Approximate Bayesian inference for dynamic sparsity in time-varying parameter model

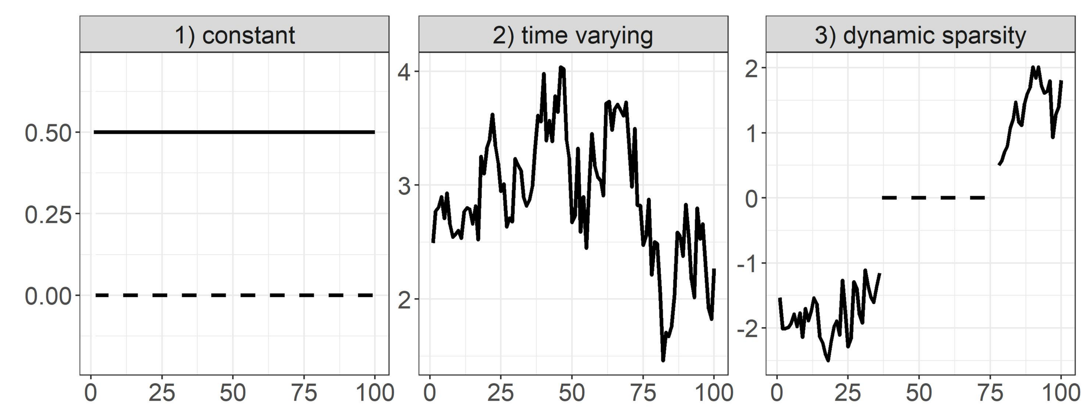
Nicolas Bianco - Department of Statistical Sciences, University of Padova

Statistical methods and models for complex data — Poster Session: 21 September 2022

SCAN ME!
Follow the poster with multimedia contents

Motivation

Dynamic modeling accounts for different behaviour of the regression coefficients:



State of the art Bayesian statistics tackle this problem via dynamic shrinkage prior (see Kalli and Griffin, 2014; Koval et al, 2019) or dynamic continuous spike and slab prior (see Koop and Korobilis, 2020; Ročková and McAlinn, 2021):

✗ Complex hyper-parameter tuning ✗ Scales bad when large number of covariates

Model and Inference

Bernoulli-Gaussian specification (Ormerod et al., 2017) for time-varying parameter regression model:

$$y_t = \sum_{j=1}^{p} x_{j,t}\gamma_{j,t}\beta_{j,t} + \varepsilon_t, \qquad \varepsilon_t \sim \mathsf{N}(0, \sigma_t^2),$$

where $y_t \in \mathbb{R}$ is the response and $\mathbf{x}_t \in \mathbb{R}^p$ is a set of known covariates.

- Assume a random walk dynamic for the time-varying coefficients and the logarithm of the time-varying variance $h_t = \log \sigma_t^2$:
$$\beta_{j,t} \sim \mathsf{N}(\beta_{j,t-1}, \eta_j^2), \qquad h_t \sim \mathsf{N}(h_{t-1}, \nu^2).$$
- The indicator variables are independent $\gamma_{j,t}|\omega_{j,t} \sim \mathsf{Bern}(p_{j,t})$ given $\omega_{j,t} = \mathrm{logit}(p_{j,t})$ and the dependence in the a priori inclusion probabilities via $\omega_{j,t} \sim \mathsf{N}(\omega_{j,t-1}, \xi_j^2)$.
- Inverse-Gamma prior for the variances parameters ν^2, η_j^2, and ξ_j^2. As an alternative, $\xi_j^2 \in (0, +\infty)$ can be fixed.
- $\beta_{j,t}$ and $\omega_{j,t}$ can be also regularised with a hierarchical shrinkage prior such as in Bitto and Frühwirth-Schnatter (2019).

Semiparametric Variational Bayes

The goal is to find the best approximation q^* to the posterior distribution p such that

$$q^* = \arg\min_{q \in \mathcal{Q}} \mathsf{KL}(q||p).$$

Problem: $\mathcal{Q}$ is too general ⇒ make some assumptions!

