Dynamic Sparse Factor Analysis

Veronika Ročková

Joint work with
Ken McAlinn and Enakshi Saha

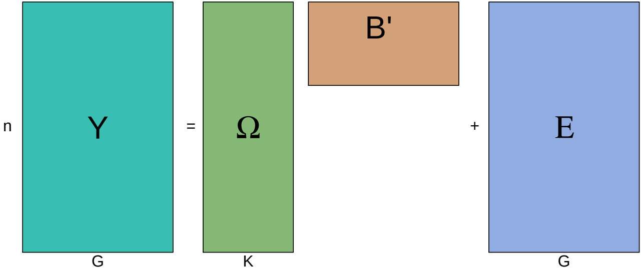
December 19^{th}, 2018

Vienna University of Economics and Business

Sparse Factor Analysis Revisited

Generic factor model for **fixed number** K of latent factors:

$$\boldsymbol{Y}_i \mid \boldsymbol{\omega}_i, \boldsymbol{B}, \Sigma \overset{\text{ind}}{\sim} \mathcal{N}_G\left(\boldsymbol{B}\boldsymbol{\omega}_i, \Sigma\right), \quad 1 \leq i \leq n, \tag{1}$$

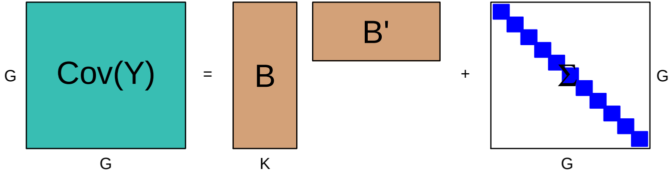


- ⇝ $\boldsymbol{E} = [\epsilon_1, \ldots, \epsilon_n]'$ with $\epsilon_i \overset{\text{ind}}{\sim} \mathcal{N}_G(\mathbf{0}, \Sigma)$, $\Sigma = \text{diag}\{\sigma_j^2\}_{j=1}^G$
- ⇝ $\boldsymbol{\Omega} = [\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n]'$: latent factors
- ⇝ $\boldsymbol{B} = (\beta_{jk})_{j,k=1}^{G,K}$: factor loadings

Sparse Factor Analysis Revisited

When $\omega_i \sim \mathcal{N}_K(\mathbf{0}, \mathrm{I}_K)$, marginally

$$f(\boldsymbol{y}_i \mid \boldsymbol{B}, \boldsymbol{\Sigma}) = \mathcal{N}_G(\mathbf{0}, \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{\Sigma}),\ 1 \leq i \leq n. \quad (2)$$



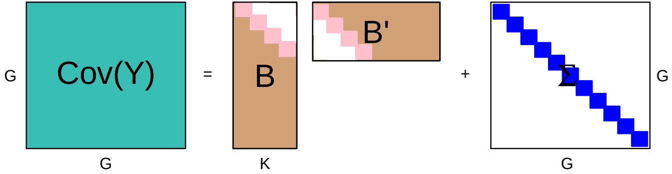
- ☹ Because $\boldsymbol{B}\boldsymbol{B}' = (\boldsymbol{B}\boldsymbol{P})(\boldsymbol{B}\boldsymbol{P})'$, for any orthogonal matrix $\boldsymbol{P}$, likelihood (3) is invariant under factor rotation.
- ☹ Identifiability constraints render responses non-exchangeable.
- ☹ Effective factor cardinality K unknown

Approach *A prior on infinite-dimensional loading matrices, which anchors interpretable factor orientations*

Sparse Factor Analysis Revisited

When $\omega_i \sim \mathcal{N}_K(\mathbf{0}, \mathrm{I}_K)$, marginally

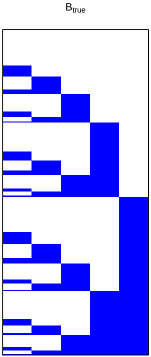
$$f(\boldsymbol{y}_i \mid \boldsymbol{B}, \boldsymbol{\Sigma}) = \mathcal{N}_G(\mathbf{0}, \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{\Sigma}), \; 1 \leq i \leq n. \tag{3}$$



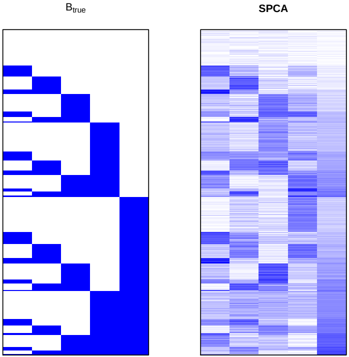
- ☹ Because $\boldsymbol{B}\boldsymbol{B}' = (\boldsymbol{B}\boldsymbol{P})(\boldsymbol{B}\boldsymbol{P})'$, for any orthogonal matrix $\boldsymbol{P}$, likelihood (3) is invariant under factor rotation.
- ☹ Identifiability constraints render responses non-exchangeable.
- ☹ Effective factor cardinality K unknown

Approach *A prior on infinite-dimensional loading matrices, which anchors interpretable factor orientations*

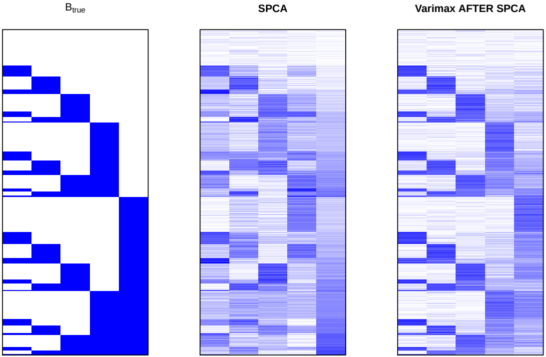
Sparsity Priors and Rotations: Motivation



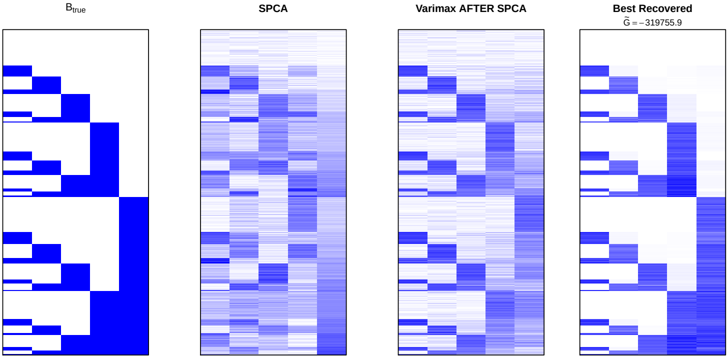
Sparsity Priors and Rotations: Motivation



Sparsity Priors and Rotations: Motivation



Sparsity Priors and Rotations: Motivation



Our Hierarchical Approach

1. Elements of our prior distribution

Prior on the loading matrix $\boldsymbol{B}_{G\times\infty}$

(1a) *Spike-and-Slab LASSO Prior*

(1b) *Indian Buffet Process Prior*

Prior on the residual variances $\boldsymbol{\Sigma} = \{\sigma_j^2\}_{j=1}^G$

Independent Inverse Gamma priors IG$(\eta/2, \eta\nu/2)$

2. Fast Bayesian computation

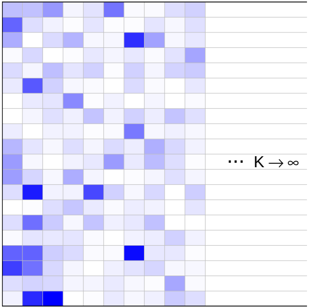
The EM algorithm

Rotations to sparsity with parameter expansion

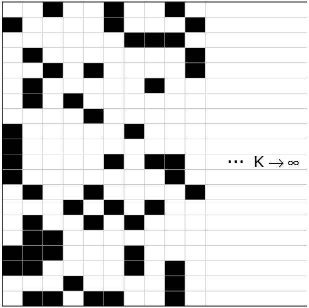
Elements of the Hierarchical Prior

The matrix $\mathbf{\Gamma} = \{\gamma_{jk}\}_{j,k=1}^{G,\infty}$ includes *binary allocation indicators* that characterize which features are associated with each response.

