Dynamic Sparse Factor Analysis

Kenichiro McAlinn[*], Enakshi Saha[†] and Veronika Rockova[‡]

June 7, 2021

Abstract

Its conceptual appeal and effectiveness has made latent factor modeling an indispensable tool for multivariate data analysis. Despite its popularity across many fields, there are outstanding methodological challenges that have hampered deployments in practice. One major challenge is the selection of the number of factors. This issue is exacerbated in dynamic factor models where factors can disappear, emerge, and/or reoccur over time. Existing models that assume a known fixed number of factors may provide a misguided data representation, especially when the factor dimension is grossly misspecified. Another challenge is interpretability which is often regarded as an unattainable objective due to the lack of identifiability. Motivated by a topical macroeconomic application, we develop a flexible Bayesian method for dynamic factor analysis (DFA) that can simultaneously accommodate a time-varying number of factors and enhance interpretability through sparse mode detection. To this end, we turn to dynamic sparsity by employing Dynamic Spike-and-Slab (DSS) priors within DFA. Scalable Bayesian EM estimation is proposed for fast posterior mode identification via rotations to sparsity, enabling Bayesian data analysis at larger scales. We study a high-dimenisional balanced panel of macroeconomic variables covering multiple facets of the US economy, with a focus on the Great Recession, to highlight the efficacy and usefulness of our proposed method.

Keywords: Dynamic Sparsity, Factor Analysis, Spike-and-Slab, Time Series.

[*] *Assistant Professor in Statistics at the Fox School of Business of Temple University*

[†] *Ph.D. Student at the Department of Statistics of the University of Chicago*

[‡] *Associate Professor in Econometrics and Statistics and James S. Kemper Faculty Scholar at the University of Chicago Booth School of Business.* The authors gratefully acknowledge support from the James S. Kemper Foundation at the Booth School of Business and the National Science Foundation (NSF: DMS 1944740).

1 Introduction

The premise of dynamic factor analysis (DFA) is fairly straightforward: there are unobservable commonalities in the variation of observable time series, which can be exploited for interpretation, forecasting, and decision making. Dating back to, at least, Burns and Mitchell (1947), the fundamental idea that a small number of indices drive co-movements of many time series has found plentiful empirical support across a wide range of applications including economics (Bai and Ng, 2002; Baumeister, Liu, and Mumtaz, Baumeister et al.; Bernanke et al., 2005; Cheng et al., 2016; Stock and Watson, 2002), finance (Aguilar et al., 1998; Aguilar and West, 2000; Carvalho et al., 2011; Diebold and Nerlove, 1989; Pitt and Shephard, 1999), and ecology (Zuur et al., 2003), to name just a few. More notably, in their seminal work on DFA, Sargent et al. (1977) showed that two dynamic factors could explain a large fraction of the variance of U.S. quarterly macroeconomic variables. Motivated by a similar (but significantly larger) application, we develop scalable Bayesian DFA methodology to glean insights into the hidden drivers of the U.S. macroeconomy before, during and after the Great Recession.

With large-scale cross sectional data becoming readily available, the need for developing scalable and reliable tools adept at capturing complex latent dynamics have spurred in both statistics and econometrics (Frühwirth-Schnatter and Lopes, 2018; Kaufmann and Beyeler, 2018; Kaufmann and Schumacher, 2017; Nakajima et al., 2017). A wide variety of factor-type models exist with varying degrees of modeling flexibility. One popular class is factor stochastic volatility models (Aguilar and West, 2000; Kastner et al., 2017; Pitt and Shephard, 1999) which, in their simplest form, assume (a) constant loadings, (b) independent factors, and (c) time-varying structures on residual variances and factor variances. Extensions to time-varying loadings that allow for more flexible correlation modeling have been considered in, e.g., Aguilar et al. (1998); Aguilar and West (1998, 2000); Lopes and Carvalho (2007) and later extended by Nakajima and West (2013a,b); Nakajima et al. (2017) to sparse factor models via latent thresholding. Sparsity has also been a key component in dynamic covariance estimation models, such as in Kastner (2019), who proposes a factor stochastic volatility model in combination with a global–local shrinkage prior. This prior is a generalization of the Bayesian Lasso (Park and Casella, 2008) and has also been adopted in the context of Bayesian vector autoregressive (VAR) models (Huber and Feldkircher, 2019; Kastner and Huber, 2020) that are capable of handling vast-dimensional time series. Other developments include large-scale Bayesian VAR methods (Bańbura et al., 2010; Giannone et al., 2014, 2015; Koop and Korobilis, 2013; Korobilis, 2013; Kuschnig and Vashold, 2019). More recently, Koop et al. (2019) and Aunsri and Taveeapiradeecharoen (2020) extended random compression dynamic regression methods (Guhaniyogi and Dunson, 2015) to the VAR framework giving rise to the Bayesian Compressed VAR (BCVAR) model that exhibits an impressive forecasting performance in high dimensions. VAR models have also been integrated within dynamic factor structures in factor augmented vector autoregressive (FAVAR) models (Bernanke et al., 2005; Stock and Watson, 2005) and in their recently introduced sparse extension (Kaufmann and Beyeler, 2018). These FAVAR models have been particularly effective in high-dimensional macroeconomic applications (Daniele and Schnaitmann, 2019; Evgenidis et al., 2019; Potjagailo, 2017; Wagan et al., 2019). A detailed discussion on FAVAR models in macroeconomics can be found in Stock and Watson (2016).

Sparsity is an indispensable tool in high dimensional inference situations where the number of parameters exceed the number of observations by a large extent. The fundamental goal of our research is to build a dynamic factor analysis method that discovers a dynamic sparse factor structure with an unknown and possibly time-varying number of latent factors and with factor loading matrices that evolve somewhat smoothly over time. There are three important ingredients of dynamic sparsity that reside at the core of our methodology.

Firstly, the latent factor loadings should account for time-varying patterns of sparsity. In (macro-)economics and finance, the sequentially observed variables may go through multiple periods of shocks, expansions, and contractions (Hamilton, 1989). It is thus expected that the underlying latent structure changes over time– either gradually or suddenly– where some factors might be active at all times, while others only at certain times. For example, in our empirical analysis we find that certain factors exert influence on some series only during a crisis and later permeate through different components of the economy as the shock spreads. Dynamic sparsity plays a very compelling role in capturing and characterizing such dynamics. Recent developments in sparse factor analysis reflect this direction of interest (Carvalho et al., 2008; Lopes et al., 2010; West, 2003; Yoshida and West, 2010). More recently, Nakajima and West (2013b); Nakajima et al. (2017) deployed the latent threshold approach of Nakajima and West (2013a) in order to induce zero loadings dynamically over time. Our methodological contribution builds on this development, but poses less practical limitations on the dimensionality of the data.

Related to the previous point is the question of selecting the number of factors. This modeling choice is traditionally determined by a combination of *a-priori* knowledge, a visual inspection of the scree plot (Onatski, 2009), and/or information criteria (Bai and Ng, 2002; Hallin and Liska, 2007). In the presence of model uncertainty, the Bayesian approach affords the opportunity to assign a probabilistic blanket over various models. Bayesian non-parametric approaches have been considered for estimating the factor dimensionality using sparsity inducing priors (Bhattacharya and Dunson, 2011; Rockova and George, 2016). The added difficulty stemming from time series data, however, is that the number of factors *may change over time* (Bai and Ng, 2002). As a remedy, we turn to dynamic sparsity as a compass for determining the number of factors without necessarily committing to one fixed number ahead of time.

The third essential requirement is accounting for structural instabilities over time with time-varying loadings and/or factors. One seemingly simple solution has been to deploy rolling/extending window approaches to obtain pseudo-dynamic loadings. These estimates, however, lack any supporting probabilistic structure that would induce smoothness and/or capture sudden dynamics. Recent DFA developments (Del Negro and Otrok, 2008; Kaufmann and Schumacher, 2019; Nakajima and West, 2013a) have treated both the factors and loadings as stochastic and dynamic. Adopting this point of view, we blend smoothness with sparsity via Dynamic Spike-and-Slab (DSS) priors on factor loadings (Rockova et al., 2020). This prior regards factor loadings as arising from a mixture of two states: an inactive state represented by very small loadings and an active state represented by smoothly evolving large loadings. The mixing weights between these two states themselves are time-varying, reflecting past information to prevent from erratic regime switching. The DSS priors allow latent factors to effectively, and smoothly, appear or disappear from each series, tracking the evolution of sparsity over time.

In this work, we develop methodology for sparse dynamic factor analysis that is built on the three principles mentioned above. Using this methodology, we examine a large-scale balanced panel of macroeconomic indices that span multiple corners of the U.S. economy from 2001 to 2015. Our method helps understand how the economy evolves over time and how shocks affect its individual components. In particular, examining the latent factor structure before, during, and after the Great Recession, we obtain insights into the channels of dependencies and we assess permanence of structural changes.

To ensure that our implementation scales with large datasets, we propose an EM algorithm for MAP estimation that recovers evolving sparse latent structures in a fast and potent manner. An important consideration for any factor analysis tool is the interpretability of the latent factors. While interpretation can be achieved with ex-post rotations (Bai and Ng, 2013; Kaufmann and Schumacher, 2017, 2019), here we deploy parameter expansion, with rotations to sparsity *inside* the EM algorithm (Section 3.1) along the lines of Rockova and George (2016) to (a) accelerate convergence and (b) obtain better oriented sparse solutions. We also provide a more traditional estimation strategy using MCMC (Section 3.2) using the conventional lower triangular identification constraint (Nakajima and West, 2013a,b) on the factor loading matrices.

The paper is structured as follows. Section 2 outlines the dynamic sparse factor model. Section 3 summarizes our estimation strategy with a parameter expanded EM algorithm, followed by an alternative MCMC implementation technique. A detailed simulation study that highlights the interpretability of our strategy relative to other methods is in Section 4, followed by an empirical study on a large-scale macroeconomic dataset in Section 5. In Section 6, we demonstrate the forecasting accuracy of our method, compared to some key competitors, on simulated and real datasets. We conclude the paper with additional comments in Section 7. Details of the implementation are in the Supplementary Materials.

2 Dynamic Sparse Factor Models

The data setup under consideration consists of a matrix of high-dimensional multivariate time series $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T] \in \mathbb{R}^{P\times T}$, where each vector $\boldsymbol{Y}_t \in \mathbb{R}^P$ contains a snapshot of continuous measurements at time t. Dynamic factor models are built on the premise that there are only a few latent factors that drive the co-movements of $\boldsymbol{Y}_t$. Evolving covariance patterns of time series can be captured with the following state space model:

$$\boldsymbol{Y}_t = \boldsymbol{B}_t\boldsymbol{\omega}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{ind}{\sim} \mathcal{N}_P(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{1}$$

$$\boldsymbol{\omega}_t = \boldsymbol{\Phi}\boldsymbol{\omega}_{t-1} + \mathbf{e}_t, \quad \mathbf{e}_t \overset{ind}{\sim} \mathcal{N}_K(\mathbf{0}, \mathbb{I}_K), \tag{2}$$

which extends the more standard dynamic factor models (Geweke, 1977; Sargent et al., 1977) in at least two ways. First, the observation equation (1) links $\boldsymbol{Y}_t$ to a vector of factors $\boldsymbol{\omega}_t$ through multivariate regression with loadings $\boldsymbol{B}_t \in \mathbb{R}^{P\times K}$ and with residual variances $\boldsymbol{\Sigma}_t = \mathrm{diag}\{\sigma^2_{1t}, \ldots, \sigma^2_{Pt}\}$, where *both* $\boldsymbol{B}_t$ and $\boldsymbol{\Sigma}_t$ are *dynamic*, i.e. are allowed to evolve over time. In this section, we tacitly assume that any location shifts in $\boldsymbol{Y}$ have been standardized away and thereby we omit an intercept in (1). The (dynamic) intercept can be however included, as we demonstrate in Section 5. Second, the transition equation (2) describes the unobserved factors $\boldsymbol{\omega}_t$ as following a stationary autoregressive process with a transition matrix $\boldsymbol{\Phi} = diag(\widetilde{\phi}_1, \ldots, \widetilde{\phi}_K)$ with $0 < \widetilde{\phi}_k < 1$ for $k = 1, \ldots, K$ and with Gaussian disturbances $\mathbf{e}_t$. As is customary with state-space models of this type, we assume that $\boldsymbol{\omega}_t$, $\mathbf{e}_t$ and $\boldsymbol{\epsilon}_t$ are cross-sectionally independent.

A related approach was proposed in Aguilar and West (2000) and Lopes and Carvalho (2007), who also permit time-varying loadings, but do not impose the AR(1) process on the factors. Instead, their factors are cross-sectionally independent and linked over time through a stochastic volatility evolution of their idiosyncratic variances. Bai and Ng (2002) and Stock and Watson (2010), on the other hand, assume that factors follow vector autoregression, but the loadings are constant over time. As in Nakajima and West (2013b), our model (1) and (2) differs from these more standard dynamic factor model formulations because it combines the AR(1) factor aspect together with dynamic loadings. A few remarks are in place. The assumption of independent and homoscedastic factor innovations may be unnecessarily restrictive. Estimating the factor covariance matrix in our framework is precluded due to lack of identification. This is because our auxiliary covariance matrices $\boldsymbol{A}_t$ (in the expanded model, to be described in Section **??**) are not linked over time. We use parameter expansion to intentionally over-parametrize (Section 3.1) as a computational trick rather than as an attempt to model the co-volatilities. However, in related work Zhou et al. (2014) assume the factor covariance matrices $\boldsymbol{A}_t$ to be non-diagonal and time varying. These matrices can be reduced to diagonal matrices $\boldsymbol{\Psi}_t$ by pre and post multiplying by lower triangular matrices $\mathbf{L_t}$, with diagonal elements equal to one, which are in fact the Cholesky factors of the matrices $\boldsymbol{A}_t$. Suitable dynamic priors are then imposed on individual elements of both $\boldsymbol{\Psi}_t$ and $\boldsymbol{L}_t$. The model is made identifiable by assuming the factor loading matrices $\boldsymbol{B}_t$ to be lower-triangular. Another way to makes the identification problem less severe would be assuming certain dynamics on $\boldsymbol{A}_t$ with identifiability inherited from the initial condition $\boldsymbol{A}_0$.

The equations (1) and (2) imply that, marginally, $\boldsymbol{Y}_t \sim \mathcal{N}_P(0, \widetilde{\boldsymbol{\Sigma}}_t)$, where $\widetilde{\boldsymbol{\Sigma}}_t = \boldsymbol{B}_t \mathbf{V} \boldsymbol{B}_t' + \boldsymbol{\Sigma}_t$ with $\mathbf{V}$ being the stationary autoregressive covariance matrix of the latent factors.[1] This decomposition provides a fundamental justification for factor-based dynamic covariance modeling. The information in high-dimensional vectors $\boldsymbol{Y}_t$ is distilled through latent factors into lower-dimensional factor loadings matrices $\boldsymbol{B}_t$, which *completely* characterize the movements of covariances over time. Other authors (Del Negro and Otrok, 2008; Lopes and Carvalho, 2007) consider a stochastic volatility (SV) evolution (either log-AR(1) or Bayesian discounting) on the variance of the latent factors and/or the innovations $\boldsymbol{\epsilon}_t$ in (1). While both are feasible within our framework, here we impose Bayesian discounting SV formulation on the innovation variances: $\sigma_{jt} = \sigma_{jt-1}\delta/\upsilon_{jt}$, where $\delta \in (0,1]$ is a discount parameter and where $\upsilon_{jt} \sim \mathcal{B}(\delta\eta_{t-1}/2, (1-\delta)\eta_{t-1}/2)$ with $\eta_t = \delta\eta_{t-1} + 1$. We use this stochastic discounting model due to its computational convenience (with Kalman filtering equation) as explained in, for example, Chapter 10 of West and Harrison (1997) and Chapter 4 of Prado and West (2010).

Parsimonious covariance estimation is only one of the objectives of dynamic factor

[1] $\boldsymbol{V}$ is the implicit solution to $\boldsymbol{V} = \boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{\Phi}' + \mathbb{I}_K$, e.g. when $\boldsymbol{\Phi} = \widetilde{\phi}\mathbb{I}_K$, $\boldsymbol{V} = \frac{1}{1-\widetilde{\phi}^2}\mathbb{I}_K$.

modeling. The more traditional objective is disentangling the covariance structure and understanding its driving forces and how they change over time. *Sparse* modeling has been indispensable for both of these objectives, where fewer estimable coefficients yield far more stable covariance estimates and where nonzero patterns in $\boldsymbol{B}_t$ yield superior interpretable characterizations (Carvalho et al., 2008; Yoshida and West, 2010). Next, we explore the role of dynamic sparsity in DFA.

2.1 Dynamic Sparsity with Shrinkage Process Priors

No assumption has been as pervasive in the analysis of high-dimensional data as the one of sparsity. Sparsity is a practical modeling choice that facilitates high-dimensional inference and/or computation. In factor model contexts, it can also be used to anchor on identifiable parametrizations (Frühwirth-Schnatter and Lopes, 2009) and/or for estimating factor dimensionality (Bhattacharya and Dunson, 2011; Rockova and George, 2016). The potential of sparsity in dynamic factor models has begun to be recognized (Kaufmann and Beyeler, 2018; Kaufmann and Schumacher, 2017, 2019; Nakajima and West, 2013b).

In this work, we complement the factor model formulation (1) with dynamic sparsity priors on the factor loadings $\boldsymbol{B}_t$ for $1 \leq t \leq T$. In other words, rather than imposing a dense model by assigning a random walk (or a stationary autoregressive) prior on the loadings (such as Del Negro and Otrok, 2008; Stock and Watson, 2002), we allow for the possibility that the loadings are zero at certain times.

We will write $\boldsymbol{B}_t = (\beta_{jk}^t)_{j,k=1}^{P,K}$ and impose a shrinkage process prior on individual time series $\{\beta_{jk}^t\}_{t=1}^T$ for each (j, k). Several authors have reported on the benefits of dynamic variable selection in the analysis of macroeconomic data (Bitto and Frühwirth-Schnatter, 2019; Frühwirth-Schnatter and Wagner, 2010; Koop et al., 2010; Lopes et al., 2010; Nakajima and West, 2013b). We build on one of the more recent developments, the Dynamic Spike-and-Slab (DSS) priors proposed by Rockova et al. (2020).

DSS priors are dynamic extensions of spike-and-slab priors for variable selection (George and McCulloch, 1993; Rockova and George, 2018). Each coefficient in DSS is thought of as arising from two latent states: (1) an *inactive* state, where the coefficient meanders randomly around zero, and (2) an *active* state, where the coefficient walks on an autoregressive path. The switching between these two states is driven by a *dynamic* mixing weight which depends on past values of the series, making the states less erratic over time.

