

Dynamic variable selection in high-dimensional predictive regressions

Forecasting @ Risk, ECB

Daniele Bianchi¹ & Mauro Bernardi² & Nicolas Bianco

¹ School of Economics and Finance, Queen Mary, University of London

² Department of Statistical Sciences, University of Padua, Italy

³ Department of Economics and Business, Universitat Pompeu Fabra, Spain.

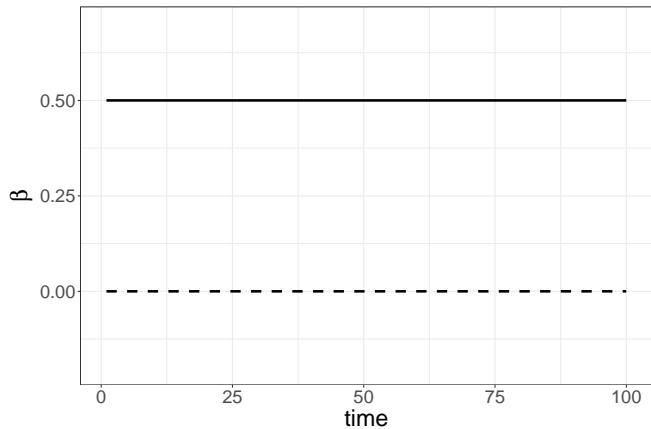
Set the stage

Set the stage

$$y_t = \beta x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2)$$

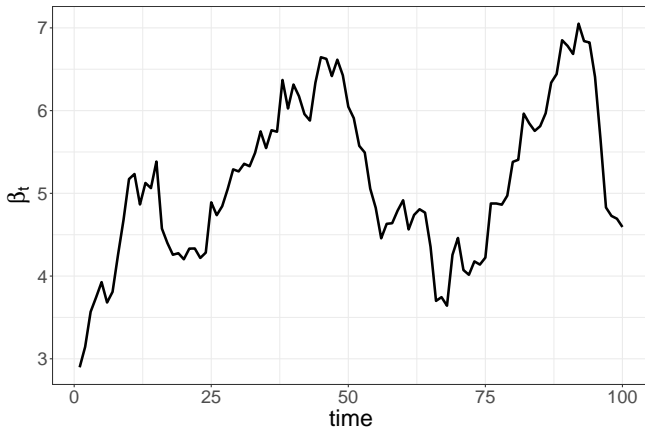
Set the stage

$$y_t = \beta x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2)$$



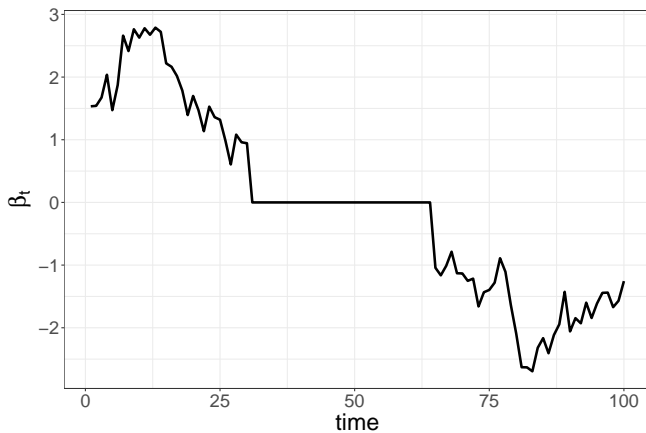
Set the stage

$$y_t = \beta_t x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2)$$



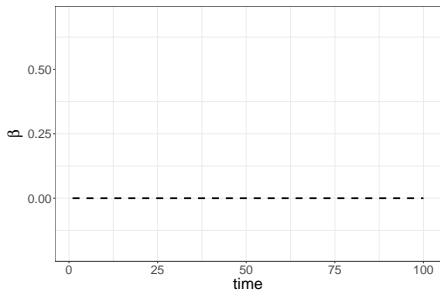
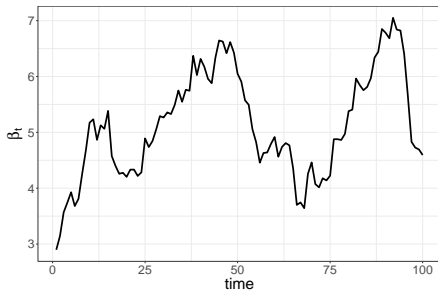
Set the stage

$$y_t = \beta_t x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$$



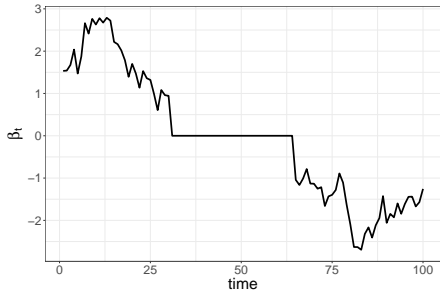
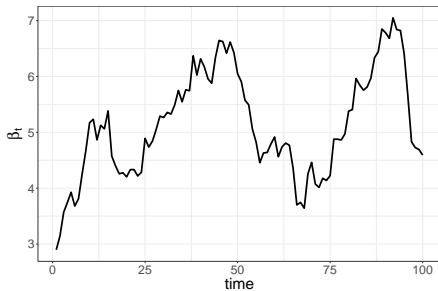
Set the stage

$$y_t = \sum_{j=1}^p \beta_{jt} x_{jt-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2)$$



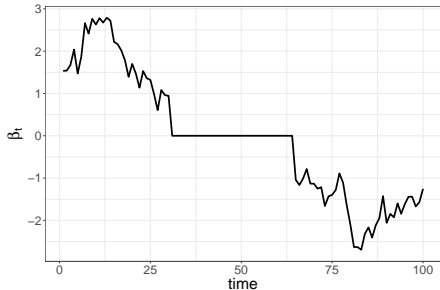
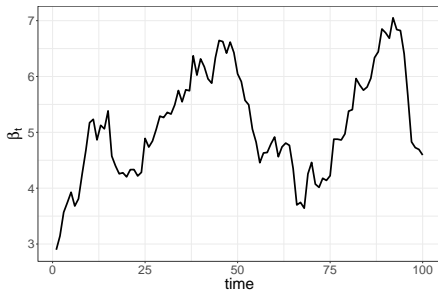
Set the stage

$$y_t = \sum_{j=1}^p \beta_{jt} x_{jt-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2)$$



Set the stage

$$y_t = \sum_{j=1}^p \beta_{jt} x_{jt-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$$



Key issues:

- ↪ The set p could be large ($p \rightarrow T$).
- ↪ Which predictor matters and when is unknown a priori...
- ↪ ...this matters even more in dynamic settings.

This paper

Bayesian method for **dynamic variable selection** in high-dimensional **time-varying** predictive regressions.

This paper

Bayesian method for **dynamic variable selection** in high-dimensional **time-varying** predictive regressions.



A single predictor can be “active” over multiple time periods.

↪ *Intensive margin.*

This paper

Bayesian method for **dynamic variable selection** in high-dimensional **time-varying** predictive regressions.



A single predictor can be “active” over multiple time periods.

↪ *Intensive margin.*



Multiple predictors can be “active” in a single time period.

↪ *Extensive margin.*

This paper

Bayesian method for **dynamic variable selection** in high-dimensional **time-varying** predictive regressions.



A single predictor can be “active” over multiple time periods.

↪ *Intensive margin.*



Multiple predictors can be “active” in a single time period.

↪ *Extensive margin.*

Method \implies Novel variational Bayes inference approach:

↪ Minimal hyper-parameters tuning.

↪ Posterior concentration properties comparable to MCMC.

↪ On-line dimension reduction (efficiency).

This paper

Bayesian method for **dynamic variable selection** in high-dimensional **time-varying** predictive regressions.



A single predictor can be “active” over multiple time periods.

↪ *Intensive margin.*



Multiple predictors can be “active” in a single time period.

↪ *Extensive margin.*

Empirical exercise(s) \implies Economic forecasting (point and density):

- ↪ Inflation predictability over several quarterly horizons based on +230 macroeconomic predictors (FRED-QD data).
- ↪ Equity risk premium predictability one-month ahead based on +150 anomaly/characteristic-based portfolios.