$\pi(\boldsymbol{B}|\mathbf{\Gamma})$



$\pi(\mathbf{\Gamma}|\boldsymbol{\theta})$



The Spike-and-Slab LASSO Prior

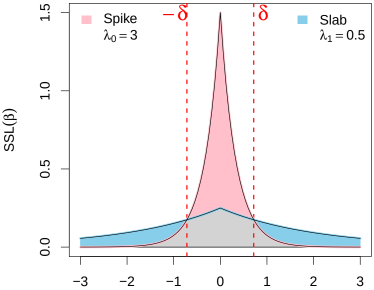
A mixture refinement of the LASSO (Laplace) prior with a mixing binary indicator $\gamma \in \{0, 1\}$

$$\pi(\beta \mid \gamma) = \gamma \mathit{Laplace}(\beta \mid \lambda_1) + (1 - \gamma)\mathit{Laplace}(\beta \mid \lambda_0)$$

λ_1 **small**: to avoid over-shrinkage of large effects

λ_0 **large**: to shrink ignorable coefficients to zero

θ Controls the sparsity, where $\mathrm{P}(\gamma = 1 \mid \theta) = \theta$



Point-mass spike-and-slab is a limiting case as $\lambda_0 \to \infty$

The Penalized Likelihood Perspective

- Conditionally on θ, the prior is an independent product
- Define by $\widehat{\boldsymbol{\beta}}$ the MAP estimator

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^n} \left[-\frac{1}{2}\sum_{i=1}^{n}(y_i - \beta_i)^2 + \sum_{i=1}^{n} pen_\theta(\beta_i) \right], \qquad (4)$$

with the separable Spike-and-Slab LASSO (SSL) penalty

$$pen_\theta(\beta_i) = \log\left[\theta\, Laplace(\beta_i \mid \lambda_1) + (1-\theta) Laplace(\beta_i \mid \lambda_0)\right]$$

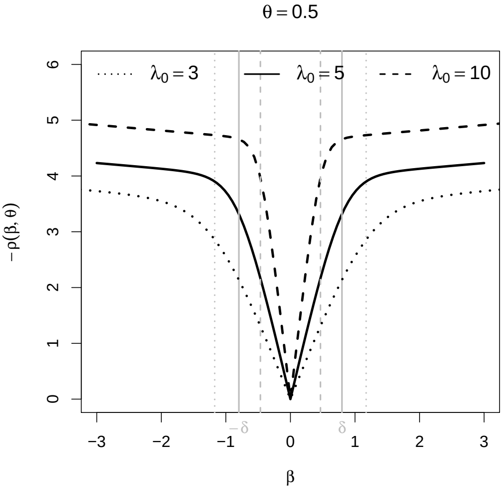
⇝ Denote by

$$p^\star_\theta(\beta_i) = \frac{\theta Laplace(\beta_i \mid \lambda_1)}{\theta Laplace(\beta_i \mid \lambda_1) + (1-\theta) Laplace(\beta_i \mid \lambda_0)}$$

a conditional inclusion probability $\mathrm{P}(\gamma_i = 1 \mid \beta_i)$.

The Spike-and-Slab LASSO (SSL) Penalty

- The SSL penalty is a smooth mix of two LASSO penalties



The Spike-and-Slab LASSO Shrinkage

The derivative of the penalty determines the amount of shrinkage

$$\frac{\partial pen_\theta(\beta_i)}{\partial|\beta_i|} = -\lambda^\star_\theta(\beta_i)$$

where

$$\lambda^\star_\theta(\beta_i) = p^\star_\theta(\beta_i)\lambda_1 + [1 - p^\star_\theta(\beta_i)]\lambda_0$$

- *The Spike-and-Slab LASSO mode satisfies*

$$\widehat{\beta}_i = \left(|y_i| - \lambda^\star_\theta(\widehat{\beta}_i)\right)_+ \mathrm{sign}(y_i) \quad (5)$$

☺ “Self-adaptive" property of the shrinkage term

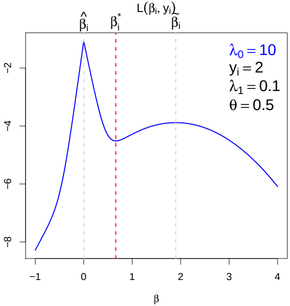
- *The LASSO mode satisfies*

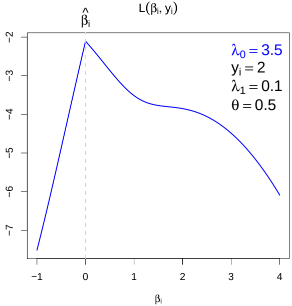
$$\widehat{\beta}_i = (|y_i| - \lambda)_+ \mathrm{sign}(y_i)$$

☹ Constant penalty regardless of the size of $|y_i|$

“Local/Global" Mode Considerations

- The *SSL* log-posterior can be multi-modal





- The log-posterior will be concave and therefore uni-modal if

$$(\lambda_0 - \lambda_1)^2 < 4$$

☹ We are interested in priors that are en-route to the point-mass mixture prior when $\lambda_0 \to \infty$

The condition (5) not sufficient to characterize the global mode.

Refined Characterization of the Global Mode

The *SSL* global mode is a thresholding rule and satisfies

$$\widehat{\beta}_j = \begin{cases} 0 & \text{when} \quad |y_j| \leq \Delta \\ [|y_j| - \lambda^\star_\theta(\widehat{\beta}_j)]_+ \text{sign}(y_j) & \text{when} \quad |y_j| > \Delta. \end{cases}$$

where

$$\Delta \sim \sqrt{2\log[1/p^\star_\theta(0)]} + \lambda_1$$

The threshold Δ depends on $(\lambda_0, \lambda_1, \theta)$ through

$$\log[1/p^\star_\theta(0)] = \log\left[\frac{1-\theta}{\theta}\frac{\lambda_0}{\lambda_1} + 1\right]$$

- $\widehat{\beta}$ is a blend of hard and soft thresholding
- The selection threshold Δ drives the properties of the mode

The Spike-and-Slab LASSO posterior keeps pace with the global mode!

The EM Approach to Sparse FA

Goal *Find $(\boldsymbol{B}, \boldsymbol{\Sigma}, \theta)$ which is the most likely (a posteriori) to have generated the data.*

parameters of interest: $\boldsymbol{B}, \boldsymbol{\Sigma}$ and θ

latent variables: $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$

Chicken If $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$ were known, $\boldsymbol{B}, \boldsymbol{\Sigma}$ and θ could be easily estimated.

Egg $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$ cannot be inferred unless $\boldsymbol{B}, \boldsymbol{\Sigma}$ and θ is known.

☺ Solution: EM algorithm of Dempster, Laird and Rubin (1977)

(E-step) Expectation of the latent data given the current parameters

(M-step) Finding the most likely parameters given the expected missing data.

The EM Algorithm for Factor Analysis

The EM algorithm locates posterior modes of

$$\log \pi(\boldsymbol{B}, \Sigma, \boldsymbol{\theta} \mid \boldsymbol{Y})$$

iteratively by *maximizing the expected augmented log posterior*:

$$Q(\boldsymbol{B}, \boldsymbol{\theta}, \Sigma) = \mathsf{E}_{\boldsymbol{\Gamma}, \boldsymbol{\Omega} \mid \cdot} \left[\log \pi \left(\underbrace{\boldsymbol{B}, \Sigma, \boldsymbol{\theta}}_{\text{unknown parameters}}, \overbrace{\boldsymbol{\Gamma}, \boldsymbol{\Omega}}^{\text{missing data}} \mid \underbrace{\boldsymbol{Y}}_{\text{observed data}} \right) \right]$$

- $\mathsf{E}_{\boldsymbol{\Gamma}, \boldsymbol{\Omega} \mid \cdot}(\cdot)$ denotes the conditional expectation given the observed data and current parameter estimates at the m-th iteration,
- Dimension of $\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}$ determined by $K^{\star}$, the order of the truncated stick-breaking approximation.

The E-Step

Using current parameters $(\boldsymbol{B}, \boldsymbol{\Sigma}, \boldsymbol{\theta}) = (\boldsymbol{B}^{(m)}, \boldsymbol{\Sigma}^{(m)}, \boldsymbol{\theta}^{(m)})$ at m-th iteration

- $\boldsymbol{\Omega}$ *Featurization step*: rows of the new features are solutions to ridge regression of $\boldsymbol{Y}\boldsymbol{\Sigma}^{-1/2}$ on the rows of $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{B}$:

$$\mathrm{E}_{\boldsymbol{\Omega}|\cdot}[\boldsymbol{\Omega}'] = \left(\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}\boldsymbol{B} + \mathrm{I}_{K^\star}\right)^{-1} \boldsymbol{B}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'$$

  Smoothness penalty matrix:

$$\mathrm{Cov}_{\boldsymbol{\Omega}|\cdot}[\omega_i] = \left(\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}\boldsymbol{B} + \mathrm{I}_{K^\star}\right)^{-1}$$

- $\boldsymbol{\Gamma}$ *Variable selection indicators*
  Mixing proportions when fitting a Laplace mixture

$$\mathrm{P}[\gamma_{jk} = 1|\beta_{jk}] = \frac{Laplace(\beta_{jk} \mid \lambda_1)\theta_k}{Laplace(\beta_{jk} \mid \lambda_1)\theta_k + Laplace(\beta_{jk} \mid \lambda_0)(1-\theta_k)},$$

  Adaptive weights determining the amount of penalization

$$p^\star_{jk} \equiv \mathrm{P}[\gamma_{jk} = 1|\beta_{jk}, \theta_k]$$

The M-Step

$\beta^{(m+1)}$ *"Adaptive" LASSO* computation:

- Denote by $\boldsymbol{y}^{j\star} = \begin{pmatrix} \boldsymbol{y}^j \\ \boldsymbol{0}_{K^\star} \end{pmatrix}$, $\boldsymbol{\Omega}^\star = \begin{pmatrix} \mathrm{E}_{\Omega|\cdot}[\Omega] \\ \boldsymbol{L}' \end{pmatrix}$, where $\mathrm{Cov}_{\Omega|\cdot}[\boldsymbol{\omega}_i] = \boldsymbol{L}\boldsymbol{L}'$.
- The j-th row of $\boldsymbol{\beta}_j$ updated as follows:

$$\boldsymbol{\beta}_j^{(m+1)} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{K\star}} \left\{ -\frac{||\boldsymbol{y}_j^\star - \boldsymbol{\Omega}^\star \boldsymbol{\beta}_j||^2}{2\sigma_j^{(m)2}} - \sum_{j=1}^{K\star} \lambda_{jk}^\star |\beta_{jk}| \right\},$$

where $$\lambda_{jk}^\star = p_{jk}^\star \lambda_1 + (1 - p_{jk}^\star)\lambda_0$$

$\sigma_j^{(m+1)}$ *Easy update* conditionally on $\boldsymbol{B}^{(m+1)}$.