We begin by reviewing the conditional specification of the DSS prior. For each coefficient β_{jk}^t, we have a binary indicator $\gamma_{jk}^t \in \{0,1\}$, which encodes the state of β_{jk}^t (the "spike" inactive state for $\gamma_{jk}^t = 0$ and the "slab" active state for $\gamma_{jk}^t = 1$). Given γ_{jk}^t and a lagged value β_{jk}^{t-1}, we assume a conditional mixture prior (independently for each (j,k)):

$$\pi(\beta_{jk}^t|\gamma_{jk}^t, \beta_{jk}^{t-1}) = (1-\gamma_{jk}^t)\psi_0(\beta_{jk}^t|\lambda_0) + \gamma_{jk}^t\psi_1\left(\beta_{jk}^t \,|\, \mu(\beta_{jk}^{t-1}), \lambda_1\right), \tag{3}$$

where

$$\mu(\beta_{jk}^{t-1}) = \phi_0 + \phi_1(\beta_{jk}^{t-1} - \phi_0) \quad \text{with} \quad |\phi_1| < 1 \tag{4}$$

and

$$\mathbb{P}(\gamma_{jk}^t = 1|\beta_{jk}^{t-1}) = \theta_{jk}^t. \tag{5}$$

The conditional prior (3) is a mixture of two components: (i) a spike Laplace/Gaussian density $\psi_0(\beta|\lambda_0)$ that is concentrated around zero and (ii) a Gaussian slab density $\psi_1(\beta_t|\mu(\beta_{jk}^{t-1}), \lambda_1)$, which is moderately peaked around its mean $\mu(\beta_{jk}^{t-1})$ with variance λ_1. This mixture formulation is an extension of existing continuous spike-and-slab priors (George and McCulloch, 1993; Ishwaran et al., 2005; Rockova, 2018), allowing the mean $\mu(\beta_{jk}^{t-1})$ of the non-negligible coefficients to evolve smoothly over time (through a stationary autoregressive process of order 1).[2] The spike distribution $\psi_0(\beta_t|\lambda_0)$, on the other hand, *does not* depend on β_{jk}^{t-1}, effectively shrinking the negligible coefficients towards zero. In this regard, the conditional prior in (3) can be seen as a "multiple shrinkage" prior (George, 1986a,b) with two centers of gravity.

In time series data (as will be seen from our empirical study), it is reasonable to expect that some factors are active only for some periods of time. Such "pockets of predictability" (Farmer et al., 2018) can be captured with spike/slab memberships γ_{jk}^t that evolve somewhat smoothly. This behavior can be encouraged with dynamic mixing weights θ_{jk}^t (defined

[2]While our framework can be extended to higher order autoregressive processes, we outline our methodology for first order autoregression with $\phi_0 = 1$ due to its universal usage in practice ((Prado and West, 2010; West and Harrison, 1997))

in (5)) that reflect past information. To this end, we deploy the deterministic construction of Rockova et al. (2020) defined, for some global balancing parameter $0 < \Theta < 1$, as follows

$$\theta_{jk}^t \equiv \theta(\beta_{jk}^t) = \frac{\Theta \psi_1^{ST}\left(\beta_{jk}^t|\lambda_1, \phi_0, \phi_1\right)}{\Theta \psi_1^{ST}\left(\beta_{jk}^t|\lambda_1, \phi_0, \phi_1\right) + (1-\Theta)\psi_0\left(\beta_{jk}^t|\lambda_0\right)}, \tag{6}$$

given $(\Theta, \lambda_0, \lambda_1, \phi_0, \phi_1)$. This mixing weight has an interesting interpretation. It is defined as the *marginal* inclusion probability $\mathbb{P}(\gamma_{jk}^{t-1} = 1 \,|\, \beta_{jk}^{t-1})$ for classifying β_{jk}^{t-1} as arising from the *stationary* slab distribution $\psi_1^{ST}\left(\beta_{jk}^t|\lambda_1, \phi_0, \phi_1\right)$, as opposed to the stationary spike distribution $\psi_0\left(\beta_{jk}^t|\lambda_0\right)$, under the prior $\mathbb{P}(\gamma_{jk}^{t-1} = 1) = \Theta$. As θ_{jk}^t's evolve over time, they project the latent state (active/inactive) of the past value onto the next values. These weights induce marginal stability in the sense that each coefficient β_{jk} has a *marginal spike-and-slab distribution*, i.e. $\pi(\beta_{jk}) = \Theta \psi_1^{ST}\left(\beta_{jk}^t|\lambda_1, \phi_0, \phi_1\right) + (1-\Theta)\psi_0\left(\beta_{jk}^t|\lambda_0\right)$, which follows from the theorem by Rockova et al. (2020) given below:

Theorem 2.1. *Assume* $\{\beta_t\}_{t=1}^T \sim DSS(\Theta, \lambda_0, \lambda_1, \phi_0, \phi_1)$ *with* $|\phi_1| < 1$. *Then* $\{\beta_t\}_{t=1}^T$ *has a stationary distribution characterized by the following univariate marginal distributions:*

$$\pi^{ST}(\beta|\Theta, \lambda_0, \lambda_1, \phi_0, \phi_1) = \Theta\, \psi_1^{ST}(\beta \,|\, \lambda_1, \phi_0, \phi_1) + (1-\Theta)\psi_0\left(\beta \,|\, \lambda_0\right), \tag{7}$$

where $\psi_1^{ST}(\beta \,|\, \lambda_1, \phi_0, \phi_1)$ *is the stationary slab distribution.*

Having introduced the DSS priors, we can now fully specify our dynamic latent factor model with (1)-(5). It is possible to extend our model to non-stationary random walk slab process, (obtained with $\phi_1 = 1$) by allowing transition weights θ_{jk}^t to be random, equal to some deterministic sequence (e.g. as in Nakajima and West (2013a)) or to a fixed value $\theta_{jk}^t = \widetilde{\theta}$ for $1 \leq t \leq T$. When treated as random, the weights may be prone to transitioning too often between the spike/slab states creating instabilities over time.

Our sparse dynamic factor model is related to the approach of Nakajima and West (2013b), who zero out loadings whenever their autoregressive path drops below a certain threshold (see Rockova et al., 2020, for comparisons). Another related approach is by Kaufmann and Beyeler (2018), who induce a point-mass spike and slab prior on the loadings. However, their approach (a) does not link the inclusion indicators and loadings over time, and (b) MCMC is deployed for calculations. Here, we develop both MCMC and an EM estimation procedure which does not rely on strict identifiability constraints.

2.2 Identifiability Considerations

Factor models are not free from identifiability problems owing to the fact that the model (1) and (2) is observationally equivalent to $\boldsymbol{Y}_t = \boldsymbol{B}_t^{\star}\boldsymbol{\omega}_t^{\star} + \boldsymbol{\epsilon}_t$ and $\boldsymbol{\omega}_t^{\star} = \boldsymbol{\Phi}\boldsymbol{\omega}_{t-1}^{\star} + \boldsymbol{e}_t$, where $\boldsymbol{\omega}_t^{\star} = \boldsymbol{A}_t\boldsymbol{\omega}_t$ and $\boldsymbol{B}_t^{\star} = \boldsymbol{B}_t\boldsymbol{A}_t'$ for any orthonormal matrix $\boldsymbol{A}_t$. Identification restrictions are particularly important for Bayesian analysis with MCMC, where meaningful interpretation of $\boldsymbol{B}_t$ could be prevented by averaging over various model orientations in the Markov Chain. To ensure identifiability, it is customary to restrict $\boldsymbol{B}_t$ to be lower-triangular, with ones on the diagonal (Aguilar and West, 2000; Lopes and Carvalho, 2007; Lopes and West, 2004; Nakajima and West, 2013b; Zhou et al., 2014) or some variant of this form (Frühwirth-Schnatter and Lopes, 2009). Identifiability in *sparse* factor models is even more delicate (Frühwirth-Schnatter and Lopes, 2009). Nevertheless, these constraints render the analysis dependent on the ordering of the responses. Even without strict identifiability constraints, one needs to verify ex-post that the estimated sparse loadings satisfy identifiability constraints (as discussed e.g. by Frühwirth-Schnatter and Lopes (2009)). Bayesian ex-post MCMC strategies have been proposed that *do not* deploy identifiability constraints during the estimation stage (Kastner et al., 2017; Kaufmann and Schumacher, 2019). Instead, posterior draws coming from potentially very different orientations (identification schemes) are rotated ex-post.

For implementing our EM algorithm, we also do not impose any strict identifiability constraints on our model. Instead, we induce soft identifiability through sparsity priors and we let the EM optimization strategy converge towards one sparse posterior mode. Unlike with MCMC (an averaging strategy mixing over various sparse orientations), the EM output is conditional on one particular orientation and can be interpreted as such. To accelerate convergence and improve the chances of reaching better local modes, we use parameter expansion with automatic rotations to sparsity, as implemented by Rockova and George (2016). Unlike the ex-post rotations deployed in Frühwirth-Schnatter and Lopes (2009), our rotations are performed *inside* the algorithm to gear the EM trajectory towards promising modes. This corresponds to a variant of the PX-EM algorithm of (Liu et al. (1998a) and the one-step late PX-EM of Van Dyk and Tang (2003) for Bayesian factor analysis, where the augmented data log likelihood is maximized as a function of the augmented parameter within each EM iteration. This is in contrast to conditional data augmentation of Meng and Van Dyk (1998), where one seeks an optimal value of the augmented parameter before starting the EM algorithm. Similar data augmentation strategies can also be used to speed-up MCMC covergence, as demonstrated by the conditional and marginal data augmentation approaches of Meng and Van Dyk (1999). In the context of Bayesian factor analysis, Ghosh and Dunson (2009) proposed a prior specification through parameter expansion that facilitates posterior computation. Yu and Meng (2011) proposed an ancillarity-sufficiency interweaving strategy for speeding-up MCMC convergence. This strategy was applied in the context of factor models in Kastner et al. (2017). For our MCMC implementation, we will impose the usual constraints on loading matrices with a block lower triangular structure and with diagonal elements strictly positive (Aguilar and West, 2000; Geweke and Zhou, 1996; Lopes and Carvalho, 2007; Lopes and Migon, 2002; Nakajima and West, 2013b; Zhou et al., 2014).

2.3 Estimating Factor Dimensionality

The factor model (1) and (2) is formulated conditionally on the number of factors $K \in \mathbb{N}$. As noted by Bai and Ng (2002), "the correct specification of the number of factors is central to both the theoretical and empirical validity of factor models." The authors propose a criterion and show that it is consistent for estimating K in high-dimensional setups. In another strand of research, sparsity has been exploited for determining the effective factor dimensionality (Frühwirth-Schnatter and Lopes, 2009). In particular, Bayesian non-parametric formulations have been proposed (Bhattacharya and Dunson, 2011; Rockova and George, 2016), where K is extended to infinity, while making sure that the number of *nonzero* columns in $\boldsymbol{B}_t$ is finite with probability one. Treating the number of *active* factors as unknown and bounded by K in this way, the posterior output under our spike-and-slab priors can be used to determine the effective dimensionality. We adopt a similar approach to Rockova and George (2016), where K in (1) is purposefully over-estimated and the number of *nonzero* columns obtained under strict sparsity priors will indicate how many *effective* factors there are.

3 Estimation Strategy

To estimate the proposed dynamic latent factor model with DSS priors, we develop two computational methods: an EM algorithm (Dempster et al., 1977), which allows for fast identification of posterior modes by iteratively maximizing the conditional expectation of the log posterior, and a standard MCMC implementation that is comparatively slower and thereby less appealing for large data applications. We describe both approaches in the following subsections.

3.1 EM Algorithm

The EM algorithm is well-suited for latent variable models, such as factor analysis, where it has been deployed by multiple authors including Rubin and Thayer (1982); Watson and Engle (1983); Zuur et al. (2003) and, more recently, Rockova and George (2016). EM can be motivated by two simple facts. First, if we knew the missing data, standard estimation techniques can be deployed to estimate model parameters. Second, once we update our beliefs about model parameters we can make a much better educated guess about the missing data. Iterating between these two steps provides a fast way of obtaining maximum likelihood estimates and posterior modes.

Our EM algorithm has a few extra features that make it particularly attractive for dynamic factor analysis. First, the DSS priors (with a Laplace spike at zero) create spiky posteriors with sparse modes at coordinate axes. These modes yield interpretable latent factor structures that are anchored on sparse representations without arbitrary identifiability constraints. Second, the number of *active* factors does not have to be pre-specified and can be inferred from the dynamically evolving sparse structure.

As we discussed in Section 2.2, the model is invariant under rotation of factor loading matrices. While this lack of identifiability has been regarded as a setback, it can also be regarded as an opportunity. Rotational invariance creates ridge-lines in the posterior

Algorithm: *EM algorithm for Automatic Rotations to Sparsity*		
	Initialize $\boldsymbol{\Delta} = (\boldsymbol{B}_{0:T}, \boldsymbol{\Sigma}_{1:T})$	
	Repeat the following E-Step, M-Step and Rotation step until convergence	
The E-Step		
		For $t = 1, \dots, T$
E1:	Latent Features:	Get $\boldsymbol{\omega}_{t\,\vert\,T}, \boldsymbol{V}_{t\,\vert\,T}$ and $\boldsymbol{V}_{t,t-1\,\vert\,T}$ from the Kalman filter and smoother
E2:	Latent Indicators	Compute $\langle \gamma^t_{jk} \rangle$ for $j = 1, \dots, P$, $k = 1, \dots, K$,
		$\langle \gamma^0_{jk} \rangle = \frac{\Theta\psi_1(\beta^0_{jk}\vert 0, \frac{\lambda_1}{1-\phi^2})}{\Theta\psi_1(\beta^0_{jk}\vert 0, \frac{\lambda_1}{1-\phi^2}) + (1-\Theta)\psi_0(\beta^0_{jk}\vert 0, \lambda_0)}$
		$\langle \gamma^t_{jk} \rangle = \frac{\theta^t_{jk}\psi_1(\beta^t_{jk}\vert \phi\beta^{t-1}_{jk}, \lambda_1)}{\theta^t_{jk}\psi_1(\beta^t_{jk}\vert \phi\beta^{t-1}_{jk}, \lambda_1) + (1-\theta^t_{jk})\psi_0(\beta^t_{jk}\vert 0, \lambda_0)}$ (8)
The M-Step		
M1:	Loadings	For $t = 0, \dots, T$
		Update $\beta^{t\star}_{jk}$, for $j = 1, \dots, P$, $k = 1, \dots, K$ following equation (13) in the Appendix A.
M2:	Rotation Matrix	Set $\boldsymbol{A}_0 = \boldsymbol{I}_K$
		For $t = 1, \dots, T$
		Update $\boldsymbol{A}_t = \boldsymbol{M}_{1t} - \boldsymbol{M}_{12t} - \boldsymbol{M}'_{12t} + \boldsymbol{M}_{2t}$, where
		$\boldsymbol{M}_{1t} = \phi^2 \left[\boldsymbol{\omega}_{t-1\,\vert\,T}\boldsymbol{\omega}'_{t-1\,\vert\,T} + \boldsymbol{V}_{t-1\,\vert\,T}\right]$
		$\boldsymbol{M}_{12t} = \phi \left[\boldsymbol{\omega}_{t-1\,\vert\,T}\boldsymbol{\omega}'_{t\,\vert\,T} + \boldsymbol{V}_{t,t-1\,\vert\,T}\right]$
		$\boldsymbol{M}_{2t} = \boldsymbol{\omega}_{t\,\vert\,T}\boldsymbol{\omega}'_{t\,\vert\,T} + \boldsymbol{V}_{t\,\vert\,T}$
M3:	Idiosyncratic Variance	Compute $\boldsymbol{\Sigma}_{1:T}$ using Forward Filtering Backward Smoothing
The Rotation Step		
R:	Rotation	For $t = 0, \dots, T$
		Get Cholesky decomposition $\boldsymbol{A}_t = \boldsymbol{A}_{tL}\boldsymbol{A}'_{tL}$
		Rotate $\boldsymbol{B}_t = \boldsymbol{B}^\star_t \boldsymbol{A}_{tL}$

Table 1: Parameter Expanded EM algorithm for sparse Bayesian dynamic factor analysis

that connect posterior modes and that can guide optimization trajectories (Rockova and George, 2016). We follow the parameter expansion approach (see also Liu et al., 1998b; Liu and Wu, 1999) that intentionally over-parametrizes the model and takes advantage of the lack of identification to speed up convergence. Similarly as Rockova and George (2016), we work with the expanded model discussed in section 2.2. We assume the initial condition $\boldsymbol{\omega}_0 \sim \mathcal{N}_K(\mathbf{0}, 1/(1-\widetilde{\phi}^2)\boldsymbol{I}_K)$ which is the stationary distribution of the latent factors when $\boldsymbol{\Phi} = \widetilde{\phi}\mathbb{I}_K$ for some $0 < \widetilde{\phi} < 1$. We impose the DSS prior on the individual entries of the *rotated* matrix $\boldsymbol{B}^\star_t = \boldsymbol{B}_t\boldsymbol{A}^{-1}_{tL}$. The idea is to rotate towards sparse orientations throughout the iterations of the EM algorithm. The key observation is as follows: while matrices $\boldsymbol{A}_t$ for $1 \leq t \leq T$ *cannot* be identified from the observed data $\boldsymbol{Y}$, they can be identified from the complete data. Denoting $\boldsymbol{\Gamma}_t = \left\{\gamma^t_{jk}\right\}_{j,k}$, we treat both $\boldsymbol{\Omega} = [\boldsymbol{\omega}_0, \dots, \boldsymbol{\omega}_T]$ and $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_0, \dots, \boldsymbol{\Gamma}_T\}$ as missing data. The reduced model is obtained by setting $\boldsymbol{A}_t = \boldsymbol{I}_K$ for all $1 \leq t \leq T$.

All the unknown components of our model can be divided into three categories: missing data ($\boldsymbol{\Omega}$, $\boldsymbol{\Gamma}$), estimated parameters ($\boldsymbol{B}_0$, $\boldsymbol{B}_{1:T}$, $\boldsymbol{\Sigma}_{1:T}$) and parameters whose values are pre-specified ($\widetilde{\phi}$, ϕ_1, λ_0, λ_1, $\mathbf{\Theta}$ and δ). Among the pre-specified parameters we set the AR coefficients $\widetilde{\phi}$ and ϕ_1 close to (i.e. slightly smaller than) 1. This is because we want the AR processes to be stationary but, at the same time, we do not want the current values to deviate too far away from the past. We recommend setting values in the range $[0.9, 1]$ for the AR parameters. For similar reasons, the discount factor δ is also set close to unity (0.95 to be precise). However, instead of being treated as fixed values, the AR coefficients ϕ_0 and ϕ_1 *can* be easily estimated by assigning suitable priors, as demonstrated in the MCMC implementation discussed in Section 3.2. Following the recommendation of (Rockova et al., 2020), we set a moderate penalty for the spike distribution $\lambda_0 = 0.9$ and a comparatively large slab variance $\lambda_1 = 10(1-\phi_1^2)$. This is to ensure that the penalty on the factor loadings is unimodal. The marginal importance weight $\mathbf{\Theta} = 0.9$ is chosen to be large because smaller values provide an overwhelming support towards zero factor loadings.