Non-parametric. Factorise the joint variational density:

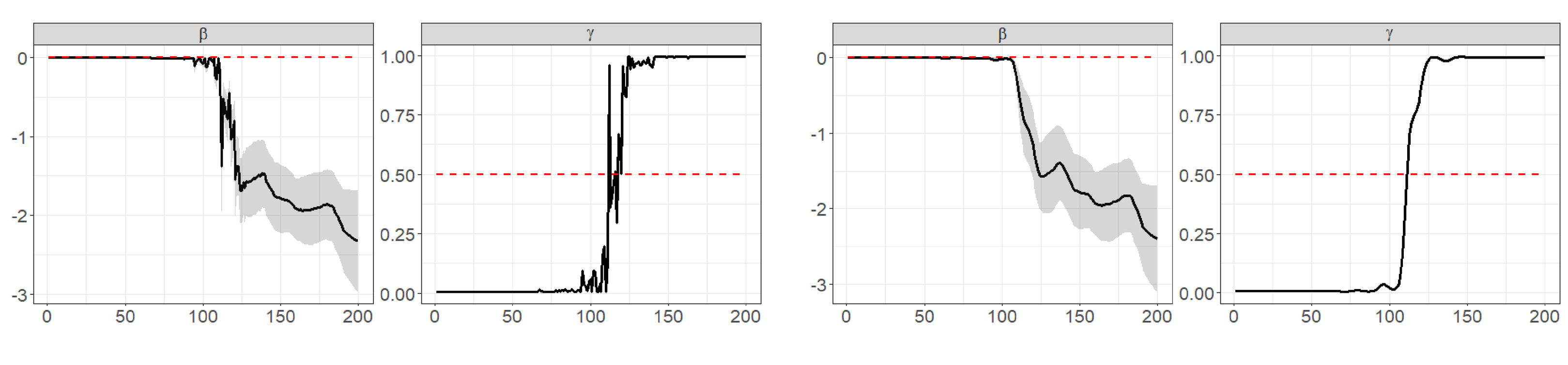
$$q(\boldsymbol{\vartheta}) = q(\mathbf{h})q(\nu^2)\prod_{j=1}^{p} q(\boldsymbol{\beta}_j)q(\boldsymbol{\omega}_j)q(\eta_j^2)q(\xi_j^2)\prod_{t=1}^{n} q(\gamma_{j,t}),$$

closed-form updates can be computed.

Parametric: impose a parametric density function on q. Here:

1) Gaussian approximation for $\mathbf{h}$.
2) Bernoulli approximation with mean constraints for $\gamma_{j,t}$.

Assumption 2) is important to get smooth trajectories for both regression coefficients and posterior inclusion probabilities:



(a) Non-parametric approximation (b) Parametric approximation

References

[1] Bitto, A., Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*.

[2] Kalli, M., Griffin, J. (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics*.

[3] Koop, G., Korobilis, D. (2020). Bayesian dynamic variable selection in high dimensions. *arXiv*.

[4] Koval, D. R., Matteson, D. S., Ruppert, D. (2019). Dynamic shrinkage processes. *J. R. Statist. Soc. B*.

[5] Ormerod, J. T., You, C., Müller, S. (2017). A variational Bayes approach to variable selection. *Electronic Journal of Statistics*.

[6] Ročková, V., McAlinn, K. (2021). Dynamic Variable Selection with Spike-and-Slab Process Priors. *Bayesian Analysis*.

Main result (extension of Result 1 in Ormerod et al., 2017)

Assume that for variable j at iteration i of the algorithm:

$$\max_t\{\mu^{(i)}_{q(\gamma_{j,t})}\} = \mu^{(i)}_{q(\gamma_{j,s_1})} = \epsilon \ll 1 \ \text{ and } \ \boldsymbol{\Sigma}^{(i)}_{q(\omega_j)} - \boldsymbol{\Sigma}^{(i-1)}_{q(\omega_j)} \text{ is a non-negative matrix,}$$

then it holds:

1. $\mu^{(i+1)}_{q(\gamma_{j,t})} = \mathrm{expit}\left\{\mu^{(i+1)}_{q(\omega_{j,t})} - \frac{1}{2}\mu^{(i+1)}_{q(1/\sigma_t^2)}x^2_{j,t}\mu^{-1(i+1)}_{q(1/\eta_j^2)}q_{t,t} + O(\epsilon)\right\}$, where $q_{t,t} = [\mathbf{Q}^{-1}]_{t,t}$.
2. $\mu^{(i+1)}_{q(\omega_{j,t})} = -\frac{1}{2}\sum_{k=1}^{n} s_{t,k} + O(\epsilon)$, where $s_{t,k} = [\boldsymbol{\Sigma}_{q(\omega_j)}]_{t,k}$.
3. $\mu^{(i+1)}_{q(\omega_{j,t})} \leq \mu^{(i)}_{q(\omega_{j,t})}$ decreases after each iteration.