$\boldsymbol{\theta}^{(m+1)}$ *Linear program*

$$\arg\max_{\boldsymbol{\theta}} \left\{ \sum_{j=1}^{G} \sum_{k=1}^{K^\star} \left[p_{jk}^\star \log \theta_k + (1 - p_{jk}^\star) \log(1 - \theta_k) \right] + (\alpha - 1) \log \theta_{K^\star} \right\},$$

subject to $$\theta_k - \theta_{k-1} \leq 0, \quad 0 \leq \theta_k \leq 1.$$

Rotational Ambiguity and Parameter Expansion

Local convergence issue exacerbated by strong coupling between Ω and $\boldsymbol{B}$

Powerful accelerations obtained with parameter expansion

PXL-EM **P**arameter e**X**pansion of the **L**ikelihood:

$$f(\boldsymbol{y}_i \mid \omega_i, \boldsymbol{B}, \boldsymbol{A}, \boldsymbol{\Sigma}) \overset{\text{ind}}{\sim} \mathcal{N}_G(\boldsymbol{B}\boldsymbol{A}_L^{-1}\omega_i, \boldsymbol{\Sigma}), \quad 1 \leq i \leq n, \tag{6}$$

where $\boldsymbol{A}_L$ is a lower Cholesky factor of $\boldsymbol{A}$ and

$$\omega_i \sim \mathcal{N}_K(\mathbf{0}, \boldsymbol{A}). \tag{7}$$

- ⇝ For each $\boldsymbol{A}$, we put the SSL prior on $\boldsymbol{B}^* = \boldsymbol{B}\boldsymbol{A}_L^{-1}$!!!
- ⇝ Original model recovered at $\boldsymbol{A}_0 = \mathrm{I}_K$.
- ⇝ The prior serves to identify sparse orientations!

The PXL-EM Algorithm

PXL-EM traverses the expanded parameter space, yielding

$$(\boldsymbol{\Sigma}^{(1)}, \boldsymbol{\theta}, \underbrace{\boldsymbol{B}^{\star(1)}, \boldsymbol{A}^{(1)}}_{\boldsymbol{B}^{(1)}}) \quad , \quad (\boldsymbol{\Sigma}^{(2)}, \boldsymbol{\theta}, \underbrace{\boldsymbol{B}^{\star(2)}, \boldsymbol{A}^{(2)}}_{\boldsymbol{B}^{(2)}}) \quad , \quad \ldots$$

which maps onto a trajectory in the original space via

$$\boldsymbol{B}^{(k)} = \boldsymbol{B}^{\star(k)} \boldsymbol{A}_L^{(k)} \tag{8}$$

E-step Operates in the reduced space, conditional on $(\boldsymbol{B}^{(k)}, \boldsymbol{A}_0)$

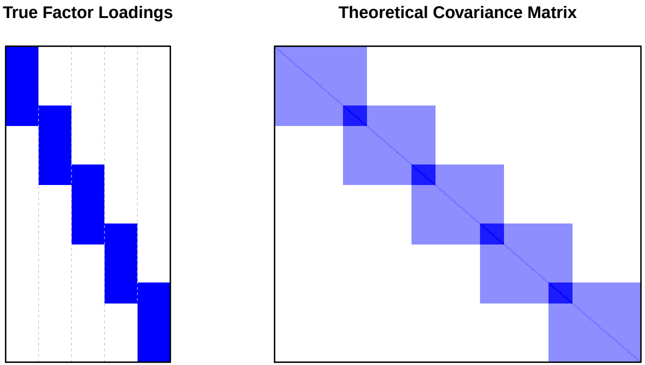
M-step Operates in the expanded space, yielding sparse $\boldsymbol{B}^{\star(k+1)}$ and

$$\boldsymbol{A}^{(k+1)} = \frac{1}{n}\langle \boldsymbol{\Omega}' \boldsymbol{\Omega} \rangle = \frac{1}{n}\langle \boldsymbol{\Omega} \rangle' \langle \boldsymbol{\Omega} \rangle + \boldsymbol{M}^{(k)}. \tag{9}$$

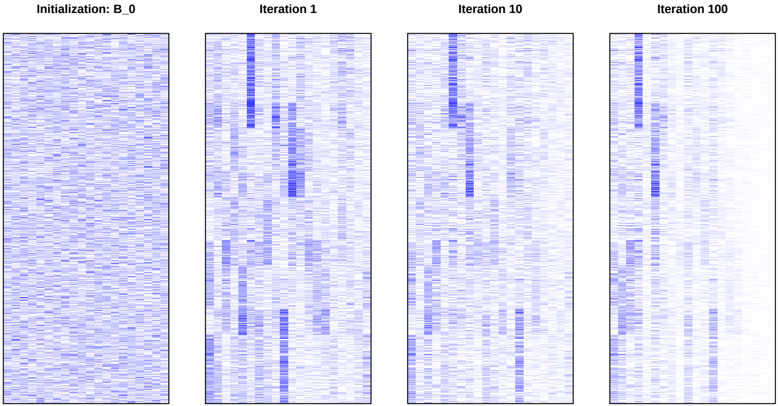
- ⇝ Upon convergence, $\frac{1}{n}\langle \boldsymbol{\Omega}' \boldsymbol{\Omega} \rangle = \boldsymbol{A} = \boldsymbol{A}_0 = \mathrm{I}$
- ⇝ Rotation to sparsity!
- ⇝ PXL-EM converges at least as fast as EM!

EM vs PXL-EM: Synthetic Data

- $n = 100$ observations generated from (1) with $G = 2\,000$ responses and $K_{true} = 5$ factors.
- $\boldsymbol{B}_{true}$ is block-diagonal with nonzero elements equal to 1
- Initialization: $\boldsymbol{B}^{(0)} \sim \mathcal{MVN}(\mathbf{0}, \mathrm{I}_G, \mathrm{I}_K^\star)$, $\boldsymbol{\Sigma}^{(0)} = \mathrm{I}_G$.
- We set $\lambda_1 = 0.001, \lambda_0 = 5, \alpha = 0.1$ and $K^\star = 20$.

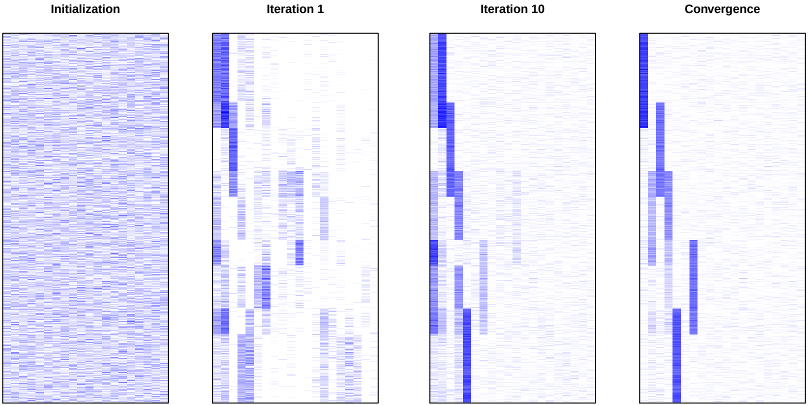


EM Trajectory



$$\lambda_1 = 0.001, \lambda_0 = 5$$

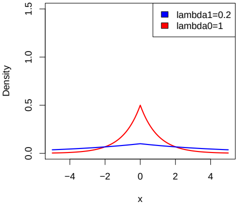
PXL-EM Trajectory



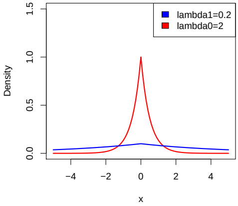
$$\lambda_1 = 0.001, \lambda_0 = 5$$

Dynamic Posterior Exploration

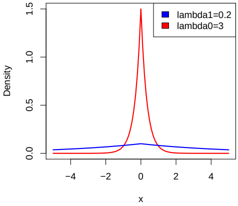
- With large differences $(\lambda_0 - \lambda_1)$, the posterior is very spiky
- We consider a *sequence of mixture priors* and compute a *solution path* indexed by λ_0 with warm starts