Let us denote by $\mathbf{\Delta} = (\boldsymbol{B}_0, \boldsymbol{B}_{1:T}, \mathbf{\Sigma}_{1:T})$ the model parameters. The matrix $\boldsymbol{B}_0$ contains the initial conditions that are assumed to arise from the stationary spike-and-slab prior distribution and $\boldsymbol{B}_{1:T}$ denotes all matrices $\boldsymbol{B}_t$ for $1 \le t \le T$. The goal of the EM algorithm is to find parameter values $\widehat{\mathbf{\Delta}}$, which are most likely (*a posteriori*) to have generated the data, i.e. $\widehat{\mathbf{\Delta}} = \arg\max_{\mathbf{\Delta}} \log \pi(\mathbf{\Delta} \mid \boldsymbol{Y})$. This is achieved indirectly by iteratively maximizing the expectation of the augmented log-posterior, treating the hidden factors $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ as missing data. Starting with an initialization $\mathbf{\Delta}^{(0)}$, the $(m+1)^{st}$ step of the EM algorithm outputs $\mathbf{\Delta}^{(m+1)} = \arg\max_{\mathbf{\Delta}} Q(\mathbf{\Delta} \mid \mathbf{\Delta}^{(m)})$, where $Q(\mathbf{\Delta} \mid \mathbf{\Delta}^{(m)}) = \mathbb{E}_{\mathbf{\Gamma},\mathbf{\Omega}|\boldsymbol{Y},\mathbf{\Delta}^{(m)}}[\log \pi(\mathbf{\Delta}, \mathbf{\Gamma}, \mathbf{\Omega} \mid \boldsymbol{Y})]$ with $\mathbb{E}_{\mathbf{\Gamma},\mathbf{\Omega}|\boldsymbol{Y},\mathbf{\Delta}^{(m)}}(.)$ denoting the conditional expectation given the observed data and current parameter estimates at the m^{th} iteration. The EM algorithm iterates between the E-step (obtaining the conditional expectation of the log-posterior) and the M-step (obtaining $\mathbf{\Delta}^{(m+1)}$). The *parameter-expanded EM* works in a slightly different manner.

The E-step of the *parameter-expanded version* operates in the reduced space (keeping $\boldsymbol{A}_t = \boldsymbol{I}_K$), while the M-step operates in the expanded space (allowing for general $\boldsymbol{A}_t$). Namely, the E-step computes the expectation $Q(\mathbf{\Delta} \mid \mathbf{\Delta}^{(m)})$ with respect to the conditional distribution of $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ under the *original model* anchoring on $\boldsymbol{B}_t$ and $\boldsymbol{A}_t = \boldsymbol{I}_K$, rather than on $\boldsymbol{B}_t^{\star}$ and unrestricted $\boldsymbol{A}_t$. The M-step, on the other hand, is performed in the *expanded* parameter space, where optimization takes place over $\boldsymbol{B}^{\star}_{0:T}$, $\boldsymbol{\Sigma}_{1:T}$, and $\boldsymbol{A}_{1:T}$. Updating $\boldsymbol{B}^{\star(m+1)}_{0:T}$ boils down to solving a series of independent penalized dynamic regressions. The idiosyncratic variances $\boldsymbol{\Sigma}_t = \mathrm{diag}\{\sigma^2_{1t}, \ldots, \sigma^2_{Pt}\}$ for $t = 1, \ldots, T$ are estimated in the M-step using Forward Filtering Backward Smoothing (Table 5 in the Appendix) (Ch. 4.3.7 Prado and West, 2010) using the discount SV specification (as discussed in the Supplemental Material). Since $\boldsymbol{A}_{1:T}$ can be inferred from the complete data, one can estimate these matrices in the M-step to leverage the information in the missing data. Nevertheless, the updated matrices $\boldsymbol{A}_{1:T}$ are *not* carried forward towards the next E-step (which uses $\boldsymbol{A}_t = \boldsymbol{I}_K$), but are used to *rotate* the solution $\boldsymbol{B}^{\star(m+1)}_{0:T}$ back towards the reduced space via $\boldsymbol{B}^{(m+1)}_t = \boldsymbol{B}^{\star(m+1)}_t \boldsymbol{A}_{tL}$. The steps of the algorithm are carefully explained in Section A.2. The computations are summarized in Table 1. The convergence of the EM algorithm with parameter expansion is provably faster (Liu et al., 1998b; Rockova and George, 2016).

3.2 MCMC

This section describes an MCMC algorithm for our dynamic factor model with dynamic spike-and-slab priors. The *DSS* prior specification here is slightly different from the setup considered for the EM algorithm. The Laplace spike distribution $\psi_0(\beta|\lambda_0) = \lambda_0/2\mathrm{e}^{-\lambda_0|\beta|}$ yields sparse posterior modes, a favorable feature for the EM algorithm. However, MCMC ultimately reports the posterior mean which is non-sparse even under the Laplace prior. We will therefore assume a Gaussian spike (instead of Laplace) to utilize its direct conditional conjugacy for posterior updating. In particular, we assume the following spike density for $\lambda_0 << \lambda_1$

$$\psi_0(\beta \mid \lambda_0) = \exp\{-\beta^2/(2\lambda_0)\}/\sqrt{2\pi\lambda_0}. \tag{9}$$

This yields the following conditional Gaussian distribution for individual factor loadings β^t_{jk}

$$\beta^t_{jk} \mid \gamma^t_{jk}, \beta^{t-1}_{jk} \sim \mathcal{N}\left(\gamma^t_{jk}\mu^t_{jk}\,,\, \gamma^t_{jk}\lambda_1 + (1-\gamma^t_{jk})\lambda_0\right)$$

and transition weights θ^t_{jk} in (6) with the Gaussian stationary spike distribution $\psi^{ST}_0(\beta^{t-1}_{jk}|\lambda_0) = \psi_0(\beta \mid \lambda_0)$. An extension to the Laplace spike is possible with an additional augmentation step, casting the Laplace distribution as a scale mixture of Gaussians with an exponential

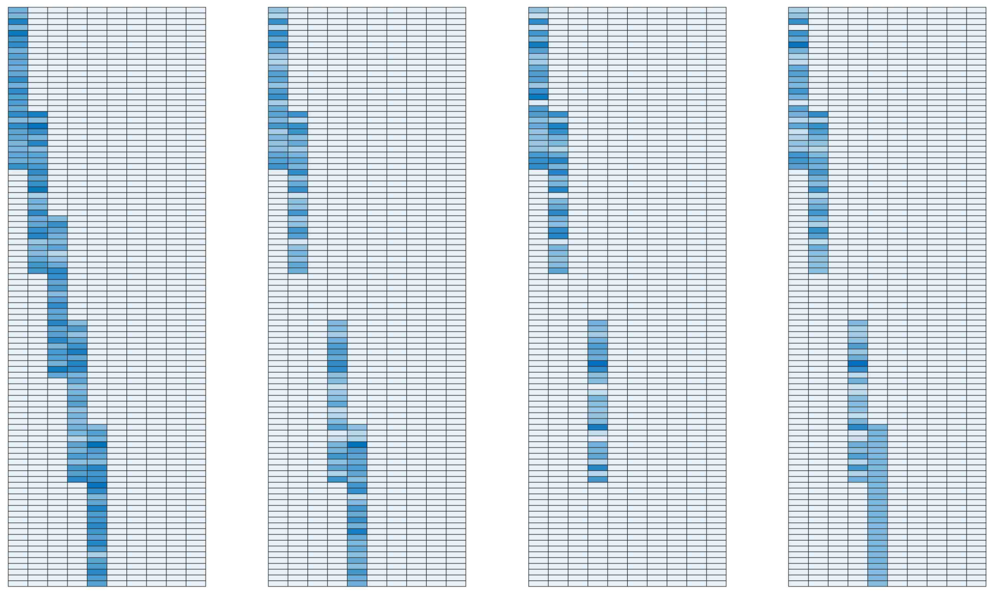


Figure 1: Simulation Study: The true latent factor loadings $\boldsymbol{B}_t^0$ at $t = 1, 101, 201, 301$.

mixing distribution (Park and Casella, 2008). The MCMC algorithm has a Gibbs structure, sampling iteratively from the conditional posteriors of the regression coefficients $\boldsymbol{B}_{0:T}$, latent indicators $\boldsymbol{\Gamma}_{\mathbf{0:T}}$ and variances $\boldsymbol{\Sigma}_{0:T}$ (Frühwirth-Schnatter 1994; Prado and West 2010, Sect 4.5).

For the stationary DSS prior, we assume that the autoregressive parameter $|\phi_1| < 1$ is assigned the following beta prior (as in (Rockova et al., 2020))

$$\pi(\phi_1) \propto \left(\frac{1+\phi_1}{2}\right)^{a0-1} \left(\frac{1-\phi_1}{2}\right)^{b0-1} \mathbb{I}(|\phi_1| < 1) \quad \text{with } a_0 = 20 \text{ and } b_0 = 1.5, \tag{10}$$

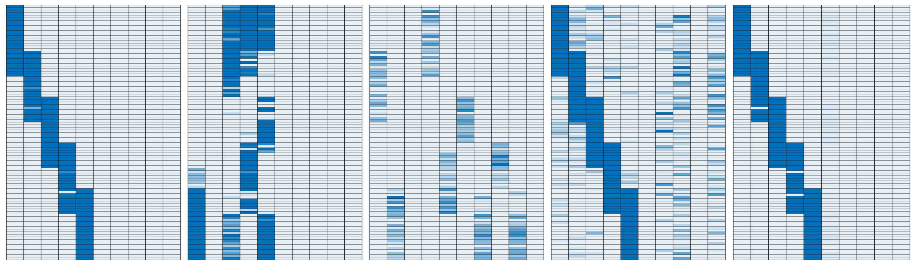
implying a prior mean of $2a_0/(a_0 + b_0) - 1 = 0.86$. We will update ϕ_1 with a Metropolis step, using a uniform proposal density on the interval $[0.8, 1]$. While we assume $\phi_0 = 0$ throughout, one can update ϕ_0 in a similar fashion. Table B1 in the appendix gives a step-by-step summary of the MCMC algorithm. A small demonstration on a simulated dataset is given in Appendix B.

4 Simulation Study

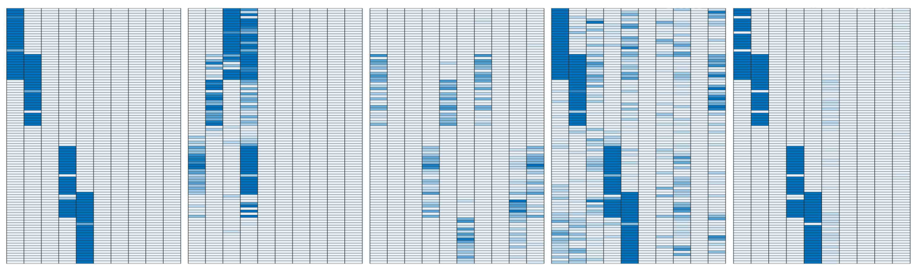
We illustrate the usefulness of our proposed approach, relative to multiple existing methods, on synthetic data reflecting the following characteristics that can occur in real applications: dynamic patterns of sparsity, smoothness, and a time-varying factor dimension.

First, we generate a single dataset with $P = 100$ responses, $K = 10$ candidate latent factors, and $T = 400$ time series observations (extra 100 data points are generated as training data for the rolling window analysis, as will be described below). The dimensionality of this example is already beyond practical limits of many Bayesian procedures. The elements of latent factors $\boldsymbol{\Omega}_t$ and idiosyncratic errors $\boldsymbol{\epsilon}_t$ are generated from a standard Gaussian distribution. Only the first five factors are potentially active over time, with the latter five being always inactive. We now describe the true loading matrices $\boldsymbol{B}^0 = [\boldsymbol{B}_1^0, \ldots, \boldsymbol{B}_T^0]$, which were used to generate the data, where $\boldsymbol{B}_t^0 = \{\beta_{jk}^{0t}\} \in \mathbb{R}^{P\times K}$. At time $t = 1$, the active latent factor loadings form a block diagonal structure with 28 active loadings per factor, of which 10 overlap with another factor. In other words, we have 60 series with only one active factor, and 40 with two active factors (see the leftmost image in Figure 1). The sparsity pattern changes structurally over time where (a) at time $t = 101$ the loadings of the third factor become inactive, (b) at $t = 201$ the loadings of the fifth factor become inactive, and (c) at $t = 301$ the loadings of the fifth factor are re-introduced and active until $T = 400$ (Figure 1). The true nonzero loadings are smooth and arrive from an autoregressive process, i.e. $\beta_{jk}^{0t} = \phi\beta_{jk}^{0t-1} + v_{jk}^t$ with $v_{jk}^t \overset{iid}{\sim} \mathcal{N}(0, 0.0025)$ for $\phi = 0.99$, initiated at $\beta_{jk}^{01} = 2$ for all $1 \le j \le P$ and $1 \le k \le 5$. When loadings β_{jk}^{0t} become inactive, they are thresholded to zero. The true factor loadings are thereby smooth until they suddenly drop out and can emerge.

We compare our proposed dynamic spike-and-slab factor selection with three other approaches. The first one is the "rolling window" version of the static factor analysis with rotations to sparsity by Rockova and George (2016) using $K = 10$ (i.e. overshooting the true factor dimensionality). We compare this approach with "Adaptive PCA" of Bai and Ng (2002), which corresponds to a rolling-window principal component analysis (PCA) with estimated number of factors, and with "Sparse PCA" using $K = 10$, which is a

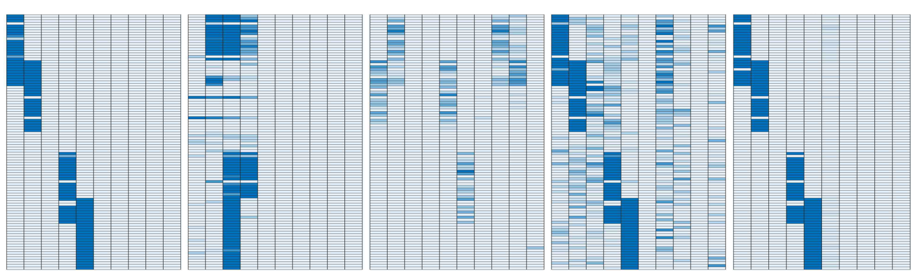


(a) $t = 100$



(b) $t = 200$

(c) $t = 300$



(d) $t = 400$

Figure 2: Simulated Example: Heatmaps of true and estimated factor loadings at $t = \{100, 200, 300, 400\}$. Comparisons are made between (from left to right), the true factor loadings, "Adaptive PCA," "Sparse PCA" ($K = 10$), rolling window spike-and-slab factor analysis ($K = 10$), and our dynamic spike-and-slab factor analysis. The first three methods are estimated using a rolling window of 100 data points. Factor loadings are absolute and capped at 0.5 for visibility.

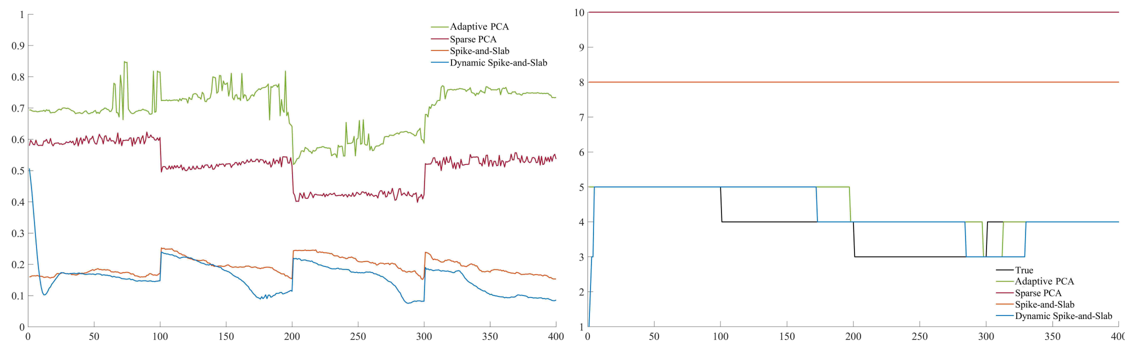
rolling-window LASSO-based regularization method with cross-validation for selecting the level of shrinkage (Witten et al., 2009). All these methods are estimated using a rolling window of size 100, where we generate extra 100 training data points using the sparsity pattern $\boldsymbol{B}_1^0$. We choose $\boldsymbol{\Phi} = \widetilde{\phi}\mathbb{I}_K$ with $\widetilde{\phi} = 0.95$ and $K = 10$. Choosing $\widetilde{\phi}$ close to 1 ensures that the latent factor processes are stationary and their means do not deviate too far away from past values. To deploy the dynamic spike-and-slab priors, we set $\phi_0 = 0$, $\phi_1 = 0.98$, $\lambda_0 = 0.9$, $\lambda_1 = 10(1 - \phi_1^2)$, and $\Theta = 0.9$. To improve the performance of our EM method, we initialize the procedure using the output from the rolling window static spike-and-slab factor model of Rockova and George (2016).

Focusing on the reconstruction of factor loadings, we take snapshots at times $t = \{100, 200, 300, 400\}$ and visually compare the output to the truth (Figure 2). We see that both spike-and-slab methods achieve good recovery. However, the static spike-and-slab cannot fully contain the dynamic loadings, where we see a lot of spillover to other factors. Dynamic spike-and-slab shrinkage, on the other hand, smooths out the sparsity over time, clearly improving on the recovery. "Adaptive PCA" performs well, correctly specifying the number of factors. However, the factor loadings are non-sparse and rotated. "Sparse PCA" with $K = 10$ is fairly successful, recovering the blocking structure correctly, but splitting the signal among multiple factors (an observation made also by Rockova and George, 2016). For the spike-and-slab methods, these patterns can be alternatively obtained by thresholding conditional inclusion probabilities rather than just looking at nonzero entries in $\widehat{\boldsymbol{B}}_{1:T}$.