Some reference literature

A non-exhaustive list of references:

- ↪ Bayesian methods for variable/model selection:
(e.g., Mitchell & Beauchamp (1988), George & McCulloch (1997), Nakajima & West (2013), Kalli & Griffin (2014), Kowal et al. (2019), Bitto & Frühwirth-Schnatter (2019), **Koop & Korobilis (2020)**, **Ročková & McAlinn (2021)**, Giannone et al. (2021), etc.)
- ↪ Economic forecasting in large-dimensional models:
(e.g., Stock & Watson (2007), Stock & Watson (2010), Faust & Wright (2013), Huber et al. (2021), Dong et al. (2022), etc.)
- ↪ Variational Bayes inference methods:
(e.g., Ormerod & Wand (2010), Ormerod et al. (2017), Gefang et al. (2019), Koop & Korobilis (2020), Chan & Yu (2022), etc.)

Model specification

Bayesian model specification

Dynamic Bernoulli-Gaussian (BG) regression specification:

$$y_t = \sum_{j=1}^p \beta_{jt} x_{jt-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2),$$

where

$$\beta_{jt} = \gamma_{jt} b_{jt}, \quad \text{and} \quad \gamma_{jt} \in \{0, 1\},$$

Bayesian model specification

Dynamic Bernoulli-Gaussian (BG) regression specification:

$$y_t = \sum_{j=1}^p \beta_{jt} x_{jt-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2),$$

where

$$\beta_{jt} = \gamma_{jt} b_{jt}, \quad \text{and} \quad \gamma_{jt} \in \{0, 1\},$$

Dynamics based on two latent processes:

↪ Time-varying coefficients b_{jt} , $j = 1, \dots, p$.

Bayesian model specification

Dynamic Bernoulli-Gaussian (BG) regression specification:

$$y_t = \sum_{j=1}^p \beta_{jt} x_{jt-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, \sigma_t^2),$$

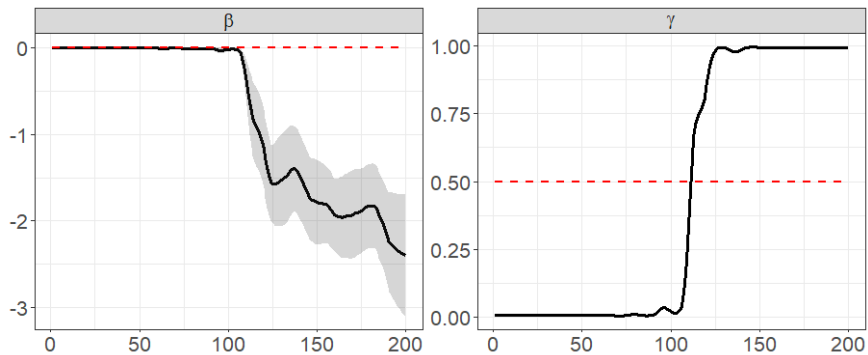
where

$$\beta_{jt} = \gamma_{jt} b_{jt}, \quad \text{and} \quad \gamma_{jt} \in \{0, 1\},$$

Dynamics based on two latent processes:

- ↪ Time-varying coefficients b_{jt} , $j = 1, \dots, p$.
- ↪ Dynamic variable indicator γ_{jt} , $j = 1, \dots, p$.

In other words...



Bayesian model specification

For the latent process b_{jt} we assume:

$$b_{jt} = b_{jt-1} + u_{jt}, \quad u_{jt} \sim \text{N}(0, \eta_j^2),$$

Bayesian model specification

For the latent process b_{jt} we assume:

$$b_{jt} = b_{jt-1} + u_{jt}, \quad u_{jt} \sim \mathbf{N}(0, \eta_j^2), \quad \implies \quad \mathbf{b}_j \sim \mathbf{N}_{n+1}(\mathbf{0}, \eta_j^2 \mathbf{Q}^{-1})$$

with $\mathbf{Q} = \begin{bmatrix} 1 + k_0^{-1} & -1 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}$

GAUSSIAN MARKOV
RANDOM FIELD (GMRF)
REPRESENTATION

Bayesian model specification

For the latent process b_{jt} we assume:

$$b_{jt} = b_{jt-1} + u_{jt}, \quad u_{jt} \sim \mathbf{N}(0, \eta_j^2), \quad \implies \quad \mathbf{b}_j \sim \mathbf{N}_{n+1}(\mathbf{0}, \eta_j^2 \mathbf{Q}^{-1})$$

with $\mathbf{Q} = \begin{bmatrix} 1 + k_0^{-1} & -1 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}$

GAUSSIAN MARKOV
RANDOM FIELD (GMRF)
REPRESENTATION

Similarly, for $h_t = \log \sigma_t^2$ we assume:

$$h_t = h_{t-1} + e_t, \quad e_t \sim \mathbf{N}(0, \nu^2),$$

Bayesian model specification

For the latent process b_{jt} we assume:

$$b_{jt} = b_{jt-1} + u_{jt}, \quad u_{jt} \sim \mathbf{N}(0, \eta_j^2), \quad \implies \quad \mathbf{b}_j \sim \mathbf{N}_{n+1}(\mathbf{0}, \eta_j^2 \mathbf{Q}^{-1})$$

with $\mathbf{Q} = \begin{bmatrix} 1 + k_0^{-1} & -1 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}$

GAUSSIAN MARKOV
RANDOM FIELD (GMRF)
REPRESENTATION

Similarly, for $h_t = \log \sigma_t^2$ we assume:

$$h_t = h_{t-1} + e_t, \quad e_t \sim \mathbf{N}(0, \nu^2), \quad \implies \quad \underbrace{\mathbf{h} \sim \mathbf{N}_{n+1}(\mathbf{0}, \nu^2 \mathbf{Q}^{-1})}_{\text{GMRF}}$$

Bayesian model specification

For the dynamics of sparsity we assume:

$$\gamma_{jt} | \omega_{jt} \sim \text{Bern}(\text{expit}(\omega_{jt}))$$

Bayesian model specification

For the dynamics of sparsity we assume:

$$\gamma_{jt} | \omega_{jt} \sim \text{Bern}(\text{expit}(\omega_{jt})) \quad \implies \quad \underbrace{\omega_j \sim N_{n+1}(\mathbf{0}, \xi_j^2 \mathbf{Q}^{-1})}_{\text{GMRF}}$$

Bayesian model specification

For the dynamics of sparsity we assume:

$$\gamma_{jt} | \omega_{jt} \sim \text{Bern}(\text{expit}(\omega_{jt})) \quad \implies \quad \underbrace{\omega_j \sim \text{N}_{n+1}(\mathbf{0}, \xi_j^2 \mathbf{Q}^{-1})}_{\text{GMRF}}$$

Such that the marginal for $\gamma_{j1}, \dots, \gamma_{jn}$

$$p(\gamma_{j1}, \dots, \gamma_{jn}) = \int p(\omega_j) \prod_{t=1}^n p(\gamma_{jt} | \omega_{jt}) d\omega_j,$$

has correlated components.

↪ i.e., time series dependence driven by ω_j .

Bayesian model specification

For the dynamics of sparsity we assume:

$$\gamma_{jt} | \omega_{jt} \sim \text{Bern}(\text{expit}(\omega_{jt})) \quad \implies \quad \underbrace{\omega_j \sim \text{N}_{n+1}(\mathbf{0}, \xi_j^2 \mathbf{Q}^{-1})}_{\text{GMRF}}$$

Such that the marginal for $\gamma_{j1}, \dots, \gamma_{jn}$

$$p(\gamma_{j1}, \dots, \gamma_{jn}) = \int p(\omega_j) \prod_{t=1}^n p(\gamma_{jt} | \omega_{jt}) d\omega_j,$$

has correlated components.

↪ i.e., time series dependence driven by ω_j .

Prior distributions for the variances parameters.

↪ $\nu^2 \sim \text{IG}(\underbrace{A_\nu}_{=0.01}, B_\nu)$, $\eta_j^2 \sim \text{IG}(\underbrace{A_\eta}_{=0.01}, B_\eta)$, and $\xi_j^2 \sim \text{IG}(A_\xi, B_\xi)$.