$\lambda_0 = 1$

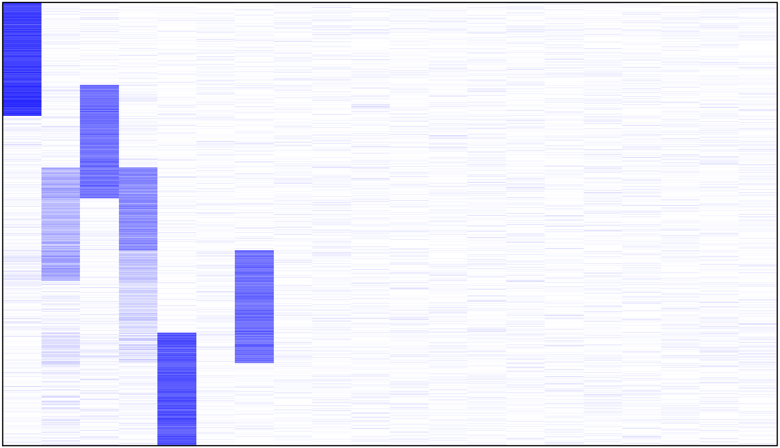


$\lambda_0 = 2$



$\lambda_0 = 3$

Dynamic Posterior Exploration in Action

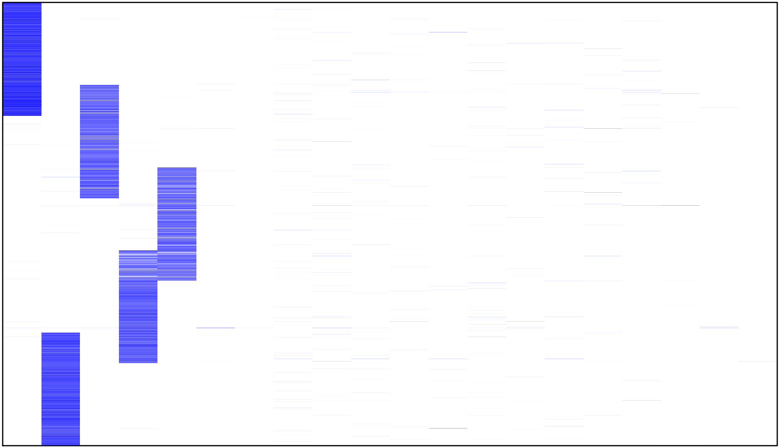


$$\lambda_1 = 0.001, \lambda_0 = 5$$

Dynamic Posterior Exploration in Action

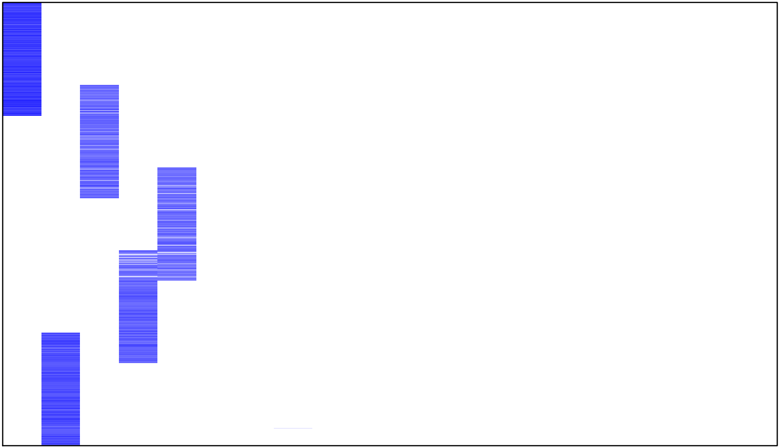
$$\lambda_1 = 0.001, \lambda_0 = 10$$

Dynamic Posterior Exploration in Action



$$\lambda_1 = 0.001, \lambda_0 = 15$$

Dynamic Posterior Exploration in Action



$$\lambda_1 = 0.001, \lambda_0 = 20$$

Dynamic Factor Analysis

Dynamic Factor Analysis

High-dimensional multivariate time series $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T] \in \mathbb{R}^{P \times T}$.

Evolving covariance patterns over time can be captured with the following *state space model*:

$$\boldsymbol{Y}_t = \boldsymbol{B}_t \boldsymbol{\omega}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{\text{ind}}{\sim} \mathcal{N}_P(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{10}$$

$$\boldsymbol{\omega}_t = \boldsymbol{\Phi} \boldsymbol{\omega}_{t-1} + \mathbf{e}_t, \quad \mathbf{e}_t \overset{\text{ind}}{\sim} \mathcal{N}_K(\mathbf{0}, \sigma_\omega^2 \mathbb{I}_K). \tag{11}$$

Stochastic volatility: $\boldsymbol{\Sigma}_t = \mathrm{diag}\{\sigma_{jt}^2\}_{j=1}^P$

$$\sigma_{jt} = \sigma_{jt-1} \delta / \upsilon_{jt},$$

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 $\eta_t = \delta \eta_{t-1} + 1$.

Parameters $\boldsymbol{\Phi} = \phi \mathrm{I}$ and σ_ω^2 are treated as known.

Related procedures: Kaufmann and Schumacher (2013), Del Negro and Otrok (2008), Nakajima and West (2016), Kastner et al. (2017)

Dynamic Spike-and-Slab Processes

Dynamic Linear Model

A *scalar* response y_t at time t is related to a vector of *known regressors* $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tp})'$ through

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t^0 + \varepsilon_t, \quad t = 1, \ldots, T, \tag{12}$$

where

- $\boldsymbol{\beta}_t^0 = (\beta_{t1}^0, \ldots, \beta_{tp}^0)'$ is a *time-varying vector* of regression coefficients
- $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ is an innovation term at time t

Motivation

By obscuring variable selection uncertainty over time, **confining to a single inferential model** *may lead to poorer predictive performance, especially when the effective subset at each time is* **sparse**.

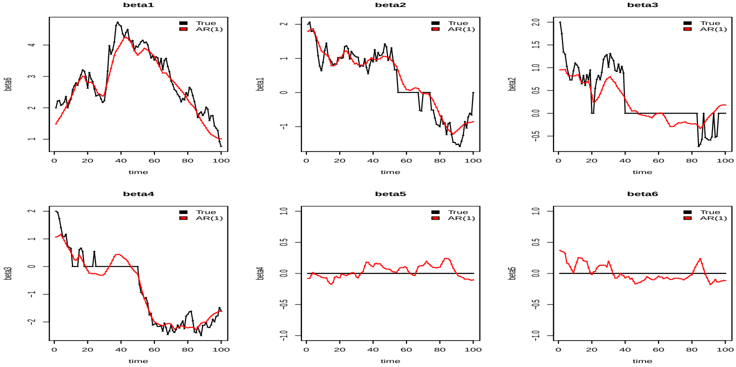
AR(1) does not capture intermittent zeroes...

Suppose that the true coefficients came from an AR(1) process

$$\beta^0_{tj} = \phi_1 \beta^0_{t-1j} + \nu_{tj}, \quad \phi_1 = 0.98, \quad \nu_{tj} \overset{\text{iid}}{\sim} \mathcal{N}[0, 10(1 - \phi_1^2)]$$

and were thresholded to *zero* if $|\beta^0_{tj}| < 0.5$.

Assume $T = 100$ and $p = 6$ and obtain y_t from (12).



Ingredients for Dynamic Variable Selection

We design dynamic priors $\pi(\{\beta_{jt}\})$ that are able to capture

(a) *Vertical sparsity* (in $\{\beta_{jt}\}_{j=1}^{p}$) : only a small portion of coefficients at time t is nonzero

(b) *Horizontal sparsity* (in $\{\beta_{jt}\}_{t=1}^{T}$): some predictors may not be important *at all times*

(c) *Smoothness* (in $\{\beta_{jt}\}_{t=1}^{T}$): the active coefficients evolve smoothly over time

We explore various **Dynamic Spike-and-Slab** *formulations for this setup.*

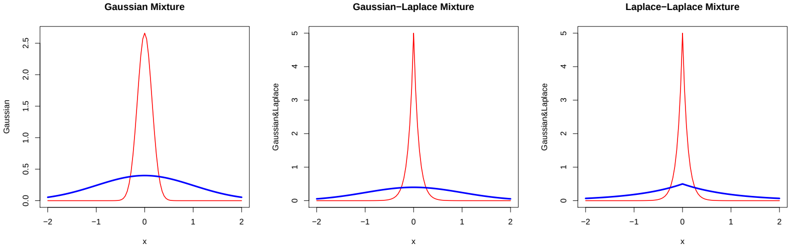
Related approaches: Bitto and Frühwirth-Schnatter (2018), Nakajima and West (2016), Kallin and Griffin (2016), Frühwirth-Schnatter and Wagner (2009)

Spike-and-Slab: Static Variable Selection

Mixtures of two densities for segregating *small* vs *large* effects

$$\pi(\beta_{tj} \mid \gamma_{tj}) = \gamma_{tj}\psi_1(\beta_{tj}) + (1 - \gamma_{tj})\psi_0(\beta_{tj}), \tag{13}$$

- $\psi_0(\beta_{tj})$ is a *spike* centered at zero (small variance)
- $\psi_1(\beta_{tj})$ is a *slab* centered at zero (large variance)
- $\mathrm{P}(\gamma_{tj} = 1 \mid \theta_{tj}) = \theta_{tj}$



Why Continuous Spike-and-Slab Priors?