We further explore how the root mean squared errors (RMSE) change over time for one of the simulations (Figure 3). This is calculated for each $t = 1 : T$ by

$$RMSE(\widehat{\boldsymbol{B}}_t) = \sqrt{\frac{tr(\boldsymbol{B}_t^0 - \widehat{\boldsymbol{B}}_t)'(\boldsymbol{B}_t^0 - \widehat{\boldsymbol{B}}_t)}{P \times K}}, \tag{11}$$

where $\widehat{\boldsymbol{B}}_t$ are the estimated factor loadings at time t. Since this comparison is not entirely meaningful due to the rotational invariance, we compute (11) for the left-ordered variants of these matrices. By looking at the speed of decrease in RMSE after a structural change, it is clear that dynamic spike-and-slab adapts faster compared to its rolling window counterpart. The drop of RMSE for "Adaptive PCA" in periods 101:200 and 201:300 can be attributed



(a) RMSE (b) $\widehat{K}$

Figure 3: Simulation Study: (Left) The root mean squared error (11) and (Right) the estimated number of factors for "Adaptive PCA," "Sparse PCA," static spike-and-slab, and dynamic spike-and-slab, calculated for each $t = 1{:}400$.

to the fact that the number of factors was estimated correctly, resulting in many true zero discoveries. On the other hand, the large estimation error of "Sparse PCA" is due to the lack of sparsity and scattered structure of the factors.

Additionally, we plot the estimated number of factors for each method and compare it to the true number of factors. "Sparse PCA" overestimates the number of factors (where we regard a factor as active if it has at least one nonzero loading). This indicates that unstructured sparsity is not enough. Looking at "Adaptive PCA" and our dynamic spike-and-slab factor model, we find that both perform similarly well in terms of estimating the number of factors. Furthermore, we note that dynamic spike-and-slab adapts faster to factors disappearing, while "Adaptive PCA" adapts faster to factors reappearing.

We repeat the experiment 10 times and report the average RMSE over each of the four stationary interim time periods in Table 2. Dynamic spike-and-slab achieves good recovery, improving upon the rolling window spike-and-slab by as much as 8% to 34% (except for the first period). Large recovery errors of the "Sparse PCA" method can be explained by factor splitting. While "Adaptive PCA" does recover the correct number of factors at each snapshot, the loadings are non-sparse, rotated and non-smooth over time.

	t=1:100			t=101:200			t=201:300			t=301:400		
	RMSE	%	$\widehat{K}$	RMSE	%	$\widehat{K}$	RMSE	%	$\widehat{K}$	RMSE	%	$\widehat{K}$
Adaptive PCA	1.0660	-266.07	5	1.0590	-400.24	4.97	0.9730	-250.38	3.97	1.033	-430.01	3.88
Sparse PCA	0.7862	-169.99	10	0.7260	-242.94	10	0.6377	-129.64	10	0.7383	-278.81	10
Spike-and-Slab	0.1919	34.10	8	0.2843	-34.29	8	0.2988	-7.60	8	0.2447	-25.60	8
Dynamic Spike-and-Slab	0.2912	-	4.89	0.2117	-	4.72	0.2777	-	3.84	0.1949	-	3.71

Table 2: Simulation Study: Performance evaluation of the latent factor methods compared to the true coefficients for $t = 1{:}400$. Performance is evaluated based on RMSE within each evaluation period. % is the performance gain compared to dynamic spike-and-slab. $\widehat{K}$ is the average number of factors estimated during that period.

5 Empirical Study

The empirical application concerns a large-scale monthly U.S. macroeconomic database, (colloquially known as the FRED-MD dataset (McCracken and Ng, 2016) in the Macroeconomics literature) comprising a balanced panel of $P = 127$ monthly macroeconomic and financial variables tracked over the period of 2001/01 to 2015/12 ($T = 180$). These variables are classified into eight main categories, depending on their economic meaning: *Output and Income*, *Labor Market*, *Consumption and Orders*, *Orders and Inventories*, *Money and Credit*, *Interest Rate and Exchange Rates*, *Prices*, and *Stock Market*. A detailed description of how variables were collected and constructed is provided in McCracken and Ng (2016). A quick table of names and groups of each variable is in the Appendix (Table B3). The variables were centered to have mean zero and standardized following the procedures in McCracken and Ng (2016).

This data, and its various subsets, have been widely studied in the literature, either as a standalone dataset (for macroeconomic forecasting and an impulse/response analysis) or as an essential part of broader data contexts. We review these analyses briefly below. For example, Stock and Watson (2018) deployed this dataset for estimation of dynamic causal effects in Macroeconomics. In other analyses, Miranda-Agrippino and Ricco (2018) extracted a set of lagged macro-financial dynamic factors to project monetary policy shocks and Gargano et al. (2019) computed the Ludvigson-Ng (LN) macro factors for predicting bond values. Using a quarterly aggregated version of this data, Huber and Feldkircher

(2019) fitted a Bayesian vector autoregressive model to forecast a subset of 21 variables. A larger forecasting exercise was conducted by Koop et al. (2019), who used 129 variables spanning over years 1960 to 2014 to predict GDP growth, inflation and short-term interest rates. A subset of this data, in conjunction with additional economic variables has been analyzed in Daniele and Schnaitmann (2019) who study the effects of a monetary policy shock through a regularized factor-augmented vector autoregressive (FAVAR) model. While the central theme of these works has been forecasting and/or impulse response analysis, the primary focus of our analysis in this section is discovering latent interpretable structures and glean insights into the interconnectivity between different sectors of the US macroeconomy, with a particular focus on the 2008 financial crisis. Forecasting will be discussed later in Section 6.

Stock and Watson (2005) analyzed a similar macroeconomic dataset (often referred to as the "Stock and Watson" dataset in Econometrics literature), containing 132 series over the sample 1959:1 to 2003:12. After performing variance decompositions, they found six factors that explain most of the variation in the data. With the same dataset, the IC1 and IC2 criteria developed in Bai and Ng (2002) find seven static factors explaining over 40 percent of the variation in the data. Bai and Ng (2013) used the same data extended to 2007:12 and showed first 7 factors still explain 45 percent of the variation in the data, though the IC2 criterion found the optimal number of factors to be 8.

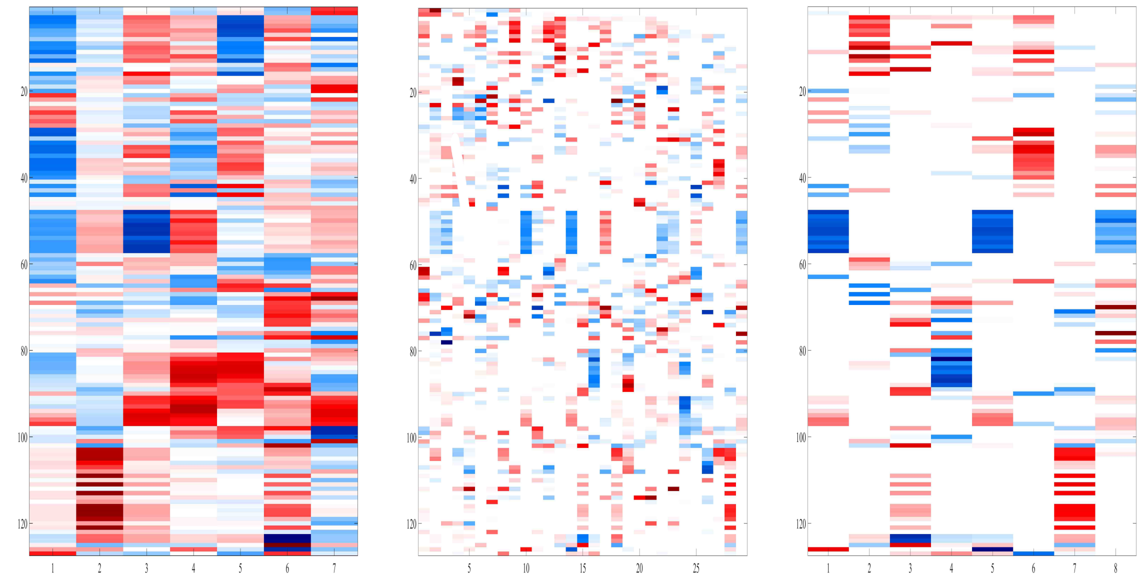
The purpose of conducting a sparse latent factor analysis on a large-scale economic dataset, such as this one, is at least twofold. Due to the group structure of the data, it is natural to assume that the measured indicators are tied via a few latent factors, the basic premise of latent factor modeling. Moreover, we expect the sparse latent structure to detect clusters of dependence structures that capture the interconnectivity of indicators spanning many *different* aspects of the economy. Sparsity will help extract such interpretable structures. Second, given the dynamic nature of the economy, there is a substantial interest in understanding how these dependencies change over time and– in particular– how they are affected by shocks. We anticipate non-negligible shifts in the economy, as the data spans over the housing bubble deflation after 2006 and the great financial crisis in late

2008, which led to the Great Recession. Understanding the interplay between contributing factors to the financial crisis has been a subject of rigorous research (see for example, Benmelech et al., 2017; Chodorow-Reich, 2014; Commission, 2011; Mian et al., 2013; Mian and Sufi, 2009, 2011; Reinhart and Rogoff, 2008). Our analysis is purely data-driven and thereby descriptive rather than causally conclusive. We attempt to characterize patterns of shock proliferation and permanence of structural changes of the economy using our dynamic factor model.

As the dataset is considerably richer than our simulated example, we expand the model (1) by incorporating a dynamic intercept to capture location shifts that could not be easily standardized away. The intercepts c_{jt} follow independent random walk evolutions with an initial condition $c_0 \sim N(0,1)$. The initial condition for the SV variances is $1/\sigma_{j0}^2 \overset{ind}{\sim} G(n_0/2, d_0/2)$ for $1 \leq j \leq P$ with $n_0 = 20$ and $d_0 = 0.002$. The discount factor is set to 0.95.

First, we examine one snapshot of the output from "Adaptive PCA" and "Sparse PCA" (described in Section 4) at time 2015/12 (Figures 4). Both methods do pick up certain groupings, but do not yield interpretable enough representations. This is likely due to overestimation of the number of factors (Figure 4 (b)), factor rotation and lack of sparsity (Figure 4 (a)) and/or factor splitting (Figure 4 (c)). Next, we deploy the rolling window spike-and-slab factor method with a training period of 10 years to obtain starting values for our dynamic factor model. Priors and their hyper-parameters were chosen as in the simulation study. We choose a generous upper bound $K = 126$ on the number of factors, letting the sparsity rule out factors that are irrelevant.

We now examine the output of our procedure at three time points: 2003/12, 2008/10, and 2015/12. These three snapshots are of particular interest as they represent three distinct states of the economy: relative stability (2003), sharp economic crisis (2008), and recovery (2015). 2008/10 is at the onset of the great financial crisis, where deflation of the housing bubble after 2006 lead to mortgage delinquencies and financial fragility (Commission, 2011). This distress permeated throughout the rest of the economy, including the labor market, leading to the deepest recession in the post-war history.



(a) “Adaptive PCA” (b) ”Sparse PCA” (c) ”Sparse PCA”

Figure 4: Macroeconomic Study: Estimated factor loadings using “Adaptive PCA” (Left), “Sparse PCA” with number of factors set as 30 (Middle), and “Sparse PCA” with number of factors set to 8 from the results of “Adaptive PCA” (Right) at t = 2015/12, with the number of series on the y-axis and the number of factors in the x-axis. The factor loading are estimated using a 10 year rolling window.

The heatmap of estimated factor loadings at time 2003/12 is in Figure 5 (left). The output has been left-ordered based on the results at 2015/12, where the more active factors are on the left, in the order of data series, and some of the less active right-most factors (with small or zero loadings) are omitted. There are 24 active factors in total (i.e. factors with at least two non-negligible non-zero factor loadings), with only 5 factors that cluster eight or more series (Factors 2, 10, 22, 23, and 25). Since the variables are grouped by their economic meaning, this type of clustering is not entirely unexpected. For example, Factor 2 includes CMRMTSPLx (real manufacturing and trade industry sales), all industrial production indices except nondurable materials, residential utilities, and fuels, CUMFNS (capacity utilization), DMANEMP (durable goods employment), and ISRATIOx (manufacturing and trade inventories to sales ratio). This factor could be interpreted as a factor for *durable goods*, which include industries that are more susceptible to economic trends,

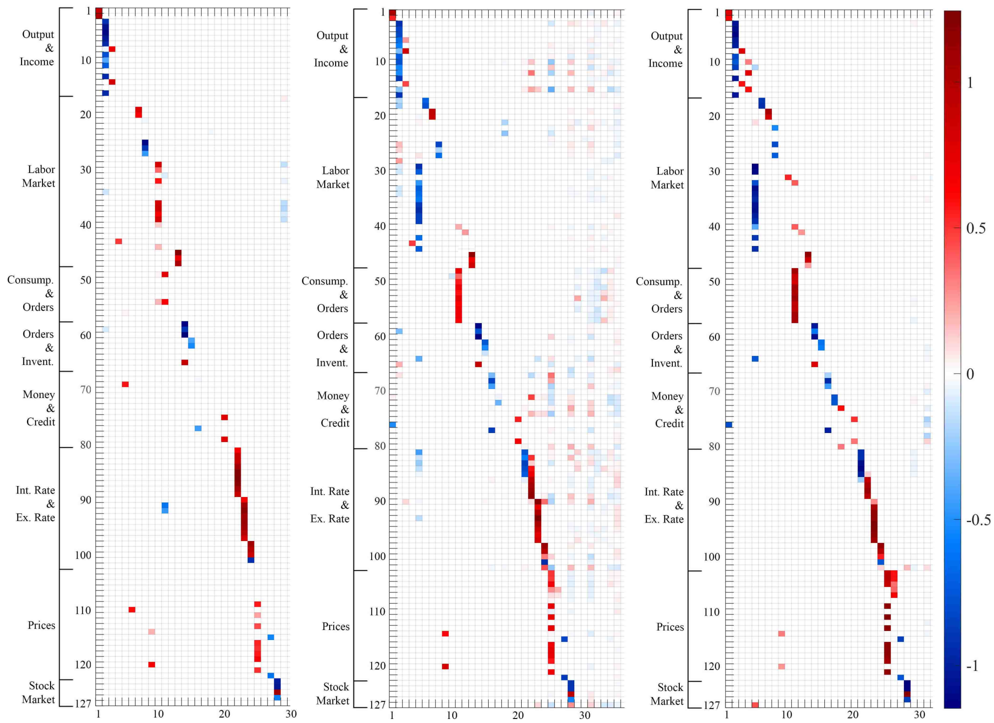


Figure 5: Macroeconomic Study: Estimated factor loadings using dynamic sparse factor analysis at t = 2003/12 (left), t = 2008/10 (center), t = 2015/12 (right), with the original series on the y-axis and the factors in the x-axis. The factor loading are estimated dynamically over the period 2001/1:2015/12.

where sales, inventories, industrial production, capacity utilization, and employment are all connected. Conversely, we expect nondurable goods, such as utilities and fuels, to have a different dynamic than durable goods, which is reflected in the exclusion of those indices in Factor 2. Similarly, Factor 10 includes employment data (except for mining and logging, manufacturing, durable goods, nondurable goods, and government), Factor 22 includes interests rates (fed funds rate, treasury bills, and bond yields), Factor 23 includes the spread between interest rates minus fed funds rate, and Factor 25 includes consumer price indices except apparel, medical care, durables, and services, as well as personal consumptions expenditures on nondurable goods. All of these factors produce meaningful and mostly separated clusters that largely conform with economic intuition.

During the crisis (Figures 5; center), radical changes occur in the factor structure.

Concerning Factor 2, the dependence structure expands, now spanning over nondurables and fuels, as well as HWI (the help wanted index), UNEMP15OV (unemployment for 15 weeks and over), CLAIMSx (unemployment insurance claims), and PAYEMS (employment, total non-farm, goods-producing, manufacturing, and durable goods). This indicates that the shock might have affected relatively stable industries and unemployment, with the co-movement across industries being largely synchronized under distress (with the exception of residential utilities). Another interesting observation is the emergence of new factors. In particular, Factor 11, which includes housing starts and new housing permits in different regions in the U.S., was *not* present pre-crisis and now surfaces as a connecting thread between housing markets across regions. While in 2003/12 the latent factors were largely separated (loadings had little overlap), we now see at least two factors (namely Factor 25 and 28), whose loadings are non-sparse and far-reaching. In particular, Factor 28 emerges as a non-sparse link between many different sectors of the economy, including retail sales, industrial production, employment (in particular financial services), real M2 money stock, loans, BAA bond yields (but not AAA), exchange rates, consumer sentiment, investment and, most importantly, the stock market indices, including the S&P 500 and the VIX (i.e. the fear index), a popular measure of the stock market's expectation of volatility. This factor loads heavily on stock market indices, which were isolated pre-crisis, but are now connected to the various corners of the economy. Factor 25, on the other hand, is driven mainly by prices (e.g. CPI). Both of these factors could potentially be interpreted as crisis factors as they permeate through various sectors of the economy, that had little interconnectivity in the pre-crisis era. The only sectors not influenced by these factors are Consumption and Orders and, more interestingly, the housing market.

There is an ongoing discussion on what were the main catalysts of the Great Recession. One line of reasoning focuses on the financial market, where the devaluation of securities, including mortgage backed securities, led to curtailed lending and thereby consumption (Benmelech et al., 2017; Chodorow-Reich, 2014). The second one focuses directly on the downturn of the housing market (Mian et al., 2013; Mian and Sufi, 2009, 2011). The "orthogonality" between the housing market factor (Factor 11) and the "crisis factors"

(Factor 25 and 28) may suggest that, while the crisis was triggered by the housing market, the main catalyst of the recession was probably the financial market. While our analysis does not necessarily prove this hypothesis, it aligns with the previous lines of reasoning.

Finally, Figure 5 (right) shows the end of the analysis at 2015/12, where the economy has mostly recovered from the Great Recession, but has fundamentally changed from what it was before. Although most of the factor overlap has dissipated, we see a notably different structure compared to 2003. In particular, Factor 5 (employment) and Factor 11 (housing) persevere from the crisis. Moreover, the "crisis factors" Factor 25 and 28, representing the prices and the stock market, are no longer strongly tied to other parts of the economy (labor, output, interest and exchange rates, etc.). In addition, the VIX indicator for market sentiment, is no longer connected to many of the key factors, even the stock market, and is only connected to Factor 5, implying that the market's anticipation of volatility is no longer severely intertwined with the rest of the economy. Factor 2 is one of the few factors that have returned back to its original structure, except for CMRMTSPLx and industrial production of nondurable consumer goods. Its dependence with the labor market (e.g. unemployment) has disappeared, suggesting that industry production is no longer in co-movement with the labor market.

We also obtain insights into the effects and duration of the crisis by looking at the evolution of the factor loadings for one of the "crisis" factors, Factor 28. Figure 6 shows a dynamic heatmap and a 3-D plot of β^t_{jk} for $1 \leq j \leq 127$ (y-axis) and $1 \leq t \leq 180$ (x-axis) with $k = 28$. For the 3-D plot, the loadings on the S&P indices are suppressed to zero in order to improve visibility. The figure reveals a spur of activity around the sharp financial crisis (late 2008 and early 2009), where the contagion battered multiple corners of the economy. The duration of the active loadings provide additional insights. For example, the loadings on VIX (series 127) emerges and disappears in a eight month span from 06/2008 to 02/2009, while the loadings on the exchange rate between U.S. and Canada lasts for 17 months. However, most factor loadings seem to only emerge for about 4-6 months.