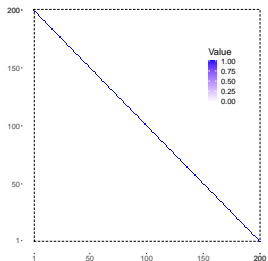
Digression: Some comparative statics on the priors

How posterior estimates are affected by the choice of (A_ξ, B_ξ) ?

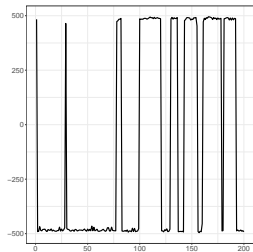
Digression: Some comparative statics on the priors

How posterior estimates are affected by the choice of (A_ξ, B_ξ) ?

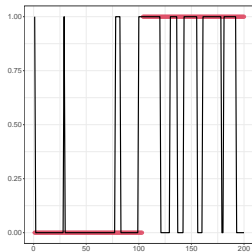
↪ **Scenario A** $\Rightarrow A_\xi$ constant and $B_\xi \rightarrow +\infty$.



(d)



(e)



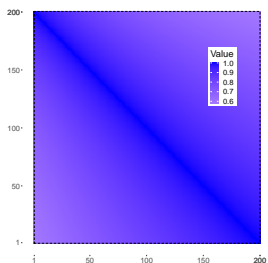
(f) $\{\mu_q(\gamma_{jt})\}_{t=1}^n$

Scenario A: $B_\xi \rightarrow +\infty$. (a) Depicts the variational correlation matrix for the process $\{\omega_{jt}\}_{t=1}^n$ obtained from $\Sigma_q(\omega_j)$. (b) Plots the trajectory of $\{\mu_q(\omega_{jt})\}_{t=1}^n$. (c) Shows the effect on the posterior inclusion probabilities estimated $\{\mu_q(\gamma_{jt})\}_{t=1}^n$ compared to the simulated (red points).

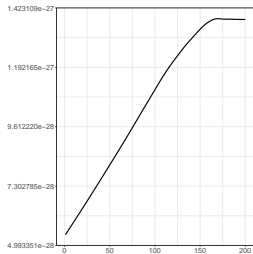
Digression: Some comparative statics on the priors

How posterior estimates are affected by the choice of (A_ξ, B_ξ) ?

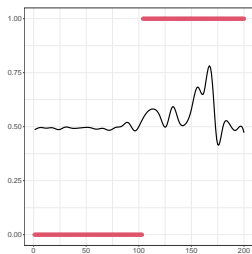
→ **Scenario B** $\Rightarrow B_\xi$ constant and $A_\xi \rightarrow +\infty$.



(g)



(h)



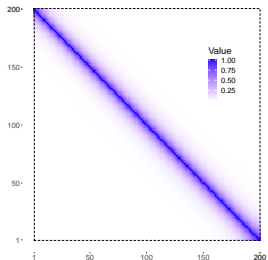
(i) $\{\mu_q(\gamma_{j,t})\}_{t=1}^n$

Scenario B: $A_\xi \rightarrow +\infty$. (a) Depicts the variational correlation matrix for the process $\{\omega_{j,t}\}_{t=1}^n$ obtained from $\Sigma_q(\omega_{j,t})$. (b) Plots the trajectory of $\{\mu_q(\omega_{j,t})\}_{t=1}^n$. (c) Shows the effect on the posterior inclusion probabilities estimated $\{\mu_q(\gamma_{j,t})\}_{t=1}^n$ compared to the simulated (red points).

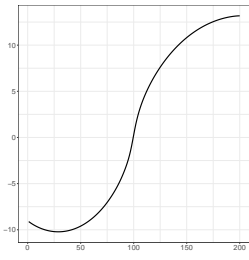
Digression: Some comparative statics on the priors

How posterior estimates are affected by the choice of (A_ξ, B_ξ) ?

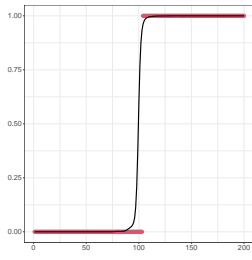
↪ **Scenario C** $\Rightarrow A_\xi/B_\xi \implies c_1, c_1 \in \mathbb{R}^+$.



(j)



(k)



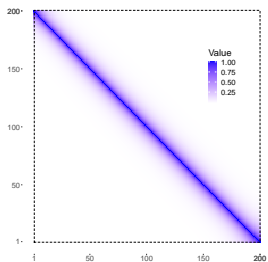
(l) $\{\mu_q(\gamma_{jt})\}_{t=1}^n$

Scenario C: $A_\xi/B_\xi \rightarrow c_1, c_1 \in \mathbb{R}^+$. (a) Depicts the variational correlation matrix for the process $\{\omega_{jt}\}_{t=1}^n$ obtained from $\Sigma_q(\omega_j)$. (b) Plots the trajectory of $\{\mu_q(\omega_{jt})\}_{t=1}^n$. (c) Shows the effect on the posterior inclusion probabilities estimated $\{\mu_q(\gamma_{jt})\}_{t=1}^n$ compared to the simulated (red points).

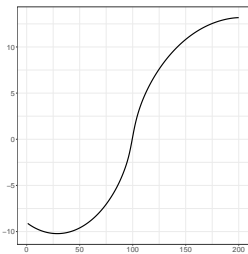
Digression: Some comparative statics on the priors

How posterior estimates are affected by the choice of (A_ξ, B_ξ) ?

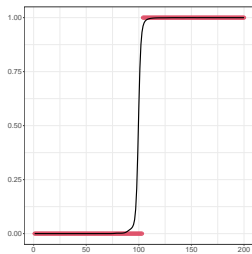
↪ Scenario C $\Rightarrow A_\xi/B_\xi \implies \xi_j^2 \sim \text{IG}(A_\xi = 2, B_\xi = 5)$



(m)



(n)



(o) $\{\mu_q(\gamma_{jt})\}_{t=1}^n$

Scenario C: $A_\xi/B_\xi \rightarrow c_1$, $c_1 \in \mathbb{R}^+$. (a) Depicts the variational correlation matrix for the process $\{\omega_{jt}\}_{t=1}^n$ obtained from $\Sigma_q(\omega_j)$. (b) Plots the trajectory of $\{\mu_q(\omega_{jt})\}_{t=1}^n$. (c) Shows the effect on the posterior inclusion probabilities estimated $\{\mu_q(\gamma_{jt})\}_{t=1}^n$ compared to the simulated (red points).

Variational Bayes inference

A re-cap on Variational Bayes (VB) inference

Minimize the Kullback-Leibler (KL) divergence between a variational density $q(\boldsymbol{\vartheta})$ and the true posterior density $p(\boldsymbol{\vartheta}|\mathbf{y})$.

A re-cap on Variational Bayes (VB) inference

Minimize the Kullback-Leibler (KL) divergence between a variational density $q(\boldsymbol{\vartheta})$ and the true posterior density $p(\boldsymbol{\vartheta}|\mathbf{y})$.

This corresponds to find $q^*(\boldsymbol{\vartheta})$ such that:

$$q^*(\boldsymbol{\vartheta}) = \arg \max_{q(\boldsymbol{\vartheta}) \in \mathcal{Q}} \log \underline{p}(\mathbf{y}; q),$$

with (see Ormerod & Wand 2010),

$$\underline{p}(\mathbf{y}; q) = \int q(\boldsymbol{\vartheta}) \log \left\{ \frac{p(\mathbf{y}, \boldsymbol{\vartheta})}{q(\boldsymbol{\vartheta})} \right\} d\boldsymbol{\vartheta},$$

the variational, or “effective”, lower bound (ELBO).

↪ N.B., both $q(\boldsymbol{\vartheta})$ and $p(\mathbf{y}, \boldsymbol{\vartheta})$ are known.

A re-cap on Variational Bayes (VB) inference (cont'd)

The choice of Q leads to different approaches.

A re-cap on Variational Bayes (VB) inference (cont'd)

The choice of \mathcal{Q} leads to different approaches.

