler
29	2012	Finding relevant variables in sparse Bayesian factor models: economic applications and simulation results	Sylvia Kaufmann Christian Schumacher

## The following Discussion Papers have been published since 2011:

### Series 1: Economic Studies

01	2011	Long-run growth expectations and “global imbalances”	M. Hoffmann M. Krause, T. Laubach
02	2011	Robust monetary policy in a New Keynesian model with imperfect interest rate pass-through	Rafael Gerke Felix Hammermann
03	2011	The impact of fiscal policy on economic activity over the business cycle – evidence from a threshold VAR analysis	Anja Baum Gerrit B. Koester
04	2011	Classical time-varying FAVAR models – estimation, forecasting and structural analysis	S. Eickmeier W. Lemke, M. Marcellino
05	2011	The changing international transmission of financial shocks: evidence from a classical time-varying FAVAR	Sandra Eickmeier Wolfgang Lemke Massimiliano Marcellino
06	2011	FiMod – a DSGE model for fiscal policy simulations	Nikolai Stähler Carlos Thomas
07	2011	Portfolio holdings in the euro area – home bias and the role of international, domestic and sector-specific factors	Axel Jochem Ute Volz
08	2011	Seasonality in house prices	F. Kajuth, T. Schmidt
09	2011	The third pillar in Europe: institutional factors and individual decisions	Julia Le Blanc
10	2011	In search for yield? Survey-based evidence on bank risk taking	C. M. Buch S. Eickmeier, E. Prieto

11	2011	Fatigue in payment diaries – empirical evidence from Germany	Tobias Schmidt
12	2011	Currency blocs in the 21 <sup>st</sup> century	Christoph Fischer
13	2011	How informative are central bank assessments of macroeconomic risks?	Malte Knüppel Guido SchulteFrankenfeld
14	2011	Evaluating macroeconomic risk forecasts	Malte Knüppel Guido SchulteFrankenfeld
15	2011	Crises, rescues, and policy transmission through international banks	Claudia M. Buch Cathérine Tahmee Koch Michael Koetter
16	2011	Substitution between net and gross settlement systems – A concern for financial stability?	Ben Craig Falko Fecht
17	2011	Recent developments in quantitative models of sovereign default	Nikolai Stähler
18	2011	Exchange rate dynamics, expectations, and monetary policy	Qianying Chen
19	2011	An information economics perspective on main bank relationships and firm R&D	D. Hoewer T. Schmidt, W. Sofka
20	2011	Foreign demand for euro banknotes issued in Germany: estimation using direct approaches	Nikolaus Bartzsch Gerhard Rösl Franz Seitz
21	2011	Foreign demand for euro banknotes issued in Germany: estimation using indirect approaches	Nikolaus Bartzsch Gerhard Rösl Franz Seitz



22	2011	Using cash to monitor liquidity – implications for payments, currency demand and withdrawal behavior	Ulf von Kalckreuth Tobias Schmidt Helmut Stix
23	2011	Home-field advantage or a matter of ambiguity aversion? Local bias among German individual investors	Markus Baltzer Oscar Stolper Andreas Walter
24	2011	Monetary transmission right from the start: on the information content of the eurosystem’s main refinancing operations	Puriya Abbassi Dieter Nautz
25	2011	Output sensitivity of inflation in the euro area: indirect evidence from disaggregated consumer prices	Annette Fröhling Kirsten Lommatzsch
26	2011	Detecting multiple breaks in long memory: the case of U.S. inflation	Uwe Hassler Barbara Meller
27	2011	How do credit supply shocks propagate internationally? A GVAR approach	Sandra Eickmeier Tim Ng
28	2011	Reforming the labor market and improving competitiveness: an analysis for Spain using FiMod	Tim Schwarzmüller Nikolai Stähler
29	2011	Cross-border bank lending, risk aversion and the financial crisis	Cornelia Düwel, Rainer Frey Alexander Lipponer
30	2011	The use of tax havens in exemption regimes	Anna Gumpert James R. Hines, Jr. Monika Schnitzer
31	2011	Bank-related loan supply factors during the crisis: an analysis based on the German bank lending survey	Barno Blaes

32	2011	Evaluating the calibration of multi-step-ahead density forecasts using raw moments	Malte Knüppel
33	2011	Optimal savings for retirement: the role of individual accounts and disaster expectations	Julia Le Blanc Almuth Scholl
34	2011	Transitions in the German labor market: structure and crisis	Michael U. Krause Harald Uhlig
35	2011	U-MIDAS: MIDAS regressions with unrestricted lag polynomials	C. Forni M. Marcellino, C. Schumacher

## Series 2: Banking and Financial Studies

01	2011	Contingent capital to strengthen the private safety net for financial institutions: Cocos to the rescue?	George M. von Furstenberg
02	2011	Gauging the impact of a low-interest rate environment on German life insurers	Anke Kablau Michael Wedow
03	2011	Do capital buffers mitigate volatility of bank lending? A simulation study	Frank Heid Ulrich Krüger
04	2011	The price impact of lending relationships	Ingrid Stein
05	2011	Does modeling framework matter? A comparative study of structural and reduced-form models	Yalin Gündüz Marliese Uhrig-Homburg
06	2011	Contagion at the interbank market with stochastic LGD	Christoph Memmel Angelika Sachs, Ingrid Stein
07	2011	The two-sided effect of financial globalization on output volatility	Barbara Meller
08	2011	Systemic risk contributions: a credit portfolio approach	Klaus Düllmann Natalia Puzanova
09	2011	The importance of qualitative risk assessment in banking supervision before and during the crisis	Thomas Kick Andreas Pfingsten
10	2011	Bank bailouts, interventions, and moral hazard	Lammertjan Dam Michael Koetter
11	2011	Improvements in rating models for the German corporate sector	Till Förstemann

12	2011	The effect of the interbank network structure on contagion and common shocks	Co-Pierre Georg
13	2011	Banks' management of the net interest margin: evidence from Germany	Christoph Memmel Andrea Schertler
14	2011	A hierarchical Archimedean copula for portfolio credit risk modelling	Natalia Puzanova
15	2011	Credit contagion between financial systems	Natalia Podlich Michael Wedow
16	2011	A hierarchical model of tail dependent asset returns for assessing portfolio credit risk	Natalia Puzanova
17	2011	Contagion in the interbank market and its determinants	Christoph Memmel Angelika Sachs
18	2011	Does it pay to have friends? Social ties and executive appointments in banking	A. N. Berger, T. Kick M. Koetter, K. Schaeck

## **Visiting researcher at the Deutsche Bundesbank**

The Deutsche Bundesbank in Frankfurt is looking for a visiting researcher. Among others under certain conditions visiting researchers have access to a wide range of data in the Bundesbank. They include micro data on firms and banks not available in the public. Visitors should prepare a research project during their stay at the Bundesbank. Candidates must hold a PhD and be engaged in the field of either macroeconomics and monetary economics, financial markets or international economics. Proposed research projects should be from these fields. The visiting term will be from 3 to 6 months. Salary is commensurate with experience.

Applicants are requested to send a CV, copies of recent papers, letters of reference and a proposal for a research project to:

Deutsche Bundesbank  
Personalabteilung  
Wilhelm-Epstein-Str. 14

60431 Frankfurt  
GERMANY