To understand the degree of connectivity/overlap between factors, we plot the average number of active factors (with absolute loadings truncated to above 0.1)[3] per series over

[3]We use the 0.1 as cutoff, because $(-0.1, 0.1)$ is approximately the shortest 10-percent confidence interval

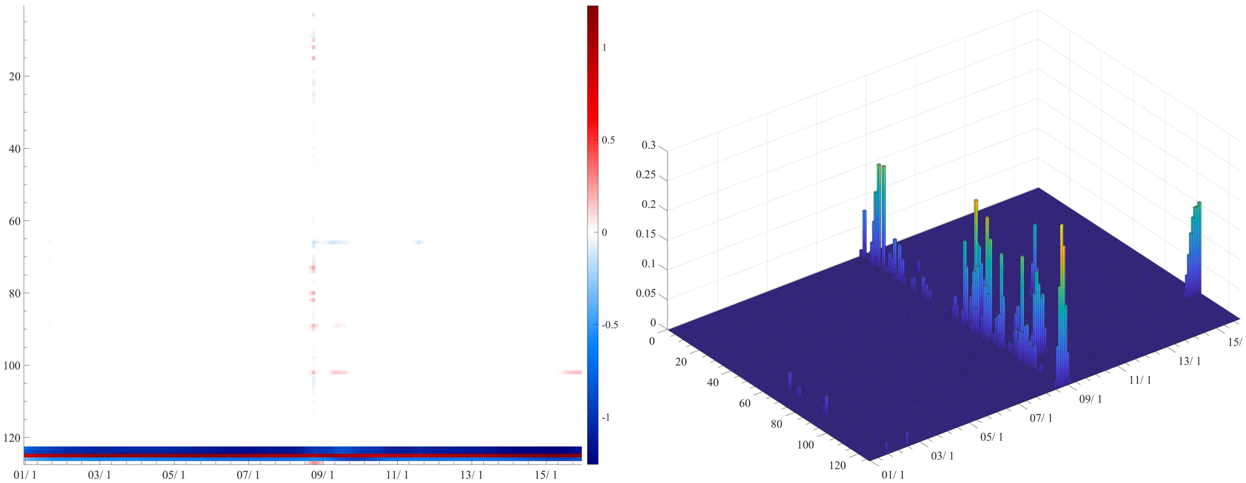


Figure 6: Macroeconomic Study: Estimated factor loadings for Factor 28 using dynamic spike-and-slab from t = 2001/12:2015/12, with a heatmap of the entire factor loadings (Left) and a 3-D plot of the factor loadings with the loadings on 123-126 (S&P related indices) set to zero to increase visibility.

time (Figure 7). More overlap indicates a more intertwined economy. We observe an increase in late 2008, reflecting the emergence of pervasive crisis factor(s), as well as its build up from mid-2006. Another point to note is that the level pre-crisis is comparatively lower than post-crisis, indicating a structural shift is the economy brought on by the crisis.

We further our analysis with a few insights into the idiosyncratic variances for variables related to the housing market: HOUST (total housing starts) and its regional variants (North East, Mid-West, South, and West). We choose the housing market for deeper analysis, because the housing market has been subjected to intense scrutiny, following the great recession of 2009, as a suspected trigger of the crisis (Mian et al., 2013; Mian and Sufi, 2009, 2011). Housing starts is the seasonally adjusted number of new residential construction projects that have begun during any particular month and, as such, is a key part of the U.S. economy, which relates to employment and many industry sectors including banking (the mortgage sector), raw materials production, construction, manufacturing, and real estate.

of the spike distribution (Laplace distribution centered at 0 and with variance being equal to 0.9) used in the dynamic spike and slab prior.

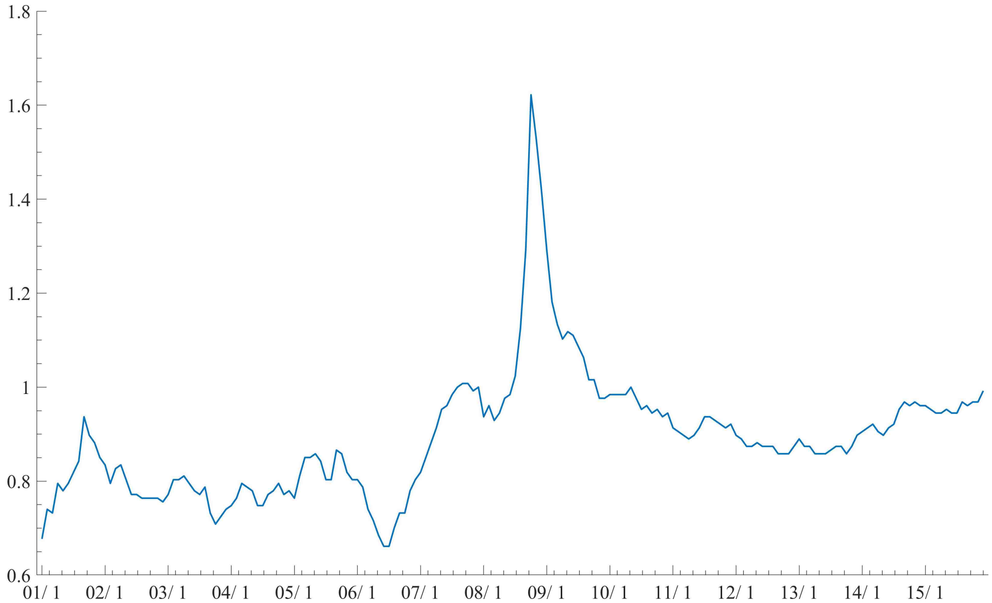


Figure 7: Macroeconomic Study: The average number of estimated active factors (with absolute loadings above 0.1)per series over the period 2001/1:2015/12.

In our earlier analysis (Figure 5) we found that, while regional indicators were not clustered pre-crisis, persistent clustering occurs post-crisis. Figure 8 portrays the series of residual uncertainties $\{\sigma_{jt}^2 : 1 \leq t \leq T\}$ for each regional housing starts indicator. We find several interesting patterns. Figure 8 indicates that increased uncertainty in housing starts is a global phenomenon but that there is heterogeneity across regions as to the magnitude and timing. For example, we find that the West region to react the earliest, followed by Mid-West and South. North-East is somewhat of an exception, as the idiosyncratic variance starts out greater than the other series, falling off pre-crisis, increasing during the crisis, and tapering off to a level similar to the other regions. The speed of mounting uncertainty could be associated with the deflation of the housing bubble after 2006 (Commission, 2011). As the economy recovers from the Great Recession, we find a steady decrease in uncertainty in all regions, except for the South region, which is persistently high throughout the analysis long after the crisis. Interestingly, the South region was one of the hardest hit regions during the great recession, with the increase in unemployment being the highest of all the regions. Possibly due to this characteristic, we find that the South region does not return

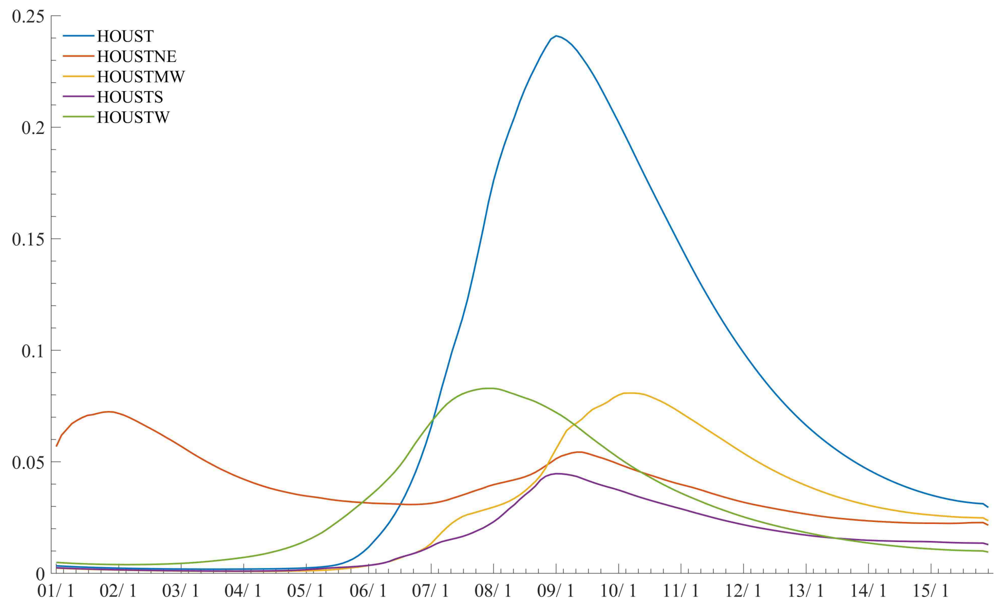


Figure 8: Macroeconomic Study: The idiosyncratic variance, $\mathbf{\Sigma}_t$, of U.S. housing starts, over the period 2001/1:2015/12.

to the pre-crisis state. This is an important insight for the decision/policy maker, as this indicates some unique circumstances in the South region that requires further investigation, where the housing bubble from 2004-2006 bursts after mid-2006.

6 Forecasting Evaluations

In this section, we compare the forecasting performance of our method (point-wise predictions as well as forecast distributions) with three alternatives. The first one is a static version of our model that assumes that factor loading matrices are constant over time, keeping all the other model assumptions the same. The second method is the sparse Bayesian latent factor stochastic volatility model implemented in the R package "factorstochvol" (Kastner, G., 2017). The third method is the hierarchical Bayesian vector autoregressive model (BVAR) of Kuschnig and Vashold (2019). It implements a hierarchical modeling approach to prior selection in the fashion of Giannone et al. (2015). For both "factorstochvol" and "BVAR" methods, we draw $15,000$ MCMC samples, of which $5,000$ samples are discarded as a burn-in. Both these methods are implemented through their corresponding R packages with default parameter settings. Forecasting comparison is conducted for four examples: (i) a smaller simulated data (described later in this section) with $p = 10$ and $T = 100$, (ii) the same simulated data as in (i) but extended to $T = 400$, (iii) the simulated data discussed in Section 4 and (iv) the macroeconomic data discussed in Section 5.

For all three factor analysis models: our "Dynamic FA", "Static FA" and "factorstochvol", we use the same upper bound on the number of latent factors (K). For the lower dimensional simulated datasets (i) and (ii) we assign $K = 6$ and for the higher dimensional simulated dataset (iii) we fix $K = 10$. In Section 5, we discovered that for the macroeconomic data, the number of active factors never exceeds $K = 28$. Therefore, for evaluating forecasting performance, we used $K = 30$ for all three factor analysis models to facilitate faster computation and higher accuracy.

Forecasting can be performed using both the EM (Section 3.1) and the MCMC (Section 3.2) implementations. Even though the EM algorithm is more practically feasible for larger datasets, a common drawback of this method is that we get only point estimates of the variables of interest over the forecast period, as opposed to the MCMC which provides the entire predictive distribution. Point forecasts for the time period $T+1$ can be obtained by training the EM algorithm with data $\boldsymbol{Y}_{1:T}$ and then computing $\widehat{\boldsymbol{Y}}_{T+1} = \widehat{\boldsymbol{B}}_{T+1}\widehat{\boldsymbol{\omega}}_{T+1}$, where $\widehat{\boldsymbol{B}}_{T+1}$ and $\widehat{\boldsymbol{\omega}}_{T+1}$ are expectations of the future value conditional on the estimates obtained from the EM algorithm. Forecasting comparisons of predictive distributions obtained from MCMC are described towards the end of this section.

For the smallest simulated data (with $p = 10$ and $T = 100$), we conduct a sequential one-step-ahead forecast for 50 consecutive time points into the future with the EM implementation of our model. The same forecasting exercise is also performed with the three competing methods under consideration. Then we plot the root mean squared prediction error (RMSE) over time for all the four methods in Figures 9(a) and 9(b). We simulate the $p = 10$-dimensional data $\boldsymbol{Y}_{1:150}$ for $T = 150$ time points as described in the Appendix (Section B1)[4].

[4]The appendix only describes the data generating process through time $T = 1 : 100$. Beyond this range, the time series undergoes two structural changes: (a) at time $t = 111$ the loadings of the third

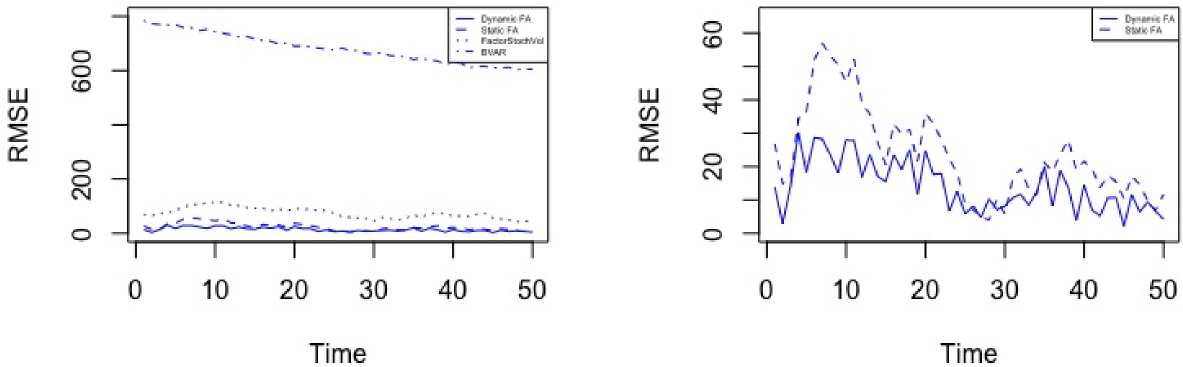


Figure 9: Root mean squared error (RMSE) computed over 50 one-step-ahead forecasts from a simulated data with $p = 10$ and $T = 100$. The dynamic sparse factor analysis model (Dynamic FA) is compared against (i) the Static spike and slab factor analysis (Static FA), (ii) FactorStochvol and (iii) Bayesian VAR (BVAR). The plot on the left shows forecasting RMSE over time for all four methods and the plot on the right zooms in on the Dynamic FA and Static FA models.

We start the forecasting exercise for the above data by first estimating the model with $\boldsymbol{Y}_{1:100}$ and then forecasting for $T = 101$. Then we incorporate the next observation $\boldsymbol{Y}_{101}$ into the training data, estimate the model again and forecast for $T = 102$. This process is repeated through $T = 101 : 150$. For each time point we compute the RMSE over the $p = 10$ simulated time series. Figure 9(a) shows how the RMSE varies over the 50 forecast points for all the four methods under consideration and Figure 9(b) zooms in onto our dynamic sparse factor analysis model and the static spike and slab factor model. We see that our dynamic factor model maintains superior forecasting performance, compared to the three competing methods throughout the forecast period. An interesting point to

factor become inactive and (d) at $t = 120$ the loadings of the third factor are re-introduced and they remain active until $T = 150$. Therefore the number of active factors K transition as follows: $K = 3$ for $T = 1 : 35$, $K = 2$ for $T = 36 : 45$, $K = 3$ for $T = 46 : 110$, $K = 2$ for $T = 111 : 120$ and finally $K = 3$ for $T = 121 : 150$. All other aspects of the simulation remain same as described in Section B1 of the appendix.

Data	Dimension	Dynamic FA	Static FA	FactorStochVol	BVAR
Simulation	p=10, T=100	**20.427**	34.359	63.287	752.157
Simulation	p=10, T=400	**1.158**	13.435	2.613	230.75
Simulation	p=100, T=400	**1.407**	15.319	2.389	392.186
Macroeconomy	p=127, T=180	**0.623**	1.018	1.384	1.3453

Table 3: Root mean squared forecasting error over five time points into the future

note here is that the forecast accuracy of our dynamic model persists even when the time series structurally changes at time points $T = 111$ and $T = 120$, where the number of factors changes from 3 to 2 and then again from 2 back to 3 respectively. In contrast, the factor models with static loadings (static FA and "factorstochvol") cannot adapt to these structural changes and their forecasting performance decline resulting in higher RMSE.

Next, we compare the five-step ahead forecasting performance of our dynamic model (referred to as "Dynamic FA") with the static model described above (referred to as "Static FA"), the "factorstochvol" method and the BVAR method in Table 3. Comparison is done with respect to the cumulative root mean squared forecasting errors (RMSE), computed over five time points into the future. This is a five-step-ahead forecasting exercise, as opposed to the one-step-ahead forecasting over 50 time points demonstrated in Figures 9(a) and 9(b). The five-step-ahead forecast is conducted by sequential point forecasts from the EM algorithm over five time points into the future. Specifically, we fit the model based on the first T observations $\boldsymbol{Y}_{1:T}$, and compute the one-step ahead forecast $\boldsymbol{Y}_{T+1}$. Then, we add this forecast for $\boldsymbol{Y}_{T+1}$ into the training data $\boldsymbol{Y}_{1:T}$ and compute a forecast $\boldsymbol{Y}_{T+2}$ based on $\boldsymbol{Y}_{1:(T+1)}$. We repeat these sequential updates for five time points to predict $\boldsymbol{Y}_{(T+1):(T+5)}$. These predicted values are compared with the observed/simulated data to get a cumulative root mean squared prediction error. Similarly, for the macroeconomic data we use the first 175 months (2001/01 to 2015/07) to get sequential one-step-ahead forecast for the last 5 months (2015/08 to 2015/12). Table 3 shows that, for all the simulated datasets, the three factor analysis models (Dynamic FA, static FA and FactorStochVol) perform significantly better than the Bayesian VAR model. For the macroeconomic data, our Dynamic factor analysis model appears to perform considerably better than the alternatives. This reiterates the merits of using dynamic factor loadings as opposed to constant loading matrices.