Mean-field variational Bayes (non-parametric):

$$\mathcal{Q} = \{q(\boldsymbol{\vartheta}) : \prod_{i=1}^M q(\boldsymbol{\vartheta}_i), \text{ for a partition } (\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_M)\}.$$

↪ e.g., in a typical linear regression $q(\boldsymbol{\beta}, \sigma^2) = q(\boldsymbol{\beta})q(\sigma^2)$.

↪ Closed-form updates based on coordinate ascent algorithm.

A re-cap on Variational Bayes (VB) inference (cont'd)

The choice of \mathcal{Q} leads to different approaches.

Mean-field variational Bayes (non-parametric):

$$\mathcal{Q} = \{q(\boldsymbol{\vartheta}) : \prod_{i=1}^M q(\boldsymbol{\vartheta}_i), \text{ for a partition } (\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_M)\}.$$

↪ e.g., in a typical linear regression $q(\boldsymbol{\beta}, \sigma^2) = q(\boldsymbol{\beta}) q(\sigma^2)$.

↪ Closed-form updates based on coordinate ascent algorithm.

Parametric variational Bayes:

$$\mathcal{Q} = \{q(\boldsymbol{\vartheta}) : q(\boldsymbol{\vartheta} | \boldsymbol{\lambda}_{q(\boldsymbol{\vartheta})}) = f(\boldsymbol{\vartheta}; \boldsymbol{\lambda}_{q(\boldsymbol{\vartheta})})\},$$

↪ e.g., $f(\cdot)$ Gaussian s.t., $\boldsymbol{\lambda}_{q(\boldsymbol{\vartheta})} = (\boldsymbol{\mu}_{q(\boldsymbol{\vartheta})}, \boldsymbol{\Sigma}_{q(\boldsymbol{\vartheta})})$.

Semi-parametric variational Bayes

We propose an **hybrid approach** which merge parametric and non-parametric VB to estimate ϑ .

Non-parametric \Rightarrow mean-field factorization of $q(\vartheta)$:

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

Semi-parametric variational Bayes

We propose an **hybrid approach** which merge parametric and non-parametric VB to estimate ϑ .

Non-parametric \Rightarrow mean-field factorization of $q(\vartheta)$:

$$q(\vartheta) = q(\mathbf{h})q(\nu^2) \prod_{j=1}^p q(\mathbf{b}_j)q(\omega_j)q(\eta_j^2)q(\xi_j^2) \prod_{t=1}^n q(\gamma_{jt}) q(z_{jt}).$$

Parametric \Rightarrow the red- q belong to a parametric family of distributions.

- \hookrightarrow Multivariate Gaussian for $q(\mathbf{h})$ (recall the GMRF representation).
- \hookrightarrow Polya-Gamma representation $\gamma_{jt}|\omega_{jt}$ s.t., $p(\gamma_{jt}|\omega_{jt}) \approx q(\gamma_{jt}) q(z_{jt})$ where $p(z_{jt}) \sim \text{PG}(1, 0)$.

Variational density of $\beta = \gamma b$

Proposition (see paper)

Define $\beta_j = \Gamma_j \mathbf{b}_j$, where $\mathbf{b}_j = (b_{j0}, b_{j1}, \dots, b_{jn})'$ and $\Gamma_j = \text{diag}(1, \gamma_{j1}, \dots, \gamma_{jn})$. The optimal variational density of β_j is given by a mixture of multivariate Gaussian distributions:

$$q^*(\beta_j) = \sum_{\mathbf{s} \in \mathcal{S}} w_{\mathbf{s}} \mathbf{N}_{n+1}(\mathbf{D}_{\mathbf{s}} \boldsymbol{\mu}_{q(\mathbf{b}_j)}, \mathbf{D}_{\mathbf{s}}^{1/2} \boldsymbol{\Sigma}_{q(\mathbf{b}_j)} \mathbf{D}_{\mathbf{s}}^{1/2}), \quad (1)$$

where \mathcal{S} a sequence of $\{0, 1\}$ of length n with cardinality $|\mathcal{S}| = 2^n$, the diagonal matrix $\mathbf{D}_{\mathbf{s}} = \text{diag}(1, s_1, \dots, s_n)$, and mixing weights:

$$w_{\mathbf{s}} = \prod_{t=1}^n \mu_{q(\gamma_{jt})}^{s_t} (1 - \mu_{q(\gamma_{jt})})^{1-s_t}, \quad (2)$$

where $\mathbf{s} = (s_1, \dots, s_t, \dots, s_n) \in \mathcal{S}$. Moreover, the mean $\boldsymbol{\mu}_{q(\beta_j)}$ and variance $\boldsymbol{\Sigma}_{q(\beta_j)}$ can be computed analytically... (see paper)

Properties of the variational updates

Proposition (see paper)

Assume for variable j at iteration i of the algorithm:

$$\hookrightarrow \max_t \{ \mu_{q(\gamma_{jt})}^{(i)} \} = \epsilon \ll 1.$$

$$\hookrightarrow \Sigma_{q(\omega_j)}^{(i)} - \Sigma_{q(\omega_j)}^{(i-1)} \text{ is a non-negative matrix.}$$

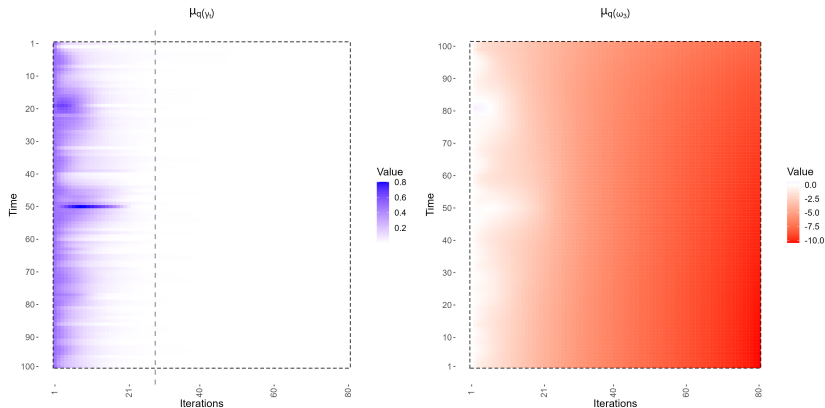
It holds that:

1. $\mu_{q(\gamma_{jt})}^{(i+1)} = \text{expit} \left\{ \mu_{q(\omega_{jt})}^{(i+1)} - \frac{1}{2} \mu_{q(1/\sigma_t^2)}^{(i+1)} x_{jt}^2 \mu_{q(1/\eta_j^2)}^{-1(i+1)} q_{t,t} + O(\epsilon) \right\},$
2. $\mu_{q(\omega_{jt})}^{(i+1)} = -\frac{1}{2} \sum_{k=1}^n s_{t,k} + O(\epsilon),$
3. $\mu_{q(\omega_{jt})}^{(i+1)} \leq \mu_{q(\omega_{jt})}^{(i)}$ decreases after each iteration,

where $\text{expit} = \text{logit}^{-1}$, $q_{tt} = [\mathbf{Q}^{-1}]_{t,t}$ and $s_{tk} = [\Sigma_{q(\omega_j)}]_{tk}$.

Properties of the variational updates

Dimension reduction



Variational update over iterations (x-axis) until convergence of the vector of posterior inclusion probabilities $(\mu_q(\gamma_{j1}), \dots, \mu_q(\gamma_{jn}))$ (left panel) and $(\mu_q(\omega_{j1}), \dots, \mu_q(\omega_{jn}))$ (right panel), for a parameter j which is always zero $\forall t$. The value of the update is given by the blue intensity. The dashed line identifies the iteration at which convergence is reached for $\epsilon = 0.01$.