- ☹ The continuous priors put zero mass on exactly sparse vectors
- ☺ However, posterior modes can be exactly sparse!
- ☺ Due to the continuity, we can implement fast optimization techniques
  - Coordinate-wise optimization (Rockova and George (2015))
  - EM (Rockova and George (2014), Ormerod et al. (2015))
  - IST, proximal methods...
- ☺ Continuous spike-and-slab priors achieve similar theoretical guarantees as point-mass mixtures (Rockova (2017), Narisetty and He (2015), Ishwaran and Rao (2005))

? **How can we make continuous Spike-and-Slab priors dynamic?**

(a) *Induce temporal dependencies in* $\{\beta_{tj}\}$
(b) *Induce temporal dependencies in* $\{\theta_{tj}\}$

Dynamic Spike-and-Slab Priors

Assume a conditional two-group prior

$$\pi(\beta_{tj} \mid \gamma_{tj}, \beta_{t-1j}) = \gamma_{tj}\psi_1(\beta_{tj} \mid \beta_{t-1j}) + (1 - \gamma_{tj})\psi_0(\beta_{tj}), \qquad (14)$$

where

- $\psi_0(\beta_{tj})$ is a spike *centered at zero* (*does not depend on* β_{t-1j})
- $\psi_1(\beta_{tj} \mid \beta_{t-1j})$ is a slab *centered around* β_{t-1j}
- $\mathrm{P}(\gamma_{tj} = 1 \mid \theta_{tj}) = \theta_{tj}$

The prior (14) can be regarded as a “multiple shrinkage prior" with *two shrinkage targets*

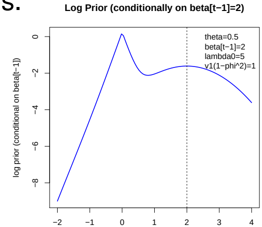
(1) zero (due to the gravitation of the spike)

(2) previous value β_{t-1j} (due to the gravitation of the slab)

Popular Spike-and-Slab Choices

⇝ *Laplace spike*: $\psi_0(\beta_{tj}) = \frac{\lambda_0}{2} e^{-|\beta_{tj}|\lambda_0}$

- ☺ The posterior has spikes at zeros!
- ☺ Automatic thresholding of small coefficients through posterior modes.



⇝ *Gaussian slab*: defined through a stationary $AR(1)$ process

$$\beta_{tj} = \phi_j \beta_{t-1j} + \nu_{tj}, \qquad \nu_{tj} \sim \mathcal{N}(0, \lambda_1(1 - \phi_j^2)) \tag{15}$$

with a stationary distribution $\mathcal{N}(0, \lambda_1)$ (when $|\phi_j| < 1$)

- ☺ Induces smoothness of the active coefficients

Dynamic Priors on Mixing Proportions

Denote by $\theta_{tj} = \mathrm{P}(\gamma_{tj} = 1 \mid \theta_{tj})$ the **random** mixing proportion

(1) *Logistic-normal AR(1) process* (Aitchison and Shen (1980))
Let $\pi(\theta_{tj}|\theta_{t-1j}, \widetilde{\phi}_j, \widetilde{\sigma})$ be distributed according to

$$\log\left[\frac{\theta_{tj}}{1-\theta_{tj}}\right] = \widetilde{\phi}_{0j} + \widetilde{\phi}_{1j} \log\left[\frac{\theta_{tj-1}}{1-\theta_{tj-1}}\right] + \widetilde{\nu}_{tj}$$

where $\widetilde{\nu}_{tj} \sim \mathcal{N}\left(0, (1-\widetilde{\phi}^2_{1j})\widetilde{\sigma}^2\right)$

(2) *Conditional Beta AR(1) process*

Let $\pi(\theta_{tj}|\theta_{t-1j}, \widetilde{\phi}_j)$ be a Beta distribution $\mathcal{B}\left(\widetilde{\mu}_t\,\widetilde{\phi}_{2j}, (1-\widetilde{\mu}_t)\widetilde{\phi}_{2j}\right)$
with expectation

$$\mathrm{E}(\theta_{tj}|\cdot) = \widetilde{\mu}_t \equiv \widetilde{\phi}_{0j} + \widetilde{\phi}_{1j}\theta_{t-1j}$$

and variance

$$\mathrm{Var}\,(\theta_{tj}|\cdot) = \widetilde{\mu}_t(1-\widetilde{\mu}_t)/(1+\widetilde{\phi}_{2j})$$

Switching type behavior when $\widetilde{\mu}_t\widetilde{\phi}_{2j} < 1$ and $(1-\widetilde{\mu}_t)\widetilde{\phi}_{2j} < 1$

Dynamic Priors on Mixing Proportions

Denote by $\theta_{tj} = \mathrm{P}(\gamma_{tj} = 1 \mid \theta_{tj})$ the **random** mixing proportion

(3) *Marginal Beta AR(1) process* (McKenzie 1985)

Conditional distribution:

$$\theta_{tj} = 1 - u_{tj}(1 - w_{tj}\theta_{t-1j})$$

where

$$u_{tj} \overset{\text{iid}}{\sim} \mathcal{B}(b_j, a_j - \phi_j) \quad \text{and} \quad w_{tj} \overset{\text{iid}}{\sim} \mathcal{B}(\phi_j, a_j - \phi_j).$$

When $\theta_{t-1j} \sim \mathcal{B}(a_j, b_j)$ then $\theta_{tj} \sim \mathcal{B}(a_j, b_j)$

Autocorrelation function

$$\rho(k) = \left[\frac{\phi_j b_j}{a_j(a_j + b_j - \phi_j)}\right]^k$$

☺ *Does imply Beta $\mathcal{B}(a_j, b_j)$ marginal distribution*

Mixture Autoregressive Priors

Can we construct a stationary time-series shrinkage prior whose marginals are the benchmark spike-and-slab priors?

The weight

$$\theta_{tj} = \mathsf{P}(\gamma_{tj} = 1|\theta_{tj})$$

is the key!

- The slab process has a *stationary distribution*

$$\psi_1^{ST}(\beta_{tj}) \sim \mathcal{N}(0, \lambda_1).$$

- The spike process has a *stationary distribution*

$$\psi_0^{ST}(\beta_{tj}) = \psi_0(\beta_{tj}).$$

How to specify the time-varying mixing weights θ_{tj}?

- Assume that $0 < \Theta_j < 1$ is a “global" mixing weight reflecting the *marginal* prior inclusion probability for j^{th} covariate
- Now let us set

$$\theta_{tj} = \frac{\Theta_j \psi_1^{ST}(\beta_{t-1j})}{\Theta_j \psi_1^{ST}(\beta_{t-1j}) + (1 - \Theta_j)\psi_0^{ST}(\beta_{t-1j})} \tag{16}$$

- It can be seen that θ_{tj} are “posterior" inclusion probabilities

$$\theta_{tj} = \mathrm{P}(\gamma_{tj} = 1 | \beta_{t-1j}, \Theta_j, \lambda_0, \lambda_1, \phi_j)$$

classifying β_{t-1j} as coming either from the spike or the slab

- The state-switching probabilities θ_{tj} thus depend on the previous value β_{t-1j} rather than θ_{t-1j}

Spike-and-Slab Autoregressive Process

Definition

Equations (14), (15) and (16) define the

Dynamic Spike-and-Slab Process (DSS)

with parameters $(\Theta_j, \lambda_0, \lambda_1, \phi_j)$. We will write

$$\{\beta_{tj}\} \sim DSS(\Theta_j, \lambda_0, \lambda_1, \phi_j)$$

- ⇝ DSS is an elaboration of mixture autoregressive (MAR) processes using *time-varying mixture weights* (Wong and Li (2000))
- ⇝ DSS is a variant of Gaussian mixture autoregressive processes (GMAR) (Kalliovirta et al. (2012))

DSS Stationarity Properties

The following result follows from Theorem 1 of Kalliovirta et al. (2012)

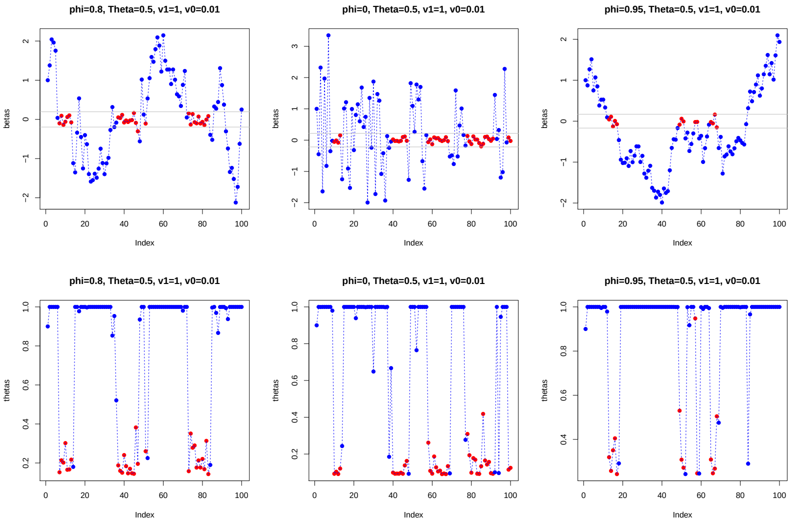
Theorem

Assume $\{\beta_{tj}\} \sim DSS(\Theta_j, \lambda_0, \lambda_1, \phi_j)$ with $|\phi_j| < 1$. Then $\{\beta_{tj}\}$ is Markov with a stationary distribution characterized by