The above observations are confirmed after computing the one-step-ahead log predictive density scores (LPDS) measuring the quality of the entire forecast distributions. For this forecasting comparison, we use our MCMC implementation (Section 3.2). As described by Kastner (2019), the one step ahead LPDS for the dynamic factor model can be computed by first drawing M MCMC samples from the distribution of $\mathbf{Y_{1:T}}$ and then averaging over $m = 1, \ldots, M$ densities of

$$\mathcal{N}_p\left(\mathbf{0}, \boldsymbol{B}^{(m)}_{(T+1):[1:T]}\boldsymbol{B}^{(m)\prime}_{(T+1):[1:T]} + \boldsymbol{\Sigma}^{(m)}_{(T+1):[1:T]}\right)$$

evaluated at $\boldsymbol{Y}_{T+1}$, where $\boldsymbol{B}^{(m)}_{T+1:[T+1]}$ and $\boldsymbol{\Sigma}^{(m)}_{(T+1):[1:T]}$ denote the m-th draw of $\boldsymbol{B}_{T+1}$ and $\boldsymbol{\Sigma}_{T+1}$ respectively, from the posterior distribution up to time T. Next, we compute the one-step ahead log predictive Bayes factor between our dynamic factor model and the static spike-and-slab factor model. Such Bayes factor between any two models $\mathcal{M}_1$ and $\mathcal{M}_2$ is defined as $\log BF(\mathcal{M}_1, \mathcal{M}_2) = \log PL_{T+1}(\mathcal{M}_1) - \log PL_{T+1}(\mathcal{M}_2)$, where $PL_t(\mathcal{M})$ denotes the predictive likelihood of model $\mathcal{M}$ at time $T+1$. When the log predictive Bayes factor is greater than zero at a given point in time, there is evidence in favor of model $\mathcal{M}_1$ as opposed to model $\mathcal{M}_2$, and vice versa. For the simulated examples with $p = 10$ and $T = 100$, the Bayes factor computed between our dynamic factor model and the static factor model turn out to be equal to

$$\log BF(\text{Dynamic FA, Static FA}) = 2.161$$

implying (strong) evidence in favor of the dynamic sparse factor model.

7 Further Comments

Motivated by a topical macroeconomic dataset, we developed a Bayesian method for dynamic sparse factor analysis for large-scale time series data. Our proposed methodology aims to tackle three challenges of dynamic factor analysis: time-varying patterns of sparsity, unknown number of factors, and identifiability constraints. By deploying dynamic sparsity, we successfully recover interpretable latent structures that automatically select the number of factors and that incorporate time-varying loadings/factors. We successfully applied our methodology on a nontrivial simulated example as well as a real dataset comprising of 127 U.S. macroeconomic indices tracked over the period of the Great Recession (and beyond) and obtained several interpretable findings.

Our methodology can be enriched/extended in many ways. One possible extension would be to develop a latent variable method that can capture within, as well as between, connectivity of several high-dimensional time series. This could be achieved with a dynamic extension of sparse canonical correlation analysis (Witten et al., 2009). Our method can also be embedded within FAVAR models (Bernanke et al., 2005) that include both observed and unobserved predictors. Additionally, throughout our analysis we have assumed the covariance of the latent factors to be fixed over time and equal to an identity matrix, one could in principle incorporate dynamic variances with stochastic volatility modeling, along the lines of Zhou et al. (2014).

One possible shortcoming of our EM-based estimation strategy, is the lack of uncertainty assessment, which is essential for forecasting. The EM algorithm, however, was the key to obtaining interpretable latent structures for very high dimensional data. To achieve uncertainty quantification along with interpretability, one could impose structural identification constraints, such as Nakajima and West (2013a,b), and perform MCMC for DSS priors, as demonstrated in Appendix B. Another approach would be to apply our method simply as a means of obtaining identifiability constraints (i.e. the sparsity pattern) and then reestimate the nonzero loadings with an MCMC strategy. While this would not quantify any sparsity-selection uncertainty, it would be an effective way to balance interpretability and forecasting/decision making. Another unavoidable feature of our method is its sensitivity to starting values. We strongly recommend using the output from the rolling window spike-and-slab factor model.

References

Aguilar, O., G. Huerta, R. Prado, and M. West (1998). Bayesian inference on latent structure in time series. *Bayesian Statistics 6*(1), 1–16.

Aguilar, O. and M. West (1998). *Bayesian dynamic factor models and variance matrix discounting for portfolio allocation*. Institute of Statistics and Decision Sciences, Duke University.

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics 18*(3), 338–357.

Aunsri, N. and P. Taveeapiradeecharoen (2020). A time-varying bayesian compressed vector autoregression for macroeconomic forecasting. *IEEE Access 8*, 192777–192786.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*(1), 18–29.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of applied Econometrics 25*(1), 71–92.

Baumeister, C., P. Liu, and H. Mumtaz. Changes in the transmission of monetary policy: Evidence from a time-varying factor-augmented VAR. *Working Paper No. 401, Bank of England*.

Benmelech, E., R. R. Meisenzahl, and R. Ramcharan (2017). The real effects of liquidity during the financial crisis: Evidence from automobiles. *The Quarterly Journal of Economics 132*(1), 317–365.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of Economics 120*(1), 387–422.

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

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

Burns, A. F. and W. C. Mitchell (1947). *Measuring Business Cycles*. The National Bureau of Economic Research.

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*(484), 1438–1456.

Carvalho, C. M., H. F. Lopes, and O. Aguilar (2011). Dynamic stock selection strategies: A structured factor model framework. *Bayesian Statistics 9*, 1–21.

Cheng, X., Z. Liao, and F. Schorfheide (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies 83*(4), 1511–1543.

Chodorow-Reich, G. (2014). Effects of unconventional monetary policy on financial institutions. Technical report, National Bureau of Economic Research.

Commission, F. C. I. (2011). *The financial crisis inquiry report, authorized edition: Final report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*. Public Affairs.

Daniele, M. and J. Schnaitmann (2019). A regularized factor-augmented vector autoregressive model. *arXiv preprint arXiv:1912.06049*.

Del Negro, M. and C. Otrok (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. *FRB of New York Staff Report No.326*.

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

Diebold, F. X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics 4*(1), 1–21.

Evgenidis, A., D. Philippas, and C. Siriopoulos (2019). Heterogeneous effects in the international transmission of the us monetary policy: a factor-augmented var perspective. *Empirical Economics 56*(5), 1549–1579.

Farmer, L., L. Schmidt, and A. Timmermann (2018). Pockets of predictability. *CEPR Discussion Paper No.DP12885*.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis 15*(2), 183–202.

Frühwirth-Schnatter, S. and H. Lopes (2009). Parsimonious Bayesian factor analysis when the number of factors is unknown. Technical report, University of Chicago Booth School of Business.

Frühwirth-Schnatter, S. and H. F. Lopes (2018). Sparse Bayesian Factor Analysis when the number of factors is unknown. *arXiv:1804.04231*.

Frühwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics 154*(1), 85–100.

Gargano, A., D. Pettenuzzo, and A. Timmermann (2019). Bond return predictability: Economic value and links to the macroeconomy. *Management Science 65*(2), 508–540.

George, E. I. (1986a). Combining minimax shrinkage estimators. *Journal of the American Statistical Association 81*(394), 437–445.

George, E. I. (1986b). Minimax multiple shrinkage estimation. *The Annals of Statistics*, 188–205.

George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series. *In: Aigner, D.J. and Goldberger, A.S., Eds., Latent Variables in Socio-Economic Models 1, North-Holland, Amsterdam.*.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *The review of financial studies 9*(2), 557–587.

Ghosh, J. and D. B. Dunson (2009). Default prior distributions and efficient posterior computation in bayesian factor analysis. *Journal of Computational and Graphical Statistics 18*(2), 306–320.

Giannone, D., M. Lenza, D. Momferatou, and L. Onorante (2014). Short-term inflation projections: A bayesian vector autoregressive approach. *International journal of forecasting 30*(3), 635–644.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Guhaniyogi, R. and D. B. Dunson (2015). Bayesian compressed regression. *Journal of the American Statistical Association 110*(512), 1500–1514.

Hallin, M. and R. Liska (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*(478), 603–617.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Huber, F. and M. Feldkircher (2019). Adaptive shrinkage in bayesian vector autoregressive models. *Journal of Business & Economic Statistics 37*(1), 27–39.

Ishwaran, H., J. S. Rao, et al. (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *The Annals of Statistics 33*(2), 730–773.

Kastner, G. (2019). Sparse bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics 210*(1), 98–115.

Kastner, G., S. Frühwirth-Schnatter, and H. F. Lopes (2017). Efficient bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics 26*(4), 905–917.

Kastner, G. and F. Huber (2020). Sparse bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*.

Kastner, G. (2017). Bayesian estimation of (sparse) latent factor stochastic volatility models. R package version 0.8.4.

Kaufmann, S. and S. Beyeler (2018). Factor augmented var revisited-a sparse dynamic factor model approach.

Kaufmann, S. and C. Schumacher (2017). Identifying relevant and irrelevant variables in sparse factor models. *Journal of Applied Econometrics 32*(6), 1123–1144.

Kaufmann, S. and C. Schumacher (2019). Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification. *Journal of Econometrics 210*(1), 116–134.

Koop, G. and D. Korobilis (2013). Large time-varying parameter vars. *Journal of Econometrics 177*(2), 185–198.

Koop, G., D. Korobilis, et al. (2010). Bayesian multivariate time series methods for empirical Macroeconomics. *Foundations and Trends® in Econometrics 3*(4), 267–358.

Koop, G., D. Korobilis, and D. Pettenuzzo (2019). Bayesian compressed vector autoregressions. *Journal of Econometrics 210*(1), 135–154.

Korobilis, D. (2013). Var forecasting using bayesian variable selection. *Journal of Applied Econometrics 28*(2), 204–230.

Kuschnig, N. and L. Vashold (2019). BVAR: bayesian vector autoregressions with hierarchical prior selection in R.

Liu, C., D. B. Rubin, and Y. N. Wu (1998a). Parameter expansion to accelerate em: the px-em algorithm. *Biometrika 85*(4), 755–770.

Liu, C., D. B. Rubin, and Y. N. Wu (1998b). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika 85*(4), 755–770.

Liu, J. S. and Y. N. Wu (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association 94*(448), 1264–1274.

Lopes, H. F. and C. M. Carvalho (2007). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference 137*(10), 3082–3091.

Lopes, H. F., R. McCulloch, and R. Tsay (2010). Cholesky stochastic volatility. *Unpublished Technical Report, University of Chicago, Booth Business School 2.*

Lopes, H. F. and H. S. Migon (2002). Comovements and contagion in emergent markets: stock indexes volatilities. In *Case studies in Bayesian statistics*, pp. 285–300. Springer.

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

McCracken, M. W. and S. Ng (2016). FRED-MD: a monthly database for Macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Meng, X.-L. and D. Van Dyk (1998). Fast em-type implementations for mixed effects models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(3), 559–578.

Meng, X.-L. and D. A. Van Dyk (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika 86*(2), 301–320.

Mian, A., K. Rao, and A. Sufi (2013). Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics 128*(4), 1687–1726.

Mian, A. and A. Sufi (2009). The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *The Quarterly Journal of Economics 124*(4), 1449–1496.

Mian, A. and A. Sufi (2011). House prices, home equity-based borrowing, and the US household leverage crisis. *American Economic Review 101*(5), 2132–56.

Miranda-Agrippino, S. and G. Ricco (2018). The transmission of monetary policy shocks.

Nakajima, J. and M. West (2013a). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics 31*(2), 151–164.

Nakajima, J. and M. West (2013b). Dynamic factor volatility modeling: A Bayesian latent threshold approach. *Journal of Financial Econometrics 11*(1), 116–153.

Nakajima, J., M. West, et al. (2017). Dynamics & sparsity in latent threshold factor models: A study in multivariate EEG signal processing. *Brazilian Journal of Probability and Statistics 31*(4), 701–731.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica 77*(5), 1447–1479.

Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Pitt, M. and N. Shephard (1999). Time varying covariances: a factor stochastic volatility approach. *Bayesian Statistics 6*, 547–570.

Potjagailo, G. (2017). Spillover effects from euro area monetary policy across europe: A factor-augmented var approach. *Journal of International Money and Finance 72*, 127–147.

Prado, R. and M. West (2010). *Time Series: Modelling, Computation & Inference*. Chapman & Hall/CRC Press.

Reinhart, C. M. and K. S. Rogoff (2008). Is the 2007 US sub-prime financial crisis so different? An international historical comparison. *American Economic Review 98*(2), 339–44.

Rockova, V. (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *The Annals of Statistics 46*(1), 401–437.

Rockova, V. and E. I. George (2016). Fast Bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association 111*(516), 1608–1622.

Rockova, V. and E. I. George (2018). The Spike-and-Slab LASSO. *Journal of the American Statistical Association 113*, 431–444.

Rockova, V., K. McAlinn, et al. (2020). Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis*.

Rubin, D. B. and D. T. Thayer (1982). EM algorithms for ML factor analysis. *Psychometrika 47*(1), 69–76.

Sargent, T. J., C. A. Sims, et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in Business cycle research 1*, 145–168.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for var analysis. Technical report, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (2010). Modeling inflation after the crisis. Technical report, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, Volume 2, pp. 415–525. Elsevier.

Stock, J. H. and M. W. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal 128*(610), 917–948.

Van Dyk, D. A. and R. Tang (2003). The one-step-late pxem algorithm. *Statistics and Computing 13*(2), 137–152.

Wagan, Z. A., Z. Chen, H. Wagan, et al. (2019). A factor-augmented vector autoregressive approach to analyze the transmission of monetary policy. *Prague Economic Papers 28*(6), 709–728.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics 23*(3), 385–400.

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

West, M. and P. J. Harrison (1997). *Bayesian Forecasting & Dynamic Models* (2nd ed.). Springer Verlag.

Witten, D. M., R. Tibshirani, and T. Hastie (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics 10*(3), 515–534.

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

Yu, Y. and X.-L. Meng (2011). To center or not to center: That is not the questionan ancillarity–sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics 20*(3), 531–570.

Zhou, X., J. Nakajima, and M. West (2014). Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. *International Journal of Forecasting 30*(4), 963–980.

Zuur, A. F., R. Fryer, I. Jolliffe, R. Dekker, and J. Beukema (2003). Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics 14*(7), 665–685.

Dynamic Sparse Factor Analysis

Supplementary Material

A Appendix

A.1 Derivation of the E-step

We now outline the steps of the parameter expanded EM algorithm. In the E-step, we compute the conditional expectation of the augmented and expanded log-posterior with respect to the missing data $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$, given observed data $\boldsymbol{Y}$ and the parameter values $\boldsymbol{\Delta}^{(m)}$ obtained at the previous M-step setting $\boldsymbol{A}_t = \boldsymbol{I}_K$. We can write

$$\mathbb{E}_{\boldsymbol{\Gamma},\boldsymbol{\Omega}\,|\,\boldsymbol{Y},\Delta^{(m)}}[\log \pi(\boldsymbol{B}^\star_{0:T}, \boldsymbol{\Sigma}_{1:T}, \boldsymbol{A}_{1:T}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}\,|\,\boldsymbol{Y})] = Q_1(\boldsymbol{B}^\star_{0:T}\,|\,\boldsymbol{\Sigma}_{1:T}) + Q_2(\boldsymbol{\Sigma}_{1:T}) + Q_3(\boldsymbol{A}_{1:T}) + C. \tag{12}$$

Define $\boldsymbol{\omega}_{t|T} = \mathbb{E}_{\boldsymbol{\Omega}}[\boldsymbol{\omega}_t\,|\,\boldsymbol{Y}, \boldsymbol{\Delta}^{(m)}]$, $\boldsymbol{V}_{t|T} = \text{cov}[\boldsymbol{\omega}_t\,|\,\boldsymbol{Y}, \boldsymbol{\Delta}^{(m)}]$. The terms $\boldsymbol{\omega}_{t|T}$ and $\boldsymbol{V}_{t|T}$ represent the best linear estimator for $\boldsymbol{\omega}_t$ using all observations and the corresponding covariance matrix, respectively. With $\boldsymbol{V}_{t,t-1|T} = \text{cov}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_{t-1}\,|\,\boldsymbol{Y}, \boldsymbol{\Delta}^{(m)}]$ we denote the covariance matrix of $\boldsymbol{\omega}_t$ and $\boldsymbol{\omega}_{t-1}$ given the data $\boldsymbol{Y}$ and $\boldsymbol{\Delta}^{(m)}$. These quantities can be obtained from the Kalman Filter and Smoother Algorithm (Table 4).

The functions $Q_1(\cdot)$, $Q_2(\cdot)$ and $Q_3(\cdot)$ in (12) can be written as follows:

$$\begin{aligned} -Q_1(\boldsymbol{B}^\star_{0:T}\,|\,\boldsymbol{\Sigma}_{1:T}) =& C + \frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{P}\log\sigma^2_{jt} \\ &+ tr\left\{\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\Sigma}_t^{-1}\left[(\boldsymbol{Y}_t - \boldsymbol{B}^\star_t\boldsymbol{\omega}_{t|T})(\boldsymbol{Y}_t - \boldsymbol{B}^\star_t\boldsymbol{\omega}_{t|T})' + \boldsymbol{B}^\star_t\boldsymbol{V}_{t|T}\boldsymbol{B}^{\star'}_t\right]\right\} \\ &+ \sum_{j=1}^{P}\sum_{k=1}^{K}\left[\frac{\langle\gamma^0_{jk}\rangle(\beta^{0*}_{jk})^2}{2\lambda_1/(1-\phi^2)} + (1-\langle\gamma^0_{jk}\rangle)|\beta^{0*}_{jk}|\lambda_0\right] \\ &+ \sum_{t=1}^{T}\sum_{j=1}^{P}\sum_{k=1}^{K}\left[\frac{\langle\gamma^t_{jk}\rangle(\beta^{t*}_{jk} - \phi\beta^{t-1*}_{jk})^2}{2\lambda_1} + (1-\langle\gamma^t_{jk}\rangle)|\beta^{t*}_{jk}|\lambda_0\right], \end{aligned}$$

Algorithm: *Kalman Filter and Smoother*	
Initialize $\boldsymbol{\omega}_{0\,|\,0} = \mathbf{0}$ and $\boldsymbol{V}_{0\,|\,0} = 1/(1-\widetilde{\phi}^2)\boldsymbol{I}_K$ Repeat the Prediction Step and Correction Step for $t = 1, \ldots, T$	
Prediction Step	$\boldsymbol{\omega}_{t\,|\,t-1} = \boldsymbol{\omega}_{t-1\,|\,t-1}$ $\boldsymbol{V}_{t\,|\,t-1} = \boldsymbol{V}_{t-1\,|\,t-1} + \boldsymbol{I}_K$
Correction Step	$\boldsymbol{K}_t = \boldsymbol{V}_{t\,|\,t-1}\boldsymbol{B}_t'(\boldsymbol{B}_t\boldsymbol{V}_{t\,|\,t-1}\boldsymbol{B}_t' + \boldsymbol{\Sigma}_t)^{-1}$ $\boldsymbol{\omega}_{t\,|\,t} = \boldsymbol{\omega}_{t\,|\,t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \boldsymbol{B}_t\boldsymbol{\omega}_{t\,|\,t-1})$ $\boldsymbol{V}_{t\,|\,t} = \boldsymbol{V}_{t\,|\,t-1} - \boldsymbol{K}_t\boldsymbol{B}_t\boldsymbol{V}_{t\,|\,t-1}$
Initialize $\boldsymbol{V}_{T,T-1\,|\,T} = (I - \boldsymbol{K}_T\boldsymbol{B}_T)\boldsymbol{V}_{T-1\,|\,T-1}$ Repeat the smoothing step for $t = T, \ldots, 1$	
Smoothing Step	$\boldsymbol{\omega}_{t-1\,|\,T} = \boldsymbol{\omega}_{t-1\,|\,t-1} + \boldsymbol{Z}_{t-1}(\boldsymbol{\omega}_{t\,|\,T} - \boldsymbol{\omega}_{t\,|\,t-1})$ $\boldsymbol{V}_{t-1\,|\,T} = \boldsymbol{V}_{t-1\,|\,t-1} + \boldsymbol{Z}_{t-1}(\boldsymbol{V}_{t\,|\,T} - \boldsymbol{V}_{t\,|\,t-1})\boldsymbol{Z}_{t-1}'$ $\boldsymbol{V}_{t,t-1\,|\,T} = \boldsymbol{V}_{t-1\,|\,t-1}\boldsymbol{Z}_{t-2}' + \boldsymbol{Z}_{t-1}(\boldsymbol{V}_{t,t-1\,|\,T} - \boldsymbol{V}_{t-1\,|\,t-1})\boldsymbol{Z}_{t-2}'$ where $\boldsymbol{Z}_{t-1} = \boldsymbol{V}_{t-1\,|\,t-1}\boldsymbol{V}_{t\,|\,t-1}^{-1}$