Efficient variational Bayes inference scheme

Algorithm 1: Efficient variational Bayes for dynamic sparse regressions.

```
 $q(\boldsymbol{\vartheta}), \Delta_{\boldsymbol{\vartheta}}, A_{\nu}, B_{\nu}, A_{\eta}, B_{\eta}, A_{\xi}, B_{\xi}$  while  $(\widehat{\Delta}_{\boldsymbol{\vartheta}} > \Delta_{\boldsymbol{\vartheta}})$  do  
  for  $j = 1, \dots, p$  do  
    Update  $q(\mathbf{b}_j)$  as in 2.1; and  $q(\eta_j)$  as in A.8;  
    Update  $q(\boldsymbol{\omega}_j)$  as in 2.3 and  $q(\xi_j)$  as in A.9;  
    for  $t = 1, \dots, n$  do  
      Update  $q(z_{jt})$  as in A.7;  
      Update  $q(\gamma_{jt})$  as in 2.2 (non-smooth) or 2.6 (smooth);  
    end  
  end  
  Update  $q(\boldsymbol{\sigma})$  as in A.1 (heteroskedastic) or A.2 (homoskedastic);  
  Update  $q(\nu^2)$  as in A.10;  
  if assumption in the previous Proposition holds then  
    for  $j = 1, \dots, p$  do  
      if  $\max_t \{\mu_{q(\gamma_{jt})}\} < \epsilon$  then  
        | Drop the  $j$ -th variable  
      end  
    end  
  end  
  Compute  $\widehat{\Delta}_{\boldsymbol{\vartheta}} = q(\boldsymbol{\vartheta})^{(\text{iter})} - q(\boldsymbol{\vartheta})^{(\text{iter}-1)}$  ;
```

end

Simulation study

Comparison with MCMC

Simulation setting:

- ↪ Generate 3 processes $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}_{t=1}^{100}$ such that $\beta_{1t} \neq 0, \forall t$, $\beta_{2t} = 0, \forall t$, and β_{3t} shows dynamic sparsity.
- ↪ Generate $N = 100$ replicates from $y_t = x_{1t}\beta_{1t} + x_{2t}\beta_{2t} + x_{3t}\beta_{3t} + \varepsilon_t$, with $\varepsilon_t \sim N(0, 0.25)$.
- ↪ Estimate the model with both **VB** and **MCMC**.

Comparison with MCMC

Simulation setting:

- ↪ Generate 3 processes $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}_{t=1}^{100}$ such that $\beta_{1t} \neq 0, \forall t$, $\beta_{2t} = 0, \forall t$, and β_{3t} shows dynamic sparsity.
- ↪ Generate $N = 100$ replicates from $y_t = x_{1t}\beta_{1t} + x_{2t}\beta_{2t} + x_{3t}\beta_{3t} + \varepsilon_t$, with $\varepsilon_t \sim \mathbf{N}(0, 0.25)$.
- ↪ Estimate the model with both **VB** and **MCMC**.

The accuracy of the approximation is quantified as in Wand et al. (2011):

$$ACC(\beta) = \left\{ 1 - 0.5 \int |q(\beta) - p(\beta|\mathbf{y})| d\beta \right\} \%,$$

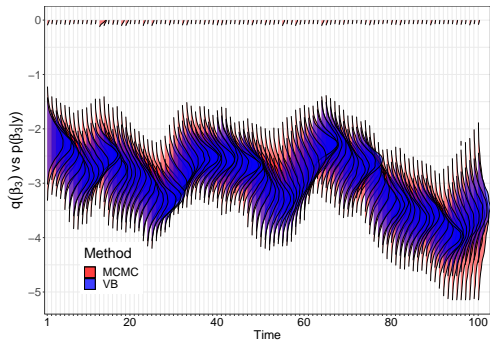
where $q(\beta)$ is the variational approximation and $p(\beta|\mathbf{y})$ is the posterior from an equivalent MCMC with a large number of draws.

Comparison with MCMC

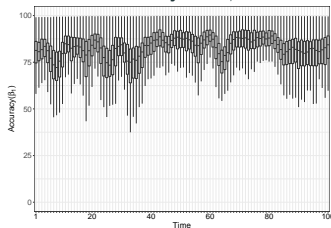
Time-varying, $\beta_{1t} \neq 0, \forall t$

Posterior densities for β_1

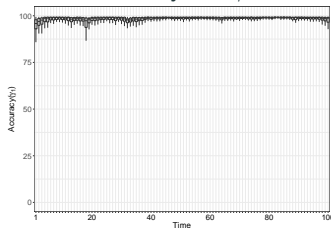
VB (blue) against MCMC (red)



Accuracy for β_1



Accuracy for γ_1

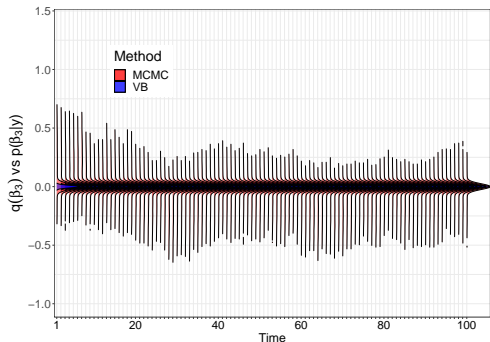


Comparison with MCMC

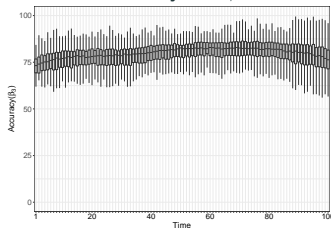
Constant at zero, i.e., $\beta_{2t} = 0, \forall t$

Posterior densities for β_2

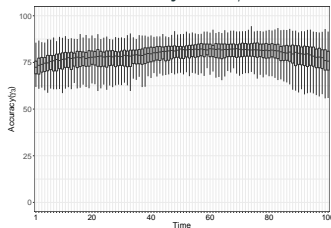
VB (blue) against MCMC (red)



Accuracy for β_2



Accuracy for γ_2

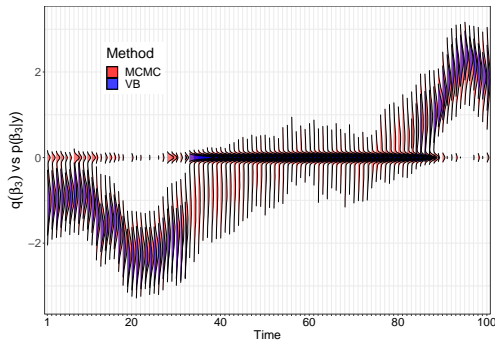


Comparison with MCMC

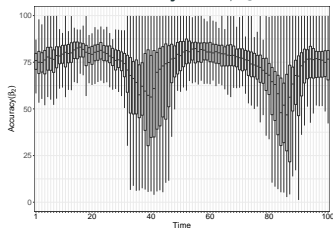
Dynamic sparsity

Posterior densities for β_3

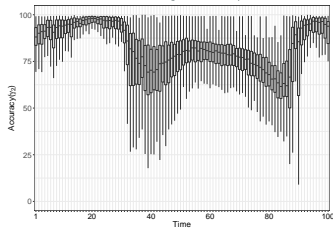
VB (blue) against MCMC (red)



Accuracy for β_3



Accuracy for γ_3



Comparison with Bayesian variable selection methods

Simulation setting

$N = 100$ replicates from the following data generating process:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, 0.25), \quad t = 1, \dots, 200,$$

The dimension of the parameter $\boldsymbol{\beta}_t$ is equal to $p = 50, 100, 200$.

- ↪ β_{1t} is always included, i.e. $\gamma_{1t} = 1, \forall t$;
- ↪ $\beta_{2:7,t}$ dynamic sparsity, i.e. $\gamma_{2:7,t}$ vary over time;
- ↪ $\beta_{8:p,t}$ is always excluded, i.e. $\gamma_{8:p,t} = 0, \forall t$.

Comparison with Bayesian variable selection methods

Simulation setting

$N = 100$ replicates from the following data generating process:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(0, 0.25), \quad t = 1, \dots, 200,$$

The dimension of the parameter $\boldsymbol{\beta}_t$ is equal to $p = 50, 100, 200$.

- ↪ β_{1t} is always included, i.e. $\gamma_{1t} = 1, \forall t$;
- ↪ $\beta_{2:7,t}$ dynamic sparsity, i.e. $\gamma_{2:7,t}$ vary over time;
- ↪ $\beta_{8:p,t}$ is always excluded, i.e. $\gamma_{8:p,t} = 0, \forall t$.