This has two main implications:

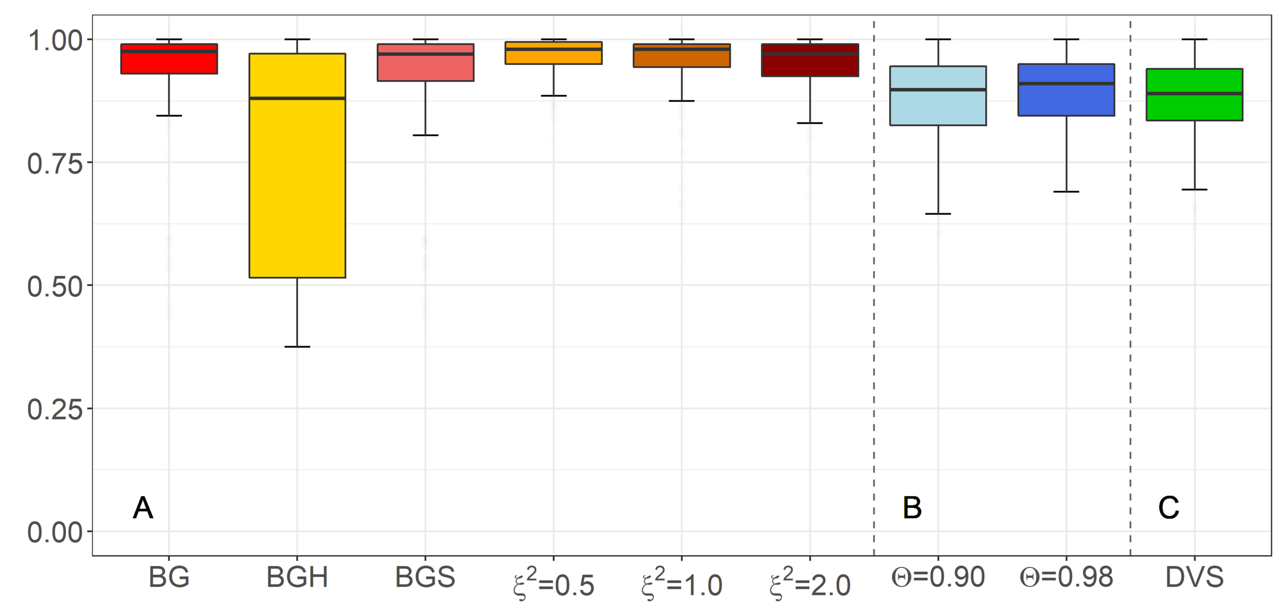
1) *Sparsity*: when ϵ is sufficiently small and $\mu^{(i+1)}_{q(\omega_{j,t})} \ll 0$, after i iterations, $\mu^{(i+1)}_{q(\gamma_{j,t})}$ can be approximated and it is represented as 0 when implemented on a computer, for all t.
2) *Dimension reduction*: if $\mu^{(i)}_{q(\gamma_{j,t})} \approx 0$, $\forall t$ at iteration i, then the successive updates remain $\mu^{(i+k)}_{q(\gamma_{j,t})} \approx 0$. Thus we can remove the j-th variable from the matrix $\mathbf{X}$.

Simulations

Samples from $y_t = \mathbf{x}_t^\intercal \boldsymbol{\Gamma}_t \boldsymbol{\beta}_t + \varepsilon_t$, where $\varepsilon_t \sim \mathsf{N}(0, 0.25)$, for $t = 1, \ldots, 200$. Here $p = 50$: the intercept is always included, 4 parameters are dynamically selected and the last 45 are always excluded.

A) Bernoulli-Gaussian models: non-parametric VB (BG), homoscedastic (BGH), parametric VB (BGS), and fixed ξ_j^2.
B) Dynamic spike-and-slab (DSS) of Ročková and McAlinn (2021).
C) Dynamic variable selection (DVS) of Koop and Korobilis (2020).