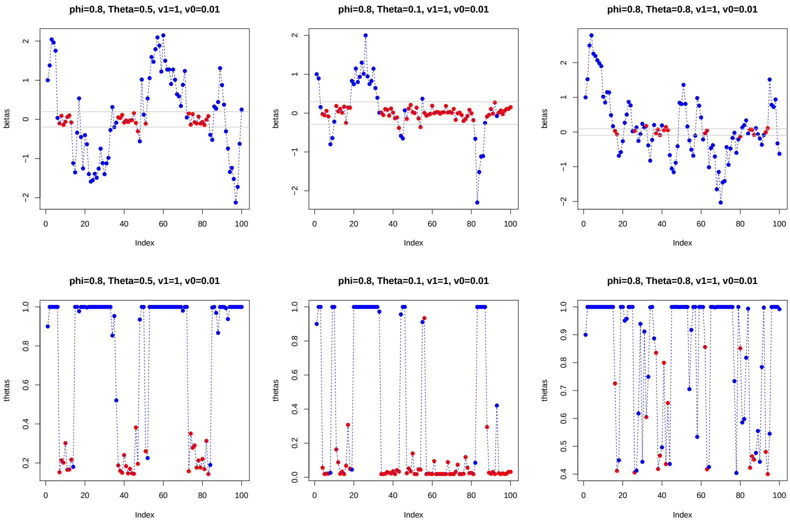
$$\pi(\beta|\Theta_j, \phi_j) = \Theta_j \psi_1^{ST}(\beta) + (1 - \Theta_j)\psi_0^{ST}(\beta)$$

- ☺ Univariate marginals of the DSS mixture process are Θ_j-weighted mixtures of marginals.
- ☺ *The marginal distribution for each β_{tj} is the spike-and-slab prior.*

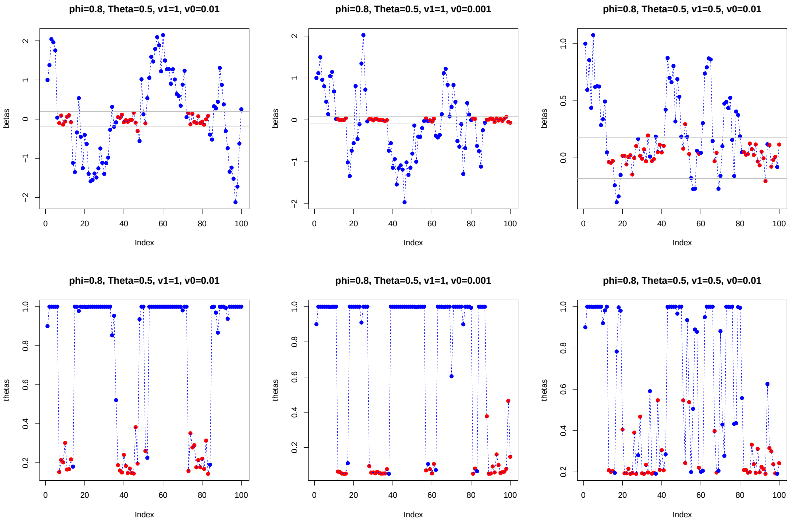
The effect of ϕ



The effect of Θ



The effects of (λ_1, λ_0)



Dynamic Penalty

Definition

For a given set of parameters $(\Theta, \lambda_0, \lambda_1, \phi_1)$, we define a *prospective* penalty function as

$$pen(\beta \mid \beta_{t-1}) = \log\left[(1-\theta_t)\,\psi_0(\beta) + \theta_t\,\psi_1(\beta \mid \beta_{t-1})\right]. \qquad (17)$$

Similarly, we define a *retrospective* penalty $pen(\beta_{t+1} \mid \beta)$ as a function of the second argument β in (17).

The Dynamic Spike-and-Slab (DSS) penalty is then defined as

$$Pen(\beta \mid \beta_{t-1}, \beta_{t+1}) = pen(\beta \mid \beta_{t-1}) + pen(\beta_{t+1} \mid \beta) + C, \qquad (18)$$

where $C \equiv -Pen(0 \mid \beta_{t-1}, \beta_{t+1})$ is a norming constant such that $Pen(0 \mid \beta_{t-1}, \beta_{t+1}) = 0$.

Penalty Plots

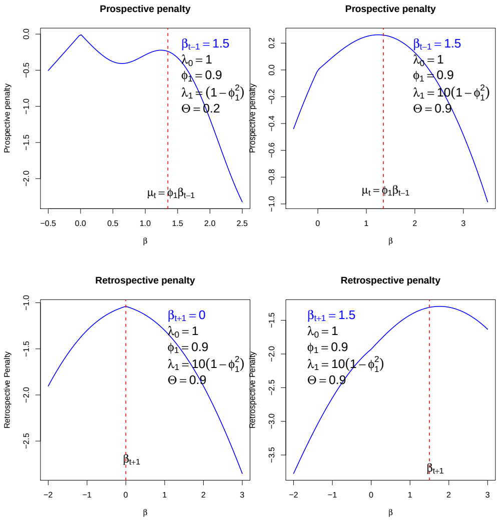


Figure: Plots of the prospective and retrospective penalty functions

Shrinkage Properties for MAP Smoothing

Shrinkage determined by

$$\frac{\partial\, Pen(\beta \mid \beta_{t-1}, \beta_{t+1})}{\partial|\beta|} \equiv -\Lambda^\star(\beta \mid \beta_{t-1}, \beta_{t+1}).$$

We will separate the term into:

$$\Lambda^\star(\beta \mid \beta_{t-1}, \beta_{t+1}) = \lambda^\star(\beta \mid \beta_{t-1}) + \widetilde{\lambda}^\star(\beta \mid \beta_{t+1}), \qquad (19)$$

- ⇝ *prospective* shrinkage effect $\lambda^\star(\beta \mid \beta_{t-1})$, driven by the past value β_{t-1}
- ⇝ *retrospective* shrinkage effect $\widetilde{\lambda}^\star(\beta \mid \beta_{t+1})$, driven by the future value β_{t+1}

where

$$\lambda^\star(\beta \mid \beta_{t-1}) = -\frac{\partial\, pen(\beta \mid \beta_{t-1})}{\partial|\beta|} \quad \text{and} \quad \widetilde{\lambda}^\star(\beta \mid \beta_{t+1}) = -\frac{\partial\, pen(\beta_{t+1}|\beta)}{\partial|\beta|}.$$

Shrinkage Properties

Prospective shrinkage

$$\lambda^\star(\beta \mid \beta_{t-1}) = -p_t^\star(\beta)\frac{\partial \log \psi_1(\beta \mid \beta_{t-1})}{\partial|\beta|} - [1 - p_t^\star(\beta)]\frac{\partial \log \psi_0(\beta)}{\partial|\beta|},$$
$$= p_t^\star(\beta)\left(\frac{\beta - \mu_t}{\lambda_1}\right)\text{sign}(\beta) + [1 - p_t^\star(\beta)]\lambda_0$$

where

$$p_t^\star(\beta) \equiv \frac{\theta_t\psi_1(\beta \mid \beta_{t-1})}{\theta_t\psi_1(\beta \mid \beta_{t-1}) + (1-\theta_t)\psi_0(\beta)}.$$

Retrospective shrinkage

We will write $p_{t+1}^\star = p_{t+1}^\star(\beta_{t+1})$.

$$\widetilde{\lambda}^\star(\beta \mid \beta_{t+1}) = \left[\lambda_0 - \text{sign}(\beta)\left(\frac{\beta}{\lambda_1}\right)\right]\left[(1 - p_{t+1}^\star)\theta_{t+1} - p_{t+1}^\star(1-\theta_{t+1})\right]$$
$$- p_{t+1}^\star\phi_1\text{sign}(\beta)\left(\frac{\beta_{t+1} - \mu_{t+1}}{\lambda_1}\right).$$

The Global Mode

Assume

$$y_t = \boldsymbol{x}_t' \beta_t^0 + \varepsilon_t, \quad t = 1, \ldots, T, \tag{20}$$

and

$$\{\beta_{tj}\} \sim DSS(\Theta_j, \lambda_0, \lambda_1, \phi_j)$$

Let $\widehat{\boldsymbol{B}} = \{\widehat{\beta}_{tj}\}_{t,j=1}^{T,p}$ denote the global mode of $\pi(\boldsymbol{B}|\boldsymbol{Y})$.

Lemma

Let $\widehat{\boldsymbol{B}}_{\backslash tj}$ denote all but the $(t,j)^{th}$ entry in $\widehat{\boldsymbol{B}}$ and by $z_{tj} = y_t - \sum_{i \neq j} x_{ti} \widehat{\beta}_{ti}$. Then $\widehat{\beta}_{tj}$ satisfies the following necessary condition

$$\widehat{\beta}_{tj} = \begin{cases} \frac{1}{x_{tj}^2} \left[x_{tj} z_{tj} - \Lambda^\star(\widehat{\beta}_{tj} \mid \widehat{\beta}_{t-1j}, \widehat{\beta}_{t-1j}) \right]_+ \operatorname{sign}(x_{tj} z_{tj}) & \text{if} \quad \Delta_{tj}^- < x_{tj} z_{tj} < \Delta_{tj}^+ \\ 0 & \text{otherwise.} \end{cases}$$

Global mode thresholds coefficients to zero.