We consider different versions of our estimation algorithm:

- ↪ BG \implies basic model, with no smoothing on $\mathbb{P}(\gamma_{jt} = 1)$;
- ↪ BGH \implies as BG, but homoschedastic assumption;
- ↪ BGS \implies model with smoothing on $\mathbb{P}(\gamma_{jt} = 1)$;
- ↪ BG with fixed latent process variances ξ_j^2 .

Comparison across methods

Benchmarks and metrics

Benchmark methods:

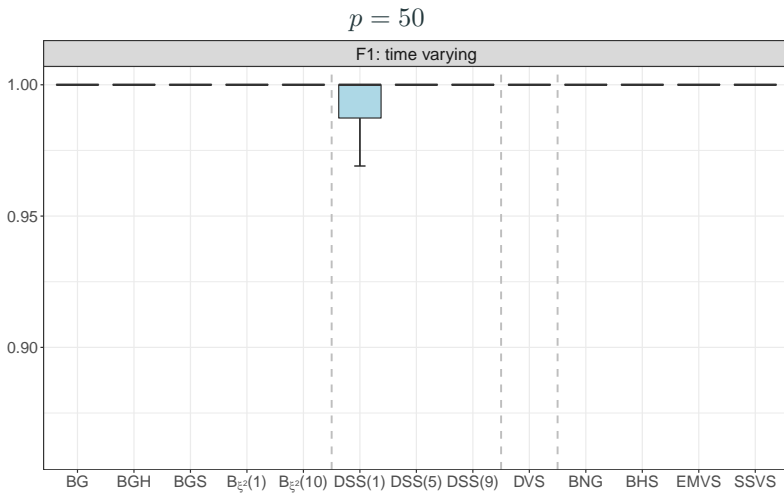
- ↪ Static models with rolling windows estimate: normal-gamma (NG), horseshoe (HS) and spike-and-slab methods (SSVS, EMVS);
- ↪ Dynamic spike-and-slab (DSS) of Ročková & McAlinn (2021), for $\Theta = \{0.1, 0.5, 0.9\}$;
- ↪ VB Dynamic variable selection (DVS) of Koop & Korobilis (2020);

Performance metrics:

- ↪ Signal/variable identification/selection (F1-score);
- ↪ Computational efficiency (running time in seconds).

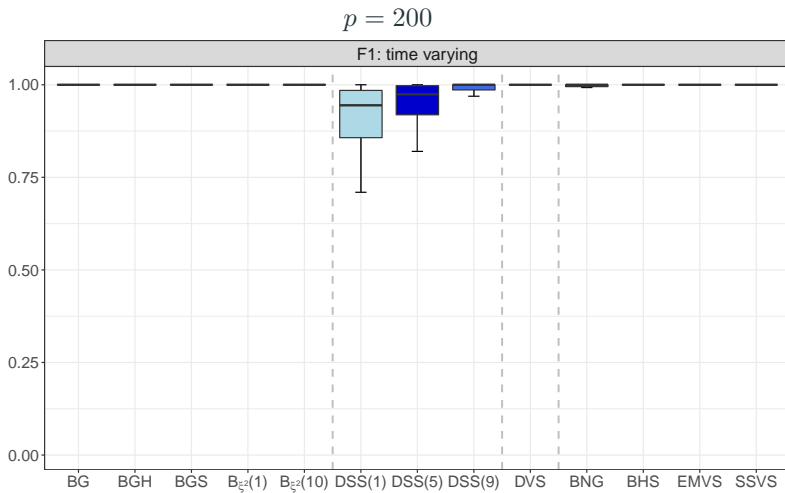
Simulation study

Scenario 1 $\implies \beta_{jt} \neq 0 \forall t$



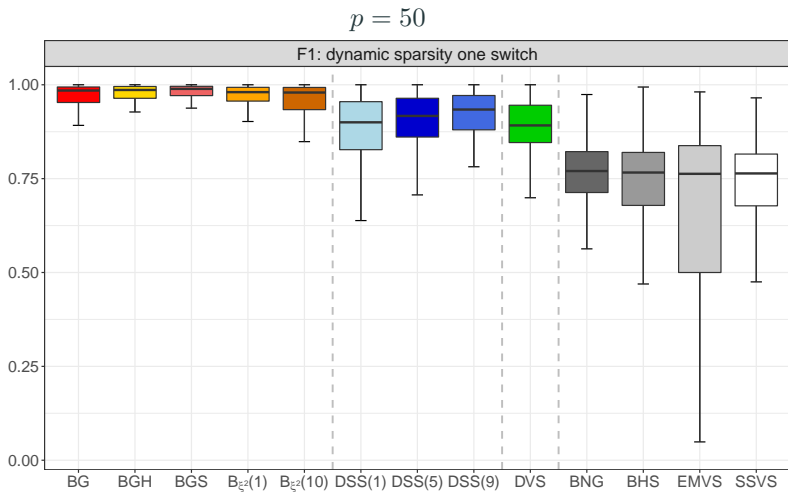
Simulation study

Scenario 1 $\implies \beta_{jt} \neq 0 \forall t$



Simulation study

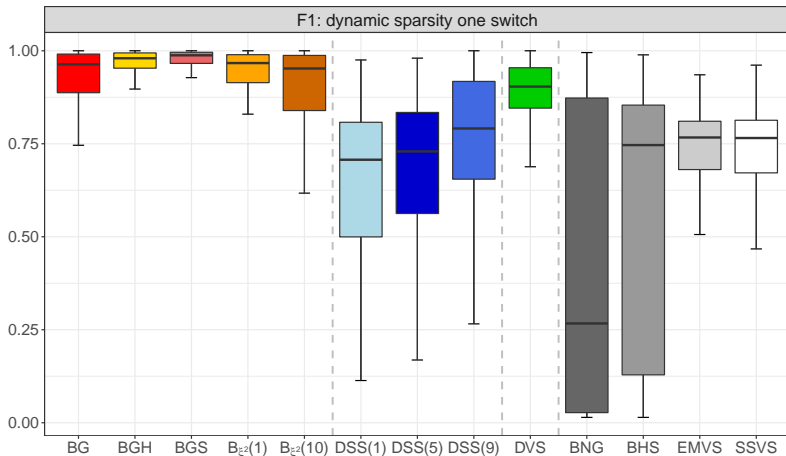
Scenario 2 \implies single switch from $\beta_{jt} = 0$ to $\beta_{jt} \neq 0$



Simulation study

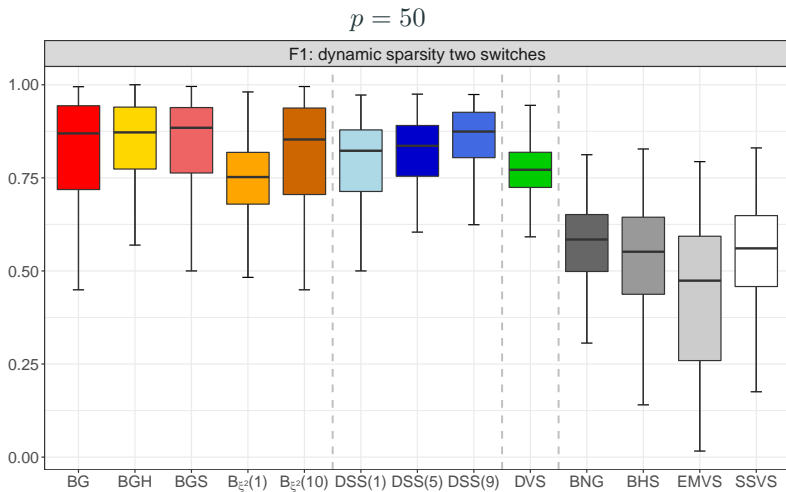
Scenario 1 \implies single switch from $\beta_{jt} = 0$ to $\beta_{jt} \neq 0$