Table 4: Kalman Filter and Smoother Algorithm for Parameter Expanded EM using rotated loading matrices $\boldsymbol{B}_{1:T}$

where

$$\langle\gamma_{jk}^0\rangle = \frac{\Theta\psi_1(\beta_{jk}^0|0, \frac{\lambda_1}{1-\phi^2})}{\Theta\psi_1(\beta_{jk}^0|0, \frac{\lambda_1}{1-\phi^2}) + (1-\Theta)\psi_0(\beta_{jk}^0|0, \lambda_0)},$$

$$\langle\gamma_{jk}^t\rangle = \frac{\theta_{jk}^t\psi_1(\beta_{jk}^t|\phi\beta_{jk}^{t-1}, \lambda_1)}{\theta_{jk}^t\psi_1(\beta_{jk}^t|\phi\beta_{jk}^{t-1}, \lambda_1) + (1-\theta_{jk}^t)\psi_0(\beta_{jk}^t|0, \lambda_0)},$$

$$-Q_2(\boldsymbol{\Sigma}_{1:T}) = \sum_{t=1}^{T-1}\sum_{j=1}^{P}\left[pen(\sigma_{jt}^2 \mid \sigma_{j(t-1)}^2) + pen(\sigma_{jt}^2 \mid \sigma_{j(t+1)}^2)\right] + \sum_{j=1}^{P} pen(\sigma_{jT}^2 \mid \sigma_{j(T-1)}^2)$$

where

$$pen(\sigma_{jt}^2 \mid \sigma_{j(t-1)}^2) = \left(\frac{\delta n_{t-1}}{2} - 1\right)\log\sigma_{jt}^2 - \left(\frac{(1-\delta)n_{t-1}}{2} - 1\right)\log\left(1 - \frac{\delta\sigma_{j(t-1)}^2}{\sigma_{jt}^2}\right),$$

$$pen(\sigma_{jt}^2 \mid \sigma_{j(t+1)}^2) = -\left(\frac{\delta n_t}{2} - 1\right)\log\sigma_{jt}^2 + \left(\frac{(1-\delta)n_t}{2} - 1\right)\log\left(1 - \frac{\delta\sigma_{jt}^2}{\sigma_{j(t+1)}^2}\right),$$

and

$$-Q_3(\boldsymbol{A}_{1:T}) = \frac{1}{2}\sum_{t=1}^{T}\log|\boldsymbol{A}_t| + \frac{1}{2}tr\{\boldsymbol{A}_t^{-1}(\boldsymbol{M}_{1t} - \boldsymbol{M}_{12t} - \boldsymbol{M}_{12t}' + \boldsymbol{M}_{2t})\},$$

where

$$\boldsymbol{M}_{1t} = (\boldsymbol{\omega}_{t-1\,|\,T}\boldsymbol{\omega}'_{t-1\,|\,T} + \boldsymbol{V}_{t-1\,|\,T}),$$
$$\boldsymbol{M}_{12t} = (\boldsymbol{\omega}_{t-1\,|\,T}\boldsymbol{\omega}'_{t\,|\,T} + \boldsymbol{V}_{t,t-1\,|\,T}),$$
$$\boldsymbol{M}_{2t} = (\boldsymbol{\omega}_{t\,|\,T}\boldsymbol{\omega}'_{t\,|\,T} + \boldsymbol{V}_{t\,|\,T}).$$

A.2 Derivation of the M-step

In the M-step, we optimize the function $Q_1(\cdot)$ with respect to $\boldsymbol{B}^\star_{0:T}$, given values of $\boldsymbol{\Sigma}_{1:T}$ from the previous M-step. Given the new values $\boldsymbol{B}^{\star(m+1)}_{0:T}$ and the posterior moment estimates of the latent factors obtained from the Kalman filter, we optimize $Q_1(\cdot)+Q_2(\cdot)$, with respect to $\boldsymbol{\Sigma}_{1:T}$. Finally, we optimize the function $Q_3(\cdot)$ with respect to $\boldsymbol{A}_{1:T}$.

Optimizing $Q_1(\cdot)$ with respect to $\boldsymbol{B}^\star_{0:T}$ boils down to solving a series of independent dynamic spike and slab LASSO regressions. This is justified by the following lemma.

Lemma A.1. *Let $\boldsymbol{Y}^t = (Y^t_1,\dots,Y^t_P)' \in \mathbb{R}^P$ denote the snapshot of the series at time t and for $1 \le j \le P$ define a zero-augmented response vector at time t with $\widetilde{\boldsymbol{Y}}^t_j = (Y^t_j, \underbrace{0,\dots,0}_{K})' \in \mathbb{R}^{K+1}$. For the SVD decomposition $\boldsymbol{V}_{t\,|\,T} = \sum_{k=1}^K s_k \boldsymbol{U}^t_k(\boldsymbol{U}^t_k)'$, we denote with $\widetilde{\boldsymbol{U}}^t_k = \sqrt{s_k}\boldsymbol{U}^t_k$ and with $\boldsymbol{\Omega}^t = [\boldsymbol{\omega}_{t|T}, \widetilde{\boldsymbol{U}}^t_1,\dots,\widetilde{\boldsymbol{U}}^t_K]' \in \mathbb{R}^{(1+K)\times K}$ and we let $\boldsymbol{\beta}^{t\star\prime}_j \in \mathbb{R}^K$ be the j^{th} row of $\boldsymbol{B}^\star_t$. Then we can decompose*

$$Q_1(\boldsymbol{B}^\star_{0:T}\,|\,\boldsymbol{\Sigma}_{1:T}) = C + \sum_{j=1}^P \left[Q_j(\boldsymbol{\beta}^{t\star}_j) + Q^0(\boldsymbol{\beta}^{0\star}_j) + \widetilde{Q}(\boldsymbol{\beta}^{1\star}_j,\dots,\boldsymbol{\beta}^{T\star}_j)\right],$$

where

$$Q^0(\boldsymbol{\beta}^{0\star}_j) = \sum_{k=1}^K \left[\frac{\langle\gamma^0_{jk}\rangle(\beta^{0*}_{jk})^2}{2\lambda_1/(1-\phi^2)} + (1-\langle\gamma^0_{jk}\rangle)|\beta^{0*}_{jk}|\lambda_0\right]$$
$$Q_j(\boldsymbol{\beta}^{t\star}_j) = \sum_{t=1}^T \left[\frac{1}{2}\log\sigma^2_{jt} + \frac{1}{2\sigma^2_{jt}}||\widetilde{\boldsymbol{Y}}^t_j - \boldsymbol{\Omega}^t\boldsymbol{\beta}^{t\star}_j||^2_2\right]$$
$$\widetilde{Q}(\boldsymbol{\beta}^{1\star}_j,\dots,\boldsymbol{\beta}^{T\star}_j) = \sum_{t=1}^T\sum_{k=1}^K \left[\frac{\langle\gamma^t_{jk}\rangle(\beta^{t*}_{jk} - \phi\beta^{t-1*}_{jk})^2}{2\lambda_1} + (1-\langle\gamma^t_{jk}\rangle)|\beta^{t*}_{jk}|\lambda_0\right].$$

Proof. Denote with

$$L \equiv tr\left\{\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\Sigma}_t^{-1}\left[(\boldsymbol{Y}_t - \boldsymbol{B}_t^{\star}\boldsymbol{\omega}_{t|T})(\boldsymbol{Y}_t - \boldsymbol{B}_t^{\star}\boldsymbol{\omega}_{t|T})' + \boldsymbol{B}_t^{\star}\boldsymbol{V}_{t|T}\boldsymbol{B}_t^{\star'}\right]\right\}.$$

Because $\boldsymbol{B}_t^{\star}\boldsymbol{V}_{t\,|\,T}\boldsymbol{B}_t^{\star'} = \boldsymbol{B}_t^{\star}\sum_{k=1}^{K} s_k \boldsymbol{U}_k^t \boldsymbol{U}_k^{t'}(\boldsymbol{B}_t^{\star})' = \sum_{k=1}^{K}(\boldsymbol{0} - \boldsymbol{B}_t^{\star}\widetilde{\boldsymbol{U}}_k^t)(\boldsymbol{0} - \boldsymbol{B}_t^{\star}\widetilde{\boldsymbol{U}}_k^t)'$, we have

$$tr\left\{\boldsymbol{\Sigma}_t^{-1}\boldsymbol{B}_t^{\star}\boldsymbol{V}_{t\,|\,T}\boldsymbol{B}_t^{\star'}\right\} = \sum_{k=1}^{K}(\boldsymbol{0} - \boldsymbol{B}_t^{\star}\widetilde{\boldsymbol{U}}_k^t)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{0} - \boldsymbol{B}_t^{\star}\widetilde{\boldsymbol{U}}_k^t).$$

Since $\boldsymbol{\Sigma}_t = \mathrm{diag}(\sigma_{1t}^2, \ldots, \sigma_{Pt}^2)$, we have

$$\begin{aligned} L &= \frac{1}{2}\sum_{j=1}^{P}\sum_{t=1}^{T}\left[\frac{(Y_j^t - \boldsymbol{\omega}_{t\,|\,T}'\boldsymbol{\beta}_j^{t\star})^2}{\sigma_{jt}^2} + \sum_{k=1}^{K}\frac{(0 - \widetilde{\boldsymbol{U}}_k^{t'}\boldsymbol{\beta}_j^{t\star})^2}{\sigma_{jt}^2}\right] \\ &= \sum_{j=1}^{P}\sum_{t=1}^{T}\frac{1}{2\sigma_{jt}^2}||\widetilde{\boldsymbol{Y}}_j^t - \boldsymbol{\Omega}^t\boldsymbol{\beta}_j^{t\star}||_2^2. \end{aligned}$$

□

Each summand $Q_j(\boldsymbol{\beta}_j^{t\star}) + Q^0(\boldsymbol{\beta}_j^{0\star}) + \widetilde{Q}(\boldsymbol{\beta}_j^{1\star}, \ldots, \boldsymbol{\beta}_j^{T\star})$ corresponds to a penalized dynamic regression with $K+1$ observations at each time t. Given $\boldsymbol{\Sigma}_t$, finding $\boldsymbol{B}^{\star(m+1)}$ thereby reduces to solving these J individual regressions. As shown in Rockova et al. (2020), each regression can be decomposed into a sequence of univariate optimization problems. We use the one-step late EM variant in Rockova et al. (2020) to obtain closed form one-site updates for each $\beta_{jk}^{\star t}$ for (j,k,t). Note that this corresponds to a generalized EM, which is aimed at improving the objective relative to the last iteration (not necessarily maximizing it).

These univariate updates are slightly different from Rockova et al. (2020), because we now have $K+1$ observations at time t, not just one. Denote with $\widehat{\beta}_{jl}^{*t}$ the most recent update of the coefficient $\beta_{jl}^{\star t}$. Let

$$z_{jk}^t = \frac{1}{\sigma_{jt}^2}\sum_{r=1}^{K+1}(\widetilde{Y}_{jr}^t - \sum_{l\neq k}\widetilde{\omega}_{rl}^t\widehat{\beta}_{jl}^{t*})\widetilde{\omega}_{rk}^t$$

and denote

$$Z_{jk}^t = z_{jk}^t + \frac{\langle\gamma_{jk}^t\rangle\phi_1}{\lambda_1}\widehat{\beta}_{jk}^{t-1} + \frac{\langle\gamma_{jk}^{t+1}\rangle\phi_1}{\lambda_1}\widehat{\beta}_{jk}^{t+1}$$

and

$$W_{jk}^t = \frac{1}{\sigma_{jt}^2}\sum_{r=1}^{K+1}(\widetilde{\omega}_{rk}^t)^2 + \frac{\langle\gamma_{jk}^t\rangle}{\lambda_1} + \frac{\langle\gamma_{jk}^{t+1}\rangle\phi_1^2}{\lambda_1}.$$

Then from the calculations in Section 5 of Rockova et al. (2020) (equations (5.16)-(5.18)) we obtain the following update for $\widehat{\beta}^{*t}_{jk}$:

$$\beta^{t\star(m+1)}_{jk} = \begin{cases} \frac{1}{W^t_{jk}+(1-\phi_1^2)/\lambda_1 M^t_{jk}}[Z^t_{jk}-\Lambda^t_{jk}]_+\text{sign}(Z^t_{jk}) & \text{for} \quad 1<t<T \\ \frac{1}{\langle\gamma^1_{jk}\rangle\phi_1^2+\langle\gamma^0_{jk}\rangle(1-\phi_1^2)}[\langle\gamma^0_{jk}\rangle\widehat{\beta}^1_{jk}\phi_1-(1-\langle\gamma^0_{jk}\rangle)\lambda_0\lambda_1]_+\text{sign}(\widehat{\beta}^1_{jk}) & \text{for} \quad t=0 \end{cases} \tag{13}$$

where $M^t_{jk}=\langle\gamma^{t+1}_{jk}\rangle(1-\theta^{t+1}_{jk})-(1-\langle\gamma^{t+1}_{jk}\rangle)\theta^{t+1}_{jk}$ and $\Lambda^t_{jk}=\lambda_0[(1-\langle\gamma^t_{jk}\rangle)-M^t_{jk}]$.

Given $\boldsymbol{B}^{\star(m+1)}$, optimizing $Q_1(\cdot)+Q_2(\cdot)$ with respect to $\boldsymbol{\Sigma}_{1:T}$ is done using the Forward Filtering Backward Smoothing algorithm (Ch. 4.3.7 Prado and West, 2010). In order to maximize the posterior log likelihood with respect to $\boldsymbol{\Sigma}_{1:T}$, we first estimate the parameters of the posterior distribution $\pi(\boldsymbol{\Sigma}_{1:T}\,|\,\Omega,\boldsymbol{Y})$, given the updated factor loading matrices $B_{1:T}$, and then calculate the mode of the posterior. Although the exact analytical posterior is unattainable, a fast Gamma approximation exists (Ch. 10.8 West and Harrison, 1997). Appropriate Gamma approximations to the posterior have the form

$$\pi(1/\sigma^2_{j,T-k}\,|\,\Omega,\boldsymbol{Y})=\boldsymbol{G}[\eta_{jT}(-k)/2,d_{jT}(-k)/2],$$

where $d_{jT}(-k)=\eta_{jT}(-k)s_{jT}(-k)$, with

$$s_{jT}(-k)^{-1}=(1-\delta)s^{-1}_{j,T-k}+\delta s_{jT}(-k+1)^{-1}$$

, and filtered degrees of freedom defined by

$$\eta_{jT}(-k)=(1-\delta)\eta_{j,T-k}+\delta\eta_{j,T-k+1},$$

initialized at $\eta_{jT}(0)=\eta_{jT}$. Here $s_{j,T-k}$ denotes $\mathbb{E}(\sigma^2_{j,T-k}\mid\Omega_{T-k},\boldsymbol{Y}_{T-k})$. The details of the algorithm is given in Algorithm 5. In the algorithm we denote the diagonal matrices with diagonal entries $\eta_{j,T-k}$ by $\boldsymbol{\eta}_{T-k}$ and analogously define matrices $\boldsymbol{D}_T(-k)$, $\boldsymbol{S}_{T-k}$ and $S_T(-k)$ for $k=0,1,\ldots,T-1$ so that we can update the parameters of the posterior distribution simultaneously for all j and fixed t. In our study, we set the prior degrees of freedom η_0 to its limit $\eta_0=(1-\delta)^{-1}$ in order to achieve stability and efficiency. Given the parameters of the posterior distribution (the expectation and degrees of freedom), computing the posterior mode is straight forward.

Algorithm: *Forward Filtering Backward Smoothing*	
Input: $\boldsymbol{B}_{1:T}$ and $\boldsymbol{\Sigma}_{1:T}$ from previous iteration Initialize $\boldsymbol{\eta}_0$, $\boldsymbol{D}_0$, $\boldsymbol{S}_0 = \boldsymbol{D}_0\boldsymbol{\eta}_0^{-1}$ Repeat the Forward Step for $t = 1, \dots, T$	
Forward Step where	$\boldsymbol{\eta}_t = \delta\boldsymbol{\eta}_{t-1} + \mathrm{I}$ $\boldsymbol{D}_t = \delta\boldsymbol{D}_{t-1} + \boldsymbol{S}_{t-1}\boldsymbol{E}_t\boldsymbol{E}_t'\boldsymbol{Q}_t^{-1}$ $\boldsymbol{S}_t = \boldsymbol{D}_t\boldsymbol{\eta}_t^{-1}$ $\boldsymbol{E}_t = \boldsymbol{Y}_t - \boldsymbol{B}_t\boldsymbol{\omega}_{t\,\vert\,t-1}$ $\boldsymbol{Q}_t = \boldsymbol{B}_t'\boldsymbol{V}_{t\,\vert\,t-1}\boldsymbol{B}_t + \boldsymbol{\Sigma}_t$
Initialize $\boldsymbol{S}_T(0) = S_T$ Repeat the Backward Step for $k = 1, \dots, T-1$	
Backward Step	$\boldsymbol{\eta}_T(-k) = (1-\delta)\boldsymbol{\eta}_{T-k} + \delta\boldsymbol{\eta}_{T-k+1}$ $\boldsymbol{S}_T(-k)^{-1} = (1-\delta)\boldsymbol{S}_{T-k}^{-1} + \delta\boldsymbol{S}_T(-k+1)^{-1}$ $\boldsymbol{D}_T(-k) = \boldsymbol{\eta}_T(-k)\boldsymbol{S}_T(-k)$ $\boldsymbol{\Upsilon}_{T-k} = (\boldsymbol{\eta}_T(-k) - \mathrm{I})\boldsymbol{D}_T(-k)^{-1}$
Compute Mode	$\boldsymbol{\Sigma}_{T-k} = \boldsymbol{\Upsilon}_{T-k}^{-1}$

Table 5: Forward Filtering Backward Smoothing algorithm for estimating idiosyncratic variances.