One-Step-Late EM for Obtaining the Mode

Initial condition: Assume that $\boldsymbol{\beta}_0$ (at time $t = 0$) came from the stationary distribution.

The mode of the posterior $\pi(\boldsymbol{\beta}_0, \boldsymbol{B} \mid \boldsymbol{Y})$ can be found iteratively by maximizing

$$\log \pi(\boldsymbol{\beta}_0, \boldsymbol{B}, \boldsymbol{\gamma}_0, \boldsymbol{\Gamma} \mid \boldsymbol{Y})$$

treating $\boldsymbol{\gamma}_0$ and $\boldsymbol{\Gamma}$ as missing data.

E-step:

$$p_{tj}^{\star} = \mathrm{P}(\gamma_{tj} = 1 | \beta_{tj}^{(m)}, \beta_{t-1j}^{(m)}, \theta_{tj})$$

M-step:

$$\beta_{tj}^{(m+1)} = \frac{1}{W_{tj} + (1 - \phi_1^2)/\lambda_1 M_{tj}} \left[Z_{tj} - \Lambda_{tj} \right]_+ \operatorname{sign}(Z_{tj}), \quad \text{for} \quad 1 < t < T,$$

where $M_{tj} = p_{t+1j}^{\star}(1 - \theta_{t+1j}) - \theta_{t+1j}(1 - p_{t+1j}^{\star})$, $\Lambda_{tj} = \lambda_0[(1 - p_{tj}^{\star}) - M_{tj}]$.
$Z_{tj} = x_{tj} z_{tj} + \frac{p_{tj}^{\star}\phi_1}{\lambda_1}\beta_{t-1j}^{(m+1)} + \frac{p_{t+1j}^{\star}\phi_1}{\lambda_1}\beta_{t+1j}^{(m+1)}$ and $W_{tj} = \left(x_{tj}^2 + \frac{p_{tj}^{\star}}{\lambda_1} + \frac{p_{t+1j}^{\star}\phi_1^2}{\lambda_1} \right)$.

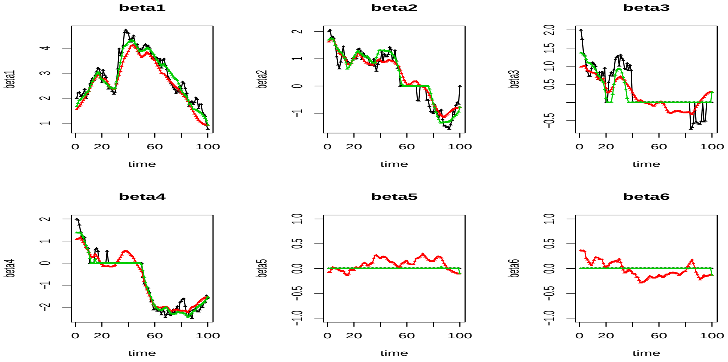
DSS prior captures intermittent zeros

Assume that the true coefficients came from an AR(1) process

$$\beta_{tj}^0 = \phi_1 \beta_{t-1j}^0 + \nu_{tj}, \quad \phi_1 = 0.98, \quad \nu_{tj} \overset{\text{iid}}{\sim} \mathcal{N}[0, 10(1-\phi_1^2)]$$

and were thresholded to *zero* if $|\beta_{tj}^0| < 0.5$. We apply the DSS prior with $\Theta = 0.9, \lambda_1 = 10(1-\phi_1)^2, \phi_1 = 0.98, \lambda_0 = 1$.

Assume $T = 100$ and $p = 6$ and obtain y_t from (12).



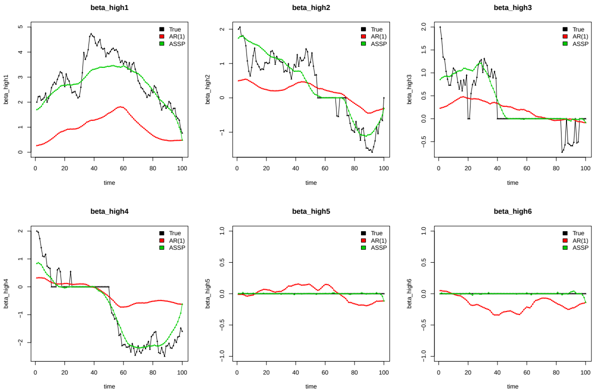
The impact is more pronounced in higher dimensions

Assume that the true nonzero coefficients came from AR(1)

$$\beta_{tj}^0 = \phi_1 \beta_{t-1j}^0 + \nu_{tj}, \quad \phi_1 = 0.98, \quad \nu_{tj} \stackrel{\text{iid}}{\sim} \mathcal{N}[0, 10(1-\phi_1^2)]$$

and were thresholded to *zero* if $|\beta_{tj}^0| < 0.5$.

Assume $T = 100$ and $p = 50$ and obtain y_t from (12).



Folding DSS within DFA

High-dimensional multivariate time series $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T] \in \mathbb{R}^{P \times T}$.

Evolving covariance patterns over time can be captured with the following *state space model*:

$$\boldsymbol{Y}_t = \boldsymbol{B}_t \boldsymbol{\omega}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{\text{ind}}{\sim} \mathcal{N}_P(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{21}$$

$$\boldsymbol{\omega}_t = \boldsymbol{\Phi} \boldsymbol{\omega}_{t-1} + \mathbf{e}_t, \quad \mathbf{e}_t \overset{\text{ind}}{\sim} \mathcal{N}_K(\mathbf{0}, \sigma_\omega^2 \mathbb{I}_K). \tag{22}$$

- $\boldsymbol{B}_t = \{\beta_{jk}\}_{j,t=1}^{P,K}$ is *assigned a DSS prior* independently for each (j, k)
- Factors may enter and leave the model as time passes

Related procedures: Kaufmann and Schumacher (2013), Del Negro and Otrok (2008)

Expanded Model

We work with the expanded model

$$\boldsymbol{Y}_t = \boldsymbol{B}_t \boldsymbol{A}_{tL}^{-1} \boldsymbol{\omega}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{\text{ind}}{\sim} \mathcal{N}_P(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{23}$$

$$\boldsymbol{\omega}_t = \boldsymbol{\Phi} \boldsymbol{\omega}_{t-1} + \mathbf{e}_t, \quad \mathbf{e}_t \overset{\text{ind}}{\sim} \mathcal{N}_K(\mathbf{0}, \boldsymbol{A}_t), \tag{24}$$

where $\boldsymbol{A}_{tL}$ is the lower Cholesky factor of a positive semi-definite matrix $\boldsymbol{A}_t$ and $\boldsymbol{A}_t \overset{i.i.d}{\sim} \pi(\boldsymbol{A}) \propto 1$.

We assume the initial condition $\boldsymbol{\omega}_0 \sim \mathcal{N}_K(\mathbf{0}, \sigma_\omega^2/(1-\phi^2)\boldsymbol{I}_K)$ and impose the DSS prior on the individual entries of the *rotated* matrix $\boldsymbol{B}_t^\star = \boldsymbol{B}_t \boldsymbol{A}_{tL}^{-1}$.

The idea is to *rotate towards sparse orientations* throughout the iterations of the EM algorithm.

Computation: Parameter Expanded EM

The E-step operates in the reduced space (keeping $\boldsymbol{A}_t = \sigma^2_\omega \boldsymbol{I}_K$)

⇝ Kalman filter:

- $\mathrm{E}[\boldsymbol{\omega}_t \mid \boldsymbol{Y}, \boldsymbol{B}_{1:T}, \boldsymbol{\Sigma}_{1:T}]$,
- $\mathrm{var}[\boldsymbol{\omega}_t \mid \boldsymbol{Y}, \boldsymbol{B}_{1:T}, \boldsymbol{\Sigma}_{1:T}]$
- $\mathrm{cov}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_{t-1} \mid \boldsymbol{Y}, \boldsymbol{B}_{1:T}, \boldsymbol{\Sigma}_{1:T}$

⇝ Compute $\mathrm{P}(\gamma_{jk} = 1 \mid \beta_{jk})$

The M-step operates in the expanded space (allowing for general $\boldsymbol{A}_t$).