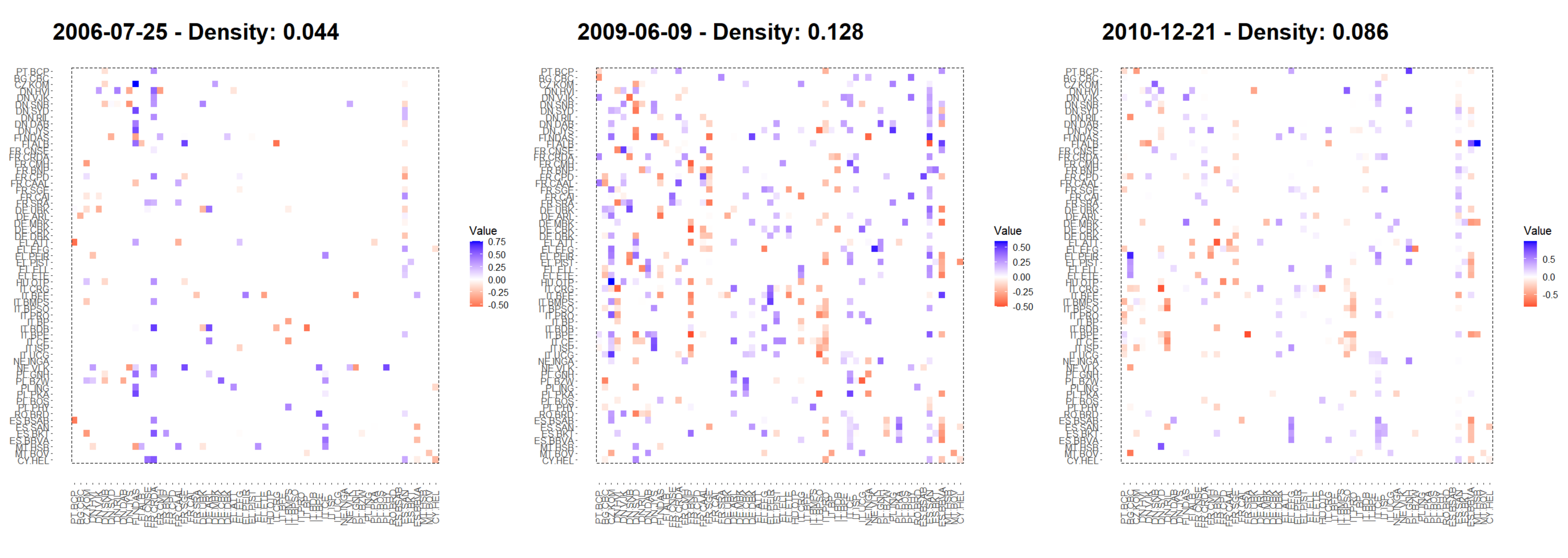
Dynamic selection accuracy



European banks' return predictability

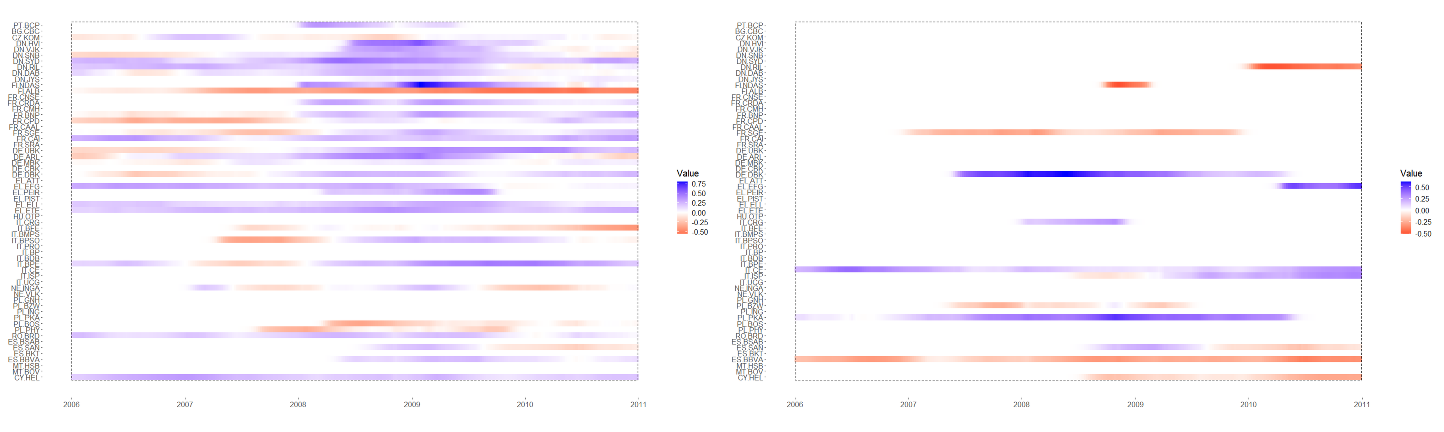
- Data: 60 European Banks from 2006 to 2012.
- Approach: sparse TVP-VAR(1) equation-by-equation.

Predictability in the financial network:



(c) Transition matrix pre-crisis (d) Transition matrix during the crisis (e) Transition matrix post-crisis

Santander Counsumer bank role in the financial network:



(f) Variable importance (g) Predictability

Contact information

Nicolas Bianco, Ph.D. student
Department of Statistical Sciences, University of Padova
nicolas.bianco@phd.unipd.it
whitenoise8.github.io