$p = 200$



Simulation study

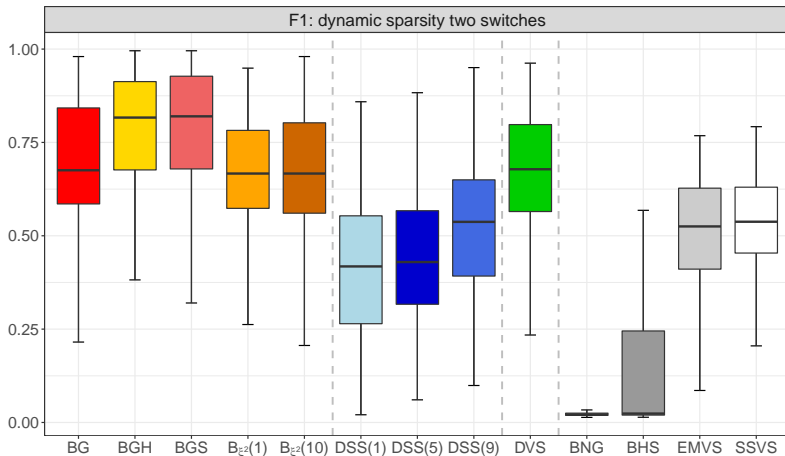
Scenario 3 – two switches from $\beta_{jt} = 0$ to $\beta_{jt} \neq 0$



Simulation study

Scenario 3 – two switches from $\beta_{jt} = 0$ to $\beta_{jt} \neq 0$

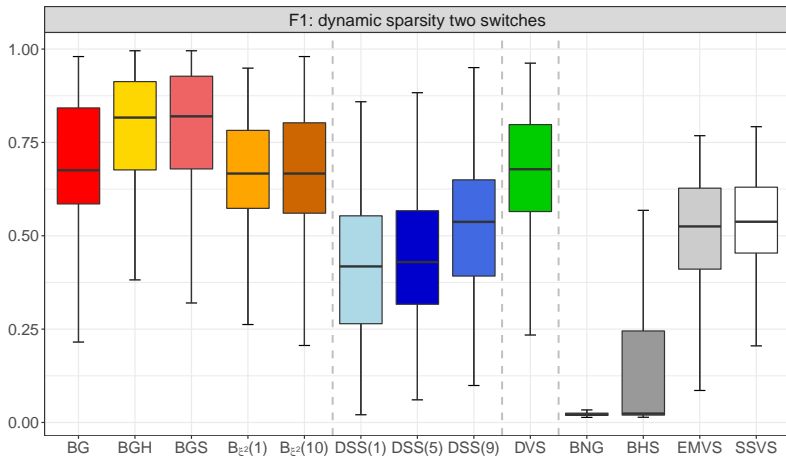
$p = 200$



Simulation study

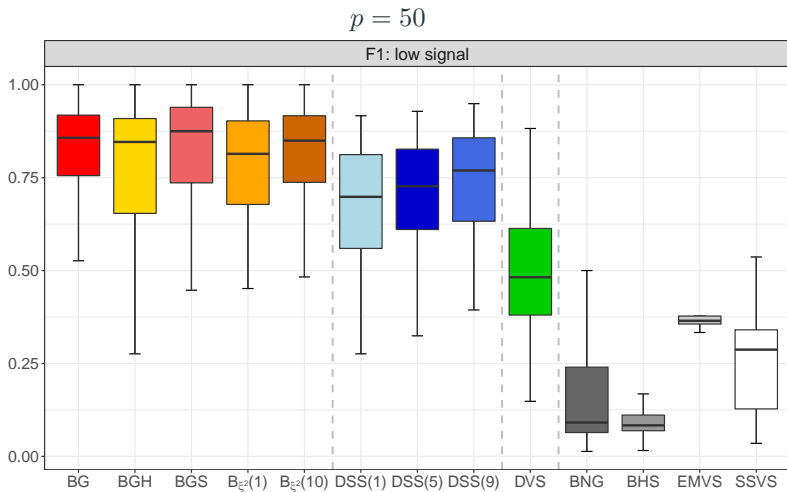
Scenario 3 – two switches from $\beta_{jt} = 0$ to $\beta_{jt} \neq 0$

$p = 200$



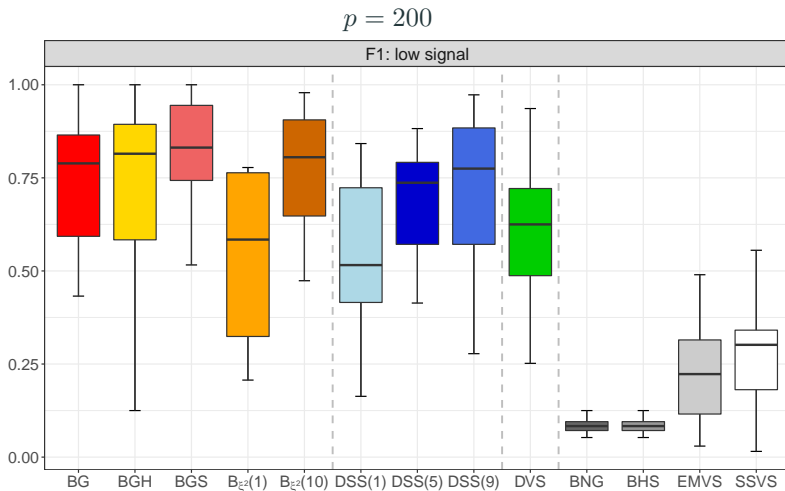
Simulation study

Scenario 4 – one short-lived switch from $\beta_{jt} = 0$ to $\beta_{jt} \neq 0$



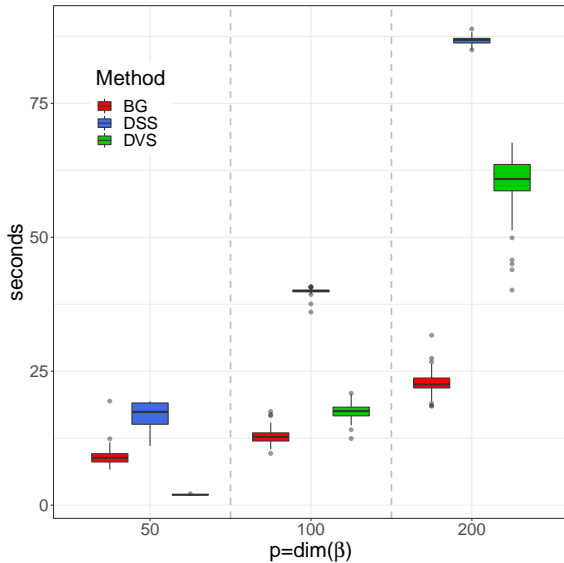
Simulation study

Scenario 4 – one short-lived switch from $\beta_{jt} = 0$ to $\beta_{jt} \neq 0$



Simulation study

Results | Running time (secs)



Inflation forecasting

Forecasting inflation based on macroeconomic variables

Empirical setting

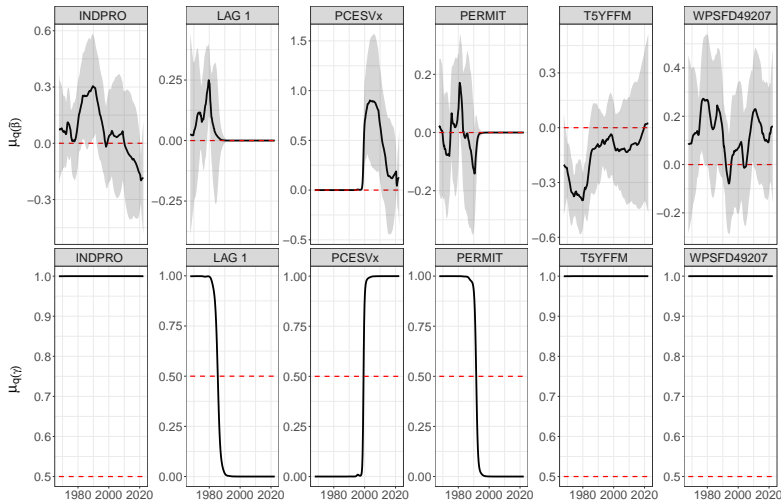
Building upon Koop & Korobilis (2020), Ročková & McAlinn (2021)

- ↪ **Target** $\Rightarrow h = 1, 2, 4, 8$ quarter-ahead inflation. Four measures of inflation: total CPI (CPIAUCSL), core CPI (CPILFESL), GDP deflator (GDPCTPI), PCE deflator (PCECTPI).
- ↪ **Predictors** \Rightarrow 229 macroeconomic variables from FRED-QD (see McCracken & Ng 2020 + 2 lags of the response (quarterly change).
- ↪ **Sample period** \Rightarrow Quarterly data 1967Q3-2022Q2
- ↪ **Forecasting benchmarks** \Rightarrow Unobserved component model (see Stock & Watson 2007) and TVP AR(1) (see Koop & Korobilis 2020).
- ↪ **Recursive forecasts** \Rightarrow 10 years “burn-in”, then recursive forecasts based on an expanding window.