⇝ Compute $\boldsymbol{\Sigma}_{1:T}$ from Forward Filtering Backward Smoothing

⇝ Compute $\boldsymbol{B}^\star_{1:T}$ by solving P independent penalized dynamic regressions

⇝ Compute rotation matrix $\boldsymbol{A}_t$

Rotation step:

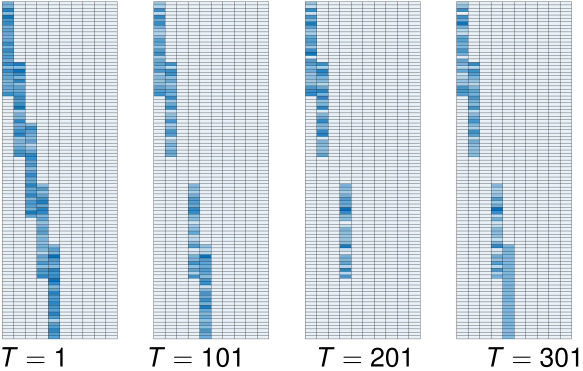
$$\boldsymbol{B}_t = \boldsymbol{B}^\star_t \boldsymbol{A}_{tL}$$

Simulated Example

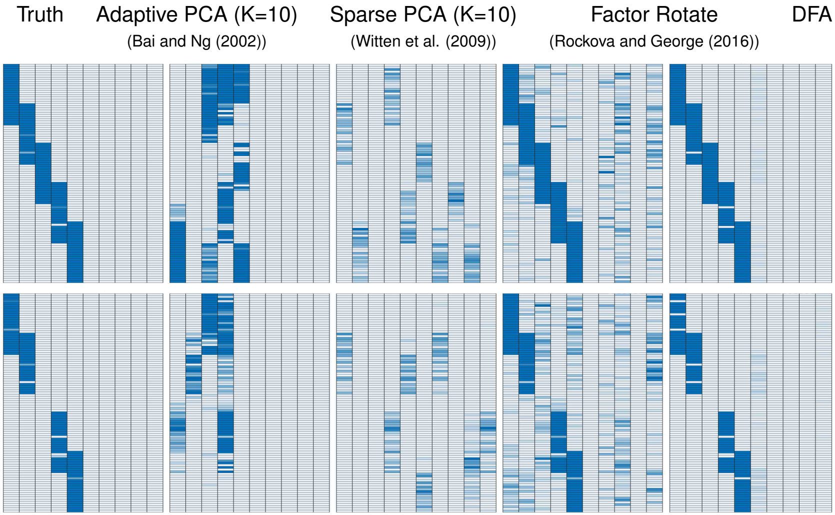
Assume $P = 100$, $K = 10$ and $T = 400$ time series observations

The true nonzero loadings are smooth: $\beta_{jk}^{0t} = \phi\beta_{jk}^{0t-1} + v_{jk}^{t}$ with $v_{jk}^{t} \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.0025)$ for $\phi = 0.99$.

When loadings β_{jk}^{0t} become inactive, they are thresholded to zero.

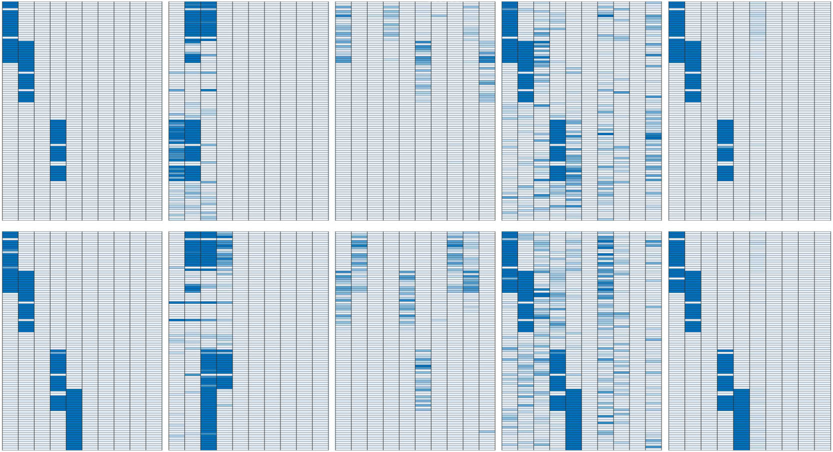


Simulated Example: $t = 100, 200$



Simulated Example: $t = 300, 400$

Truth | Adaptive PCA (K=10) (Bai and Ng (2002)) | Sparse PCA (K=10) (Witten et al. (2009)) | Factor Rotate (Rockova and George (2016)) | DFA



Macro Study

The empirical application concerns a large-scale monthly U.S. macroeconomic database.

It consists of a balanced panel of $P = 127$ variables tracked over the period of $2001/01$ to $2015/12$ ($T = 180$).

These variables are classified into eight main categories:

Output and Income	*Labor Market*	*Consumption and Orders*
Orders and Inventories	*Money and Credit*	*Interest Rate and Exchange Rates*
Prices	*Stock Market*	

We are interested in assessing the evolution of the economy: degree of connectivity and permanence of structural changes

We examine the output of our procedure at *three time points: 2003/12, 2008/10, and 2015/12.*

These represent three distinct states of the economy: relative stability (2003), sharp economic crisis (2008), and recovery (2015).

Estimated Snapshots

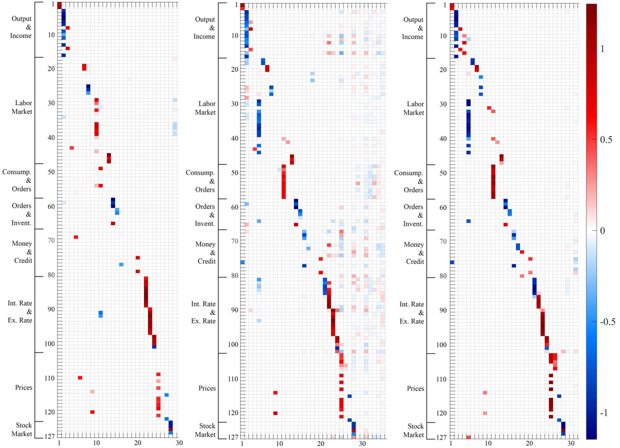


Figure: Estimated factor loadingsat $t = 2003/12$ (left), $t = 2008/10$ (center), $t = 2015/12$ (right), with the original series on the y-axis and the factors in the x-axis.

Suggested Interpretation: $t = 2003/12$

There are 24 active factors in total with only 5 factors that cluster eight or more series (Factors 2, 10, 22, 23, and 25).

Factor 2 can be interpreted as *durable goods*: includes CMRMTSPLx (real manufacturing and trade industry sales), CUMFNS (capacity utilization), DMANEMP (durable goods employment), and ISRATIOx (manufacturing and trade inventories to sales ratio).

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

Factor 25 includes consumer price indices, medical care, durables, and services, as well as personal consumptions expenditures on nondurable goods.

Suggested Interpretation: $t = 2008/10$

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).

Another interesting observation is the **emergence of new factors**.

Factor 11, which includes housing starts and new housing permits in different regions in the U.S., was *not* present pre-crisis

Factor 28 emerges as a non-sparse link between many different sectors of the economy, including retail sales, industrial production, employment, 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).

Factor 25, on the other hand, is driven mainly by prices (e.g. CPI).

Both of these factors could be potentially interpreted as crisis factors.

Suggested Interpretation: $t = 2015/12$

Although most of the factor *overlap has dissipated*.

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.).

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.

Degree of Connectivity

To understand the degree of connectivity/overlap between factors, we plot the average number of active factors per series over time.

More overlap indicates a more intertwined economy.

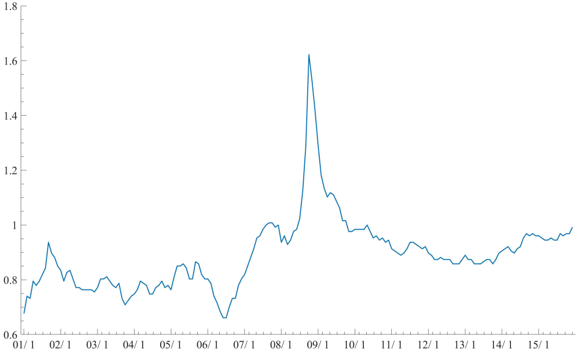


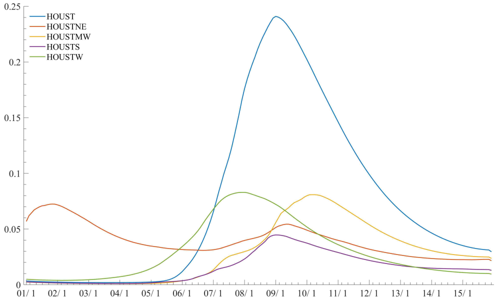
Figure: The average number of estimated active factors (with absolute loadings above 0.1) per series over the period 2001/1:2015/12.

Idiosyncratic Variances

HOUST (total housing starts) and its regional variants (North East, Mid-West, South, and West)

Seasonally adjusted number of new residential construction projects that have begun during any particular month.

Increased uncertainty in housing starts is a global phenomenon but that there is heterogeneity across regions as to the magnitude and timing.



Thank you! ☺

Some References

McAlinn, K. (2018), Ročková, V. and Saha, E.
Dynamic Sparse Factor Analysis
Working Paper

Ročková, V. and McAlinn, K. (2017+)
Dynamic Variable Selection with Spike-and-Slab Process Priors
Bayesian Anaylsis (in revision)

Ročková, V. and George, E. (2015)
The Spike-and-Slab LASSO
Journal of the American Statistical Association, 113: 431-444

Ročková, V. and George, E. (2014)
EMVS: The EM Approach to Bayesian Variable Selection
Journal of the American Statistical Association,109:828-846

Ročková, V. (2017)
Bayesian Estimation of Sparse Signals with a Continuous Spike-and-Slab Prior
The Annals of Statistics, 46:401-437