Finally, the updates for the covariance matrices $\boldsymbol{A}_{1:T}$, obtained by maximizing $Q_3(\cdot)$, have the following closed form

$$\boldsymbol{A}_t^{(m+1)} = \boldsymbol{M}_{1t} - \boldsymbol{M}_{12t} - \boldsymbol{M}_{12t}' + \boldsymbol{M}_{2t} \quad for \quad t = 1, \dots, T.$$

After completing the expanded M-step in the $(m+1)^{st}$ iteration, we perform a rotation step towards the reduced parameter space to obtain

$$\boldsymbol{B}_t{}^{(m+1)} = \boldsymbol{B}_t^{\star(m+1)}\boldsymbol{A}_{tL}{}^{(m+1)},$$

where $\boldsymbol{A}_t{}^{(m+1)} = \boldsymbol{A}_{tL}^{(m+1)}\boldsymbol{A}_{tL}{}^{(m+1)'}$ is the Cholesky decomposition. These rotated factor loading matrices are carried forward to the next E-step, where we again use the reduced parameter form by keeping $\boldsymbol{A}_t = \boldsymbol{I}_K$.

B Appendix: B

B.1 MCMC on Simulated Example

In this section, we apply our MCMC estimation procedure on a simulated example (using the lower-triangular identifiability constraint). First, we generate a single dataset with $P = 10$ responses, $K = 3$ candidate latent factors, and $T = 100$ time series observations.

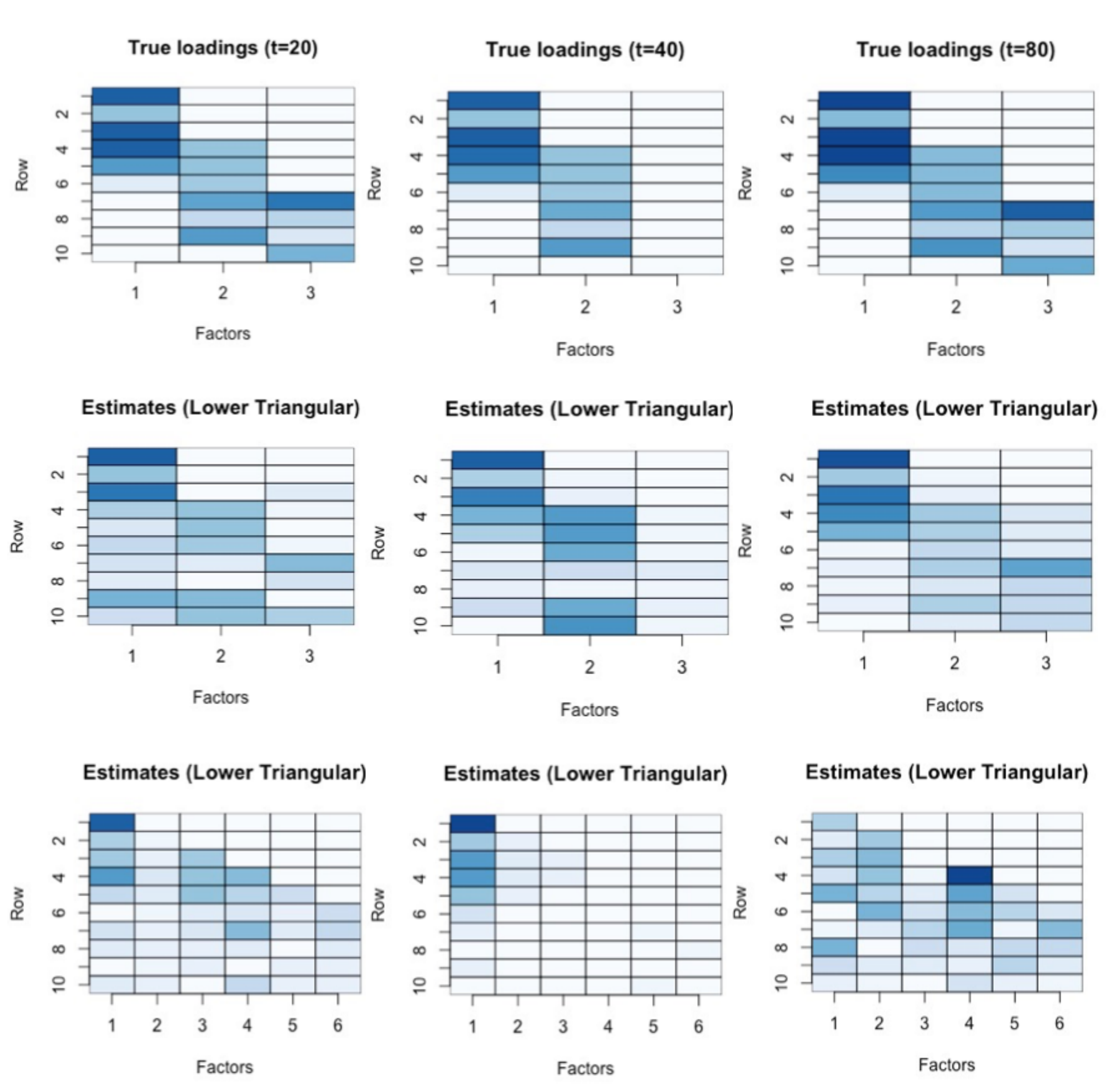


Figure B1: Simulated Example: Absolute Value of factor loading matrices estimated by MCMC (with 2000 draws after discarding 8000 burn-in samples) for a simulated data with $P = 10$ and $T = 100$. True number of nonzero factors are 3, 2 and 3 for $t = 20$, $t = 40$ and $t = 80$ respectively. (a) First row shows true factor loading matrices. (b) Second row shows estimated loading matrices with $K = 3$. (c) Third row shows estimated factor loading matrices with $K = 6$.

Algorithm: *MCMC algorithm for DSS with a Gaussian spike*	
	Initialize $(\boldsymbol{B}_{0:T}, \boldsymbol{\Sigma}_{0:T})$ and choose n_0, d_0.
	Sampling Latent Factors
Kalman Filter and Smoother	Compute $\boldsymbol{\omega}_{t\,\vert\,T}, \boldsymbol{V}_{t\,\vert\,T}$ and $\boldsymbol{V}_{t,t-1\,\vert\,T}$ for $1 \leq t \leq T$
	Sample $\boldsymbol{\omega}_t \sim \mathcal{N}_K(\boldsymbol{\omega}_{t\,\vert\,T}, \boldsymbol{V}_{t\,\vert\,T})$
	Sampling Factor Loadings
Forward filtering	For $t = 1, \dots, T$ and $j = 1, \dots, P$,
	Compute $\boldsymbol{a}_j^t = \boldsymbol{H}_j^t + \boldsymbol{\Gamma}_{j.}^t(\boldsymbol{m}_j^{t-1} - \boldsymbol{H}_j^t)$.
	Compute $\boldsymbol{R}_j^t = \boldsymbol{\Gamma}_j^t \boldsymbol{C}_j^{t-1} \boldsymbol{\Gamma}_{j.}^{t}{}' + \boldsymbol{W}_j^t$.
	Compute $f_j^t = \boldsymbol{\omega}_{t\,\vert\,T}' \boldsymbol{a}_j^t$.
	Compute $q_j^t = \boldsymbol{\omega}_{t\,\vert\,T}' \boldsymbol{R}_j^t \boldsymbol{\omega}_{t\,\vert\,T} + \sigma_{jt}^2$ and $e_j^t = y_j^t - f_j^t$.
	Compute $\boldsymbol{m}_j^t = \boldsymbol{a}_j^t + \boldsymbol{A}_j^t e_j^t$ and $\boldsymbol{C}_j^t = \boldsymbol{R}_j^t - \boldsymbol{A}_j^t \boldsymbol{A}_j^{t\prime} q_j^t$ with $\boldsymbol{A}_j^t = \boldsymbol{R}_j^t \boldsymbol{\omega}_{t\,\vert\,T}/q_j^t$.
Backward sampling	Simulate $\boldsymbol{B}_{j.}^T \sim \mathcal{N}(\boldsymbol{m}_j^T, \boldsymbol{C}_j^T)$.
	For $t = T-1, \dots, 0$ and $j = 1, \dots, P$
	Compute $\boldsymbol{a}_j^T(t-T) = \boldsymbol{m}_j^t + \boldsymbol{L}_j^t[\boldsymbol{B}_{j.}^{(t+1)\star} - \boldsymbol{a}_j^{t+1}]$.
	Compute $\boldsymbol{R}_j^T(t-T) = \boldsymbol{C}_j^t - \boldsymbol{L}_j^t \boldsymbol{R}_j^{t+1} \boldsymbol{L}_j^{t\prime}$, where $\boldsymbol{L}_j^t = \boldsymbol{C}_j^t \boldsymbol{\Gamma}_{j.}^{t+1\prime} \boldsymbol{R}_j^{t+1-1}$.
	Simulate $\boldsymbol{B}_{j.}^t \sim \mathcal{N}(\boldsymbol{a}_j^T(t-T), R_j^T(t-T))$.
	Sampling Indicators
	For $j = 1, \dots, p$ and $k = 1, \dots, K$
	Compute $\theta_{jk}^t = \theta(\beta_{jk}^{t-1})$ for $1 \leq t \leq T$ from (6).
	Compute $p_{jk}^{\star}{}^t = p_{jk}^{\star}{}^t(\beta_{jk}^t)$ for $1 \leq t \leq T$ from (8).
	Compute $p_{jk}^{\star}{}^0 = \theta(\beta_{jk}^0)$ from (6).
	Sample $\gamma_{jk}^t \sim \text{Bernoulli}[p_{jk}^{\star}{}^t(\beta_{jk}^t)]$ for $0 \leq t \leq T$.
	Sampling Precisions $\nu_j^t = 1/(\sigma_j^t)^2$
	For $t = 1, \dots, T$ and $j = 1, \dots, P$
Forward filtering	Compute $n_j^t = \delta n_j^{t-1} + 1$ and $d_j^t = \delta d_j^{t-1} + (r_j^t)^2$, where $r_j^t = y_j^t - \boldsymbol{\omega}_{t\,\vert\,T}' \boldsymbol{\beta}_j^t$.
Backward sampling	Sample $\nu_j^T \sim G(n_j^T/2, d_j^T/2)$.
	For $t = 1, \dots, T$
	Sample $\eta_j^{T-t} \sim G[(1-\delta)n_j^{T-t}/2, d_j^{T-t}/2]$.
	Set $1/(\sigma_j^{T-t})^2 = \eta_j^{T-t} + \delta/(\sigma_j^{T-t+1})^2$.

Table B1: An MCMC algorithm with DSS priors and a Gaussian spike. Note that $G(a, b)$ denotes a gamma distribution with a mean a/b.

$$\boldsymbol{H}_j^t = \phi_0 \boldsymbol{\Gamma}_j^{t}{}', \quad \boldsymbol{m}_j^0 = \phi_0 \gamma_j^0, \quad \boldsymbol{W}_j^t = diag\{\boldsymbol{\Gamma}_j^t \lambda_1 + (1 - \boldsymbol{\Gamma}_j^t)\lambda_0\}$$

We choose the dimensionality of this example to be much smaller than our in EM implementation in Section 4 because the computational times of our MCMC sampling procedure

Method	Maximum No. of Factors	Elapsed Time (sec)
MCMC	3	2439.3
MCMC	4	2526.9
MCMC	6	2852.2
EM	6	302.5

Table B2: Elapsed computation time for MCMC (10000 draws) and EM algorithm

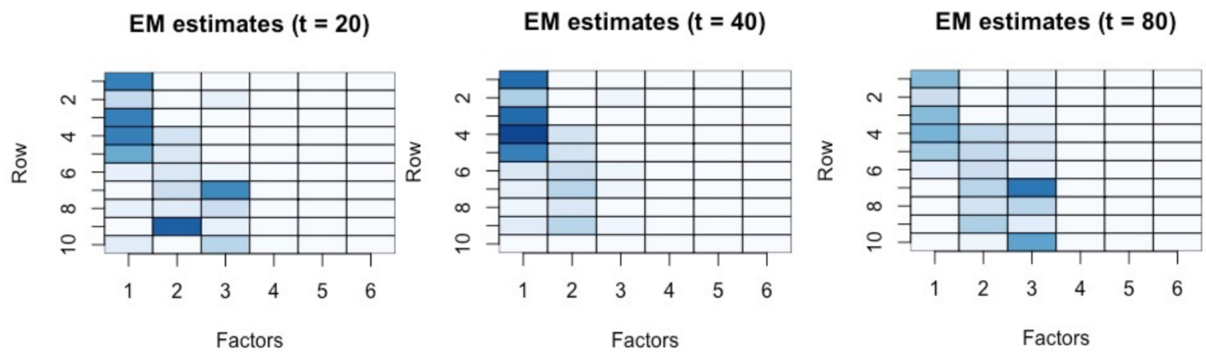


Figure B2: Simulated Example: Absolute Value of factor loading matrices estimated by EM algorithm with $K = 6$.

are less favorable compared to our EM approach. Table B2 shows a comparison between computation times for MCMC and EM for this example.

The elements of latent factors $\boldsymbol{\Omega}_t$ and idiosyncratic errors $\boldsymbol{\epsilon}_t$ are generated from a standard Gaussian distribution. We now describe the true loading matrices $\boldsymbol{B}^0 = [\boldsymbol{B}_1^0, \ldots, \boldsymbol{B}_T^0]$, which were used to generate the data, where $\boldsymbol{B}_t^0 = \{\beta_{jk}^{0t}\} \in \mathbb{R}^{P\times K}$. At time $t = 1$, the active latent factor loadings form a block diagonal structure with 5, 6 and 4 active loadings for factor 1, factor 2 and factor 3 respectively, of which factor 1 and factor 2 overlap for 2 series, while factor 2 and factor 3 overlap for 3 series. Factor 1 and factor 3 never overlap.

The sparsity pattern changes structurally over time where (a) at time $t = 35$ the loadings of the third factor become inactive, (b) at $t = 45$ the loadings of the third factor are re-introduced and active until $T = 100$. The true nonzero loadings are smooth and arrive from an autoregressive process, i.e. $\beta_{jk}^{0t} = \phi\beta_{jk}^{0t-1} + v_{jk}^t$ with $v_{jk}^t \overset{iid}{\sim} \mathcal{N}(0, 0.0025)$ for

$\phi = 0.99$, initiated at $\beta_{jk}^{01} \overset{iid}{\sim} \mathcal{N}(2, 1)$ for all $1 \leq j \leq P$ and $1 \leq k \leq 5$. When loadings β_{jk}^{0t} become inactive, they are thresholded to zero. The true factor loadings are thereby smooth until they suddenly drop out and can re-emerge.

Figure B2 shows a heatmap of the (absolute values of) true and estimated factor loadings. Row 1 shows the true factor loading matrices for times $t = 20$, $t = 40$ and $t = 80$ respectively. Row 2 and 3 show estimated factor loadings when upper limit of latent factors are set to $K = 3$ and $K = 6$ respectively. Posterior mean estimates of factor loadings are obtained from an MCMC sample of size 2000, after discarding 8000 burn-in samples. We see that for $K = 6$, MCMC is not able to identify actual number of factors due to factor splitting. However estimated loading structures are close to the true loadings for $K = 3$.

	Output and Income		Labor Market		Consumption and Orders		Orders and Inventories		Money and Credit		Interest rate and Exchange Rates		Prices		Stock Market
1	RPI	17	HWI	48	HOUST	58	DPCERA3M086SBEA	67	M1SL	81	FEDFUNDS	103	WPSFD49207	123	S&P 500
2	W875RX1	18	HWIURATIO	49	HOUSTNE	59	CMRMTSPLx	68	M2SL	82	CP3Mx	104	WPSFD49502	124	S&P: indust
3	INDPRO	19	CLF16OV	50	HOUSTMW	60	RETAILx	69	M2REAL	83	TB3MS	105	WPSID61	125	S&P div yield
4	IPFPNSS	20	CE16OV	51	HOUSTS	61	AMDMNOx	70	AMBSL	84	TB6MS	106	WPSID62	126	S&P PE ratio
5	IPFINAL	21	UNRATE	52	HOUSTW	62	ANDENOx	71	TOTRESNS	85	GS1	107	OILPRICEx	127	VXOCLSx
6	IPCONGD	22	UEMPMEAN	53	PERMIT	63	AMDMUOx	72	NONBORRES	86	GS5	108	PPICMM		
7	IPDCONGD	23	UEMPLT5	54	PERMITNE	64	BUSINVx	73	BUSLOANS	87	GS10	109	CPIAUCSL		
8	IPNCONGD	24	UEMP5TO14	55	PERMITMW	65	ISRATIOx	74	REALLN	88	AAA	110	CPIAPPSL		
9	IPBUSEQ	25	UEMP15OV	56	PERMITS	66	UMCSENTx	75	NONREVSL	89	BAA	111	CPITRNSL		
10	IPMAT	26	UEMP15T26	57	PERMITW			76	CONSPI	90	COMPAPFFx	112	CPIMEDSL		
11	IPDMAT	27	UEMP27OV					77	MZMSL	91	TB3SMFFM	113	CUSR0000SAC		
12	IPNMAT	28	CLAIMSx					78	DTCOLNVHFNM	92	TB6SMFFM	114	CUSR0000SAD		
13	IPMANSICS	29	PAYEMS					79	DTCTHFNM	93	T1YFFM	115	CUSR0000SAS		
14	IPB51222S	30	USGOOD					80	INVEST	94	T5YFFM	116	CPIULFSL		
15	IPFUELS	31	CES1021000001							95	T10YFFM	117	CUSR0000SA0L2		
16	CUMFNS	32	USCONS							96	AAAFFM	118	CUSR0000SA0L5		
		33	MANEMP							97	BAAFFM	119	PCEPI		
		34	DMANEMP							98	TWEXMMTH	120	DDURRG3M086SBEA		
		35	NDMANEMP							99	EXSZUSx	121	DNDGRG3M086SBEA		
		36	SRVPRD							100	EXJPUSx	122	DSERRG3M086SBEA		
		37	USTPU							101	EXUSUKx				
		38	USWTRADE							102	EXCAUSx				
		39	USTRADE												
		40	USFIRE												
		41	USGOVT												
		42	CES0600000007												
		43	AWOTMAN												
		44	AWHMAN												
		45	CES0600000008												
		46	CES2000000008												
		47	CES3000000008												

Table B3: Macroeconomic Study: The list of economic variables used in the study.