Forecasting inflation based on macroeconomic variables

In-sample analysis: **Total CPI**

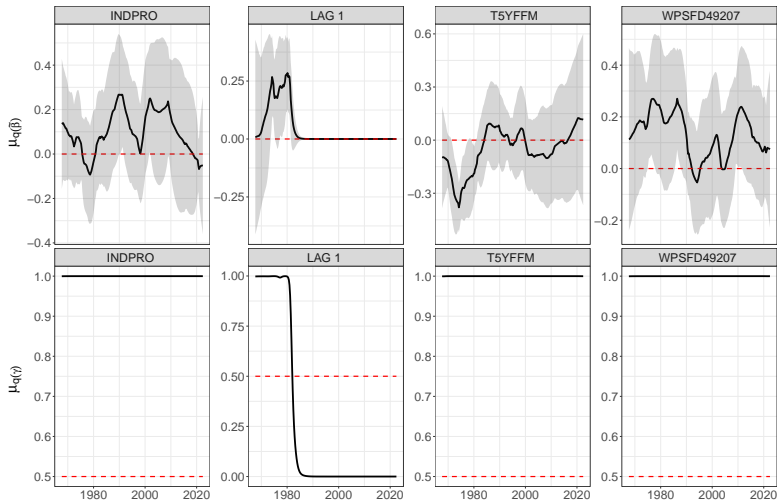
Dynamics of the **selected** regression coefficients $\gamma\beta$



Forecasting inflation based on macroeconomic variables

In-sample analysis: PCE deflator

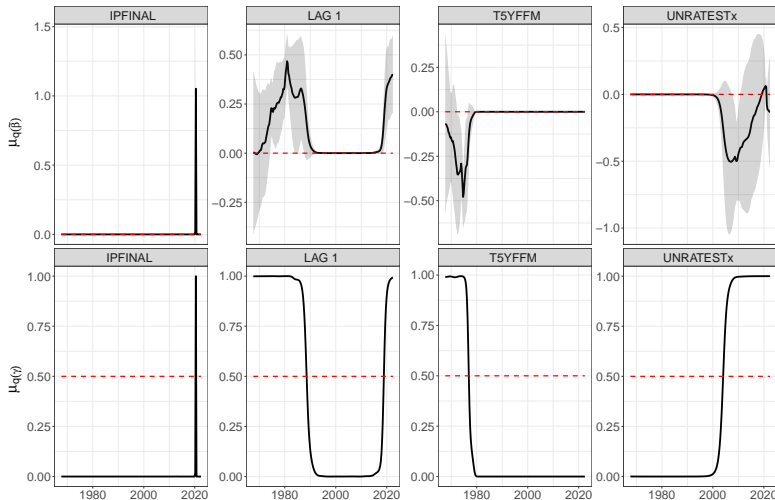
Dynamics of the **selected** regression coefficients $\gamma\beta$



Forecasting inflation based on macroeconomic variables

In-sample analysis: GDP deflator

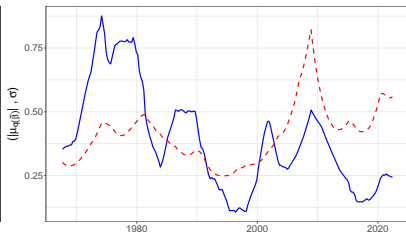
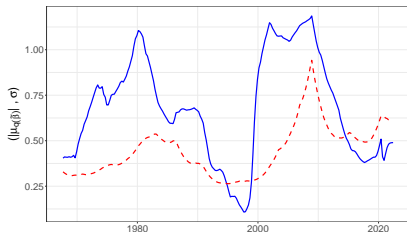
Dynamics of the selected regression coefficients $\gamma\beta$



Forecasting inflation based on macroeconomic variables

In-sample analysis: Total CPI and PCE deflator

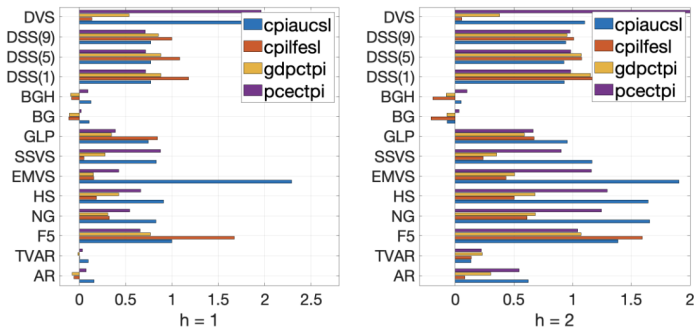
Signal $\sum_{j=1}^p |\mu_q(\beta_{jt})|$ vs $\hat{\sigma}_t$



Forecasting inflation based on macroeconomic variables

Relative mean Squared Error (benchmark unobserved component model).

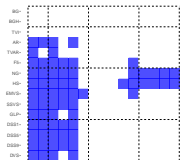
$$\sum_{t=\tau}^T e_{i,t}^2 - \sum_{t=\tau}^T e_{\text{bench},t}^2$$



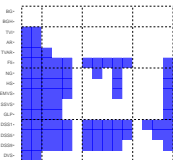
(p) Relative mean squared error

Forecasting inflation based on macroeconomic variables

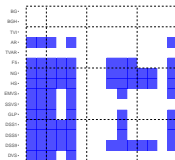
Diebold-Mariano tests



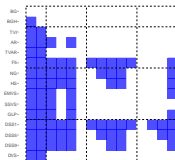
(q) CPIAUCSL



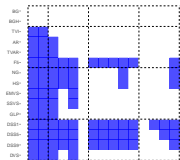
(r) CPILFESL



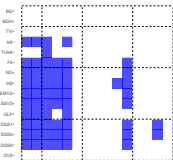
(u) CPIAUCSL



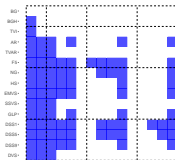
(v) CPILFESL



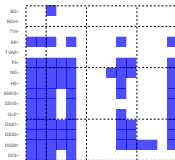
(s) GDPCTPI



(t) PCECTPI



(w) GDPCTPI



(x) PCECTPI

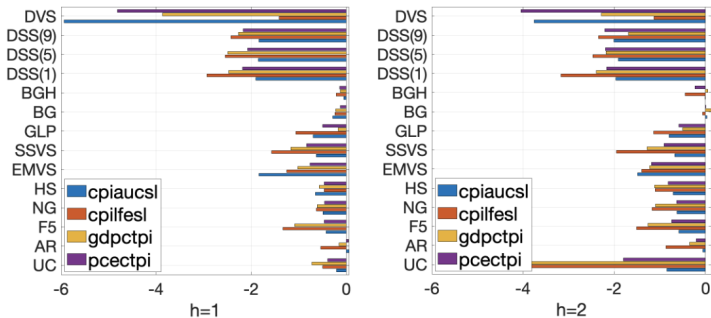
Horizon $h = 1$

Horizon $h = 2$

Forecasting inflation based on macroeconomic variables

Relative log predictive score (benchmark TV-AR(2)).

$$\frac{1}{T-\tau-1} \sum_{t=\tau}^T (\log(S_{i,t}) - \log(S_{\text{bench},t}))$$



(y) Relative log predictive score

Conclusion & what's next

This paper:

- ↪ Dynamic variable selection in large-scale time-varying predictive regressions.
- ↪ Fast and scalable semi-parametric variational Bayes algorithm.
- ↪ Competitive compared to existing variable selection methods and MCMC.

Future research:

- ↪ Extension to Generalised Linear Models.
- ↪ Dynamic group variable selection.
- ↪ Change the dependence:
 - ↪ irregular time points, spatial data, data over